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

Presented by
Hernando Martínez Vergara
born in Valencia, Spain
Oral-examination: 26th of October, 2016

Descriptive and functional approaches for a system-level understanding of *Platynereis dumerilii* and the evolution of locomotor circuits in Bilateria.

Referees:

Dr. Alexander Aulehla

Prof. Dr. Hannah Monyer

Dedicated to my parents, Ana and Hernando,
to whom I owe everything

"La libertad, Sancho, es uno de los más preciosos dones que a los hombres dieron los cielos: con ella no pueden igualarse los tesoros que encierra la tierra, ni el mar encubre; por la libertad, así como por la honra, se puede y se debe aventurar la vida..."

"Freedom, Sancho, is one of the most precious gifts that heaven has bestowed upon men; no treasures that the earth holds buried or the sea conceals can compare with it; for freedom, as for honour, life may and should be ventured..."

El Ingenioso Hidalgo Don Qvixote de la Mancha

Miguel de Cervantes

Summary

The nervous system, built by the powerful capability of neurons to intercommunicate, is the most fascinating biological achievement, which through evolution has given rise to highly complex structures responsible for remarkable animal behaviours. It is therefore surprising, and at the same time highly motivating for a scientist, to realise how little we know about this evolutionary process. Traditional approaches like palaeontology and phylogenomics lack the resolution to confidently solve the history of neuronal circuits. Nevertheless, thanks to modern comprehensive techniques and integrative approaches, it is possible to molecularly, morphologically and functionally describe entire nervous systems by their constituent components, the cell types. The comparison of cell types and circuits across animals will help elucidate the different evolutionary steps that led to extant nervous systems (Arendt, 2008).

Within this thesis, the reader will find a description of the research I have conducted on the implementation of system-level approaches to characterize cell types, with the goal of achieving an integrative understanding of the nervous system of *Platynereis dumerilii*, an animal suited for these system-level evolutionary studies. A special emphasis has been given to the generation of a new automatic pipeline to build gene expression atlases for complex body plans, the design of image analysis routines to monitor and quantify animal behaviour, and the implementation of the Crispr/Cas9 technique in this organism. I also describe pioneer work to reconstruct the full connectome of the larvae at six days post fertilization, the combination of light-sheet microscopy and calcium indicators to monitor neuronal activity in thousands of neurons, and the proof-of-principle of using optogenetics to manipulate neuronal activity in *Platynereis*.

Because of the relevance of the control of muscle contraction for the evolution of nervous systems (Wilkinson, 2015), and the vast amount of information collected for various animals with regards to locomotion (Goulding, 2009), I focus my analysis on the post-mitotic *Platynereis* ventral nerve cord. I show that this structure contains the circuits responsible for the crawling behaviour, and describe in detail the kinematics of these movements, which suggest a similar organization of circuits than vertebrates and segmental protostomes. Using the expression atlas, I unbiasedly unravel the molecular substructure of the ventral nerve cord, which consists of cell types grouped into general medio-lateral domains, remarkably similar to those in vertebrates, as well as different territories unique to protostomes. I further characterize in detail a commissural cell population, showing strong similarities with both vertebrates and *Drosophila* in terms of position, molecular profile, morphology and function. These findings support the idea of an ancestral cell type, specified by a newly acquired gene, which controlled the coordination between the two sides of the body during locomotion in the bilaterian ancestor.

Zusammenfassung

Die Funktionalität des Nervensystem beruht auf der Fähigkeit von Neuronen, miteinander zu kommunizieren und ist die vermutlich faszinierendste Errungenschaft der Natur. Evolutionäre Mechanismen haben zur Entstehung von dieser höchst komplexen Struktur beigetragen, die wiederum die Grundlage für bemerkenswerte tierische Verhaltensweisen bildet. Demzufolge ist es gleichermaßen überraschend und für jeden Wissenschaftler auch ein Ansporn, dass über diesen evolutionären Entstehungsprozess nur wenig bekannt ist. Traditionelle Ansätze wie die Paläontologie oder Phylogenomik sind zu unpräzise und deshalb außerstande, zuverlässige Aussagen zur Entstehungsgeschichte neuronaler Schaltkreise zu treffen. Jedoch ist es mittels neuerer umfassender und vergleichender Methodik möglich geworden, das gesamte Nervensystem auf molekularer, morphologischer und funktioneller Ebene in einzelne Zelltypen aufzuschlüsseln. Der Vergleich von Zelltypen und neuronalen Schaltkreisen verschiedener Organismen wird dazu beitragen, die Schritte nachzuvollziehen, die zur Entstehung gegenwärtiger Nervensysteme beitrugen (Arendt, 2008).

Die vorliegende Arbeit führt den Leser an die Forschung heran, im deren Rahmen ich Methodiken zur Charakterisierung von Zelltypen auf einem Systemlevel implementiert habe. Das Ziel war hierbei, ein umfassendes Verständnis des Nervensystems von *Platynereis dumerilii* zu entwickeln, einem Organimus, der sich für systematische evolutionäre Studien besonders anbietet. Von vornehmlicher Bedeutung ist dabei die Entwicklung eines automatisierten Verfahrens zur Erstellung von Genexpressionsatlanten für komplexe Baupläne, der Aufbau von Bildprozessierungspraktiken zur Überwachung und Quantifizierung von Verhalten und die Anwendung der Crispr/Cas9-Technik in diesem Organismus. Darüber hinaus beschreibe ich grundlegende Arbeiten zur Rekonstruktion des vollständigen Konnektoms der sechs Tage alten Larve, die kombinierte Anwendung von Lichtblatt-Mikroskopie und Calcium-Indikatoren zur Analyse neuronaler Aktivität in tausenden von Neuronen und den Nachweis über die Anwendbareit optogenetischer Methoden zur Manipulation neuronaler Aktivität in *Platynereis*.

In Bezugnahme auf die Relevanz von genau gesteuerten Muskelkontraktionen für die Evolution des Nervensystems (Wilkinson, 2015) und die große Menge an Informationen, die über Bewegungsteuerung in verschiedenen Organismen zur Verfügung steht (Goulding, 2009) zentriere ich meine Analyse auf den postmitotischen ventralen Nervenstrang von *Platynereis*. Ich weise nach, dass diese Struktur die neuronalen Schaltkreise beinhaltet, die das Kriechverhalten steuern und beschreibe detailliert die Kinematik dieser Bewegungen, welche eine Netzwerkarchitektur vermuten lässt, die derjenigen der Vertebraten und segmentierten Protostomen ähnelt. Mithilfe des Expressionsatlanten identifiziere ich die grundlegende molekulare Struktur des ventralen Nervenstranges, dessen Zelltypen in mediolateralen Domänen gruppiert sind, die denjenigen der Vertebraten bemerkenswert

ähnlich sind, jedoch auch in verschiedenen Areale assemblieren, die für Protostomen spezifisch sind. Weiterhin charakterisiere ich eine Population von Kommissurenzellen, für die sowohl in Vertebraten als auch *Drosophila* entsprechende Zellgruppen mit vergleichbarer Lokalisation, molekularem Profil, Morphologie und Funktion gefunden werden können. Diese Resultate unterstützen die Theorie eines ursprünglichen Zelltyps, gekennzeichnet durch Aneignung eines neuen Gens, welcher die Koordination zwischen den beiden Körperhälften während Bewegungsphasen kontrolliert.

Acknowledgements

First of all, I would like to stress my gratitude to my supervisor Detlev Arendt for many reasons. You welcomed me into your lab, where I have found not only great science but many true friends. This combination makes it feel like I haven't worked a single day during my PhD. You offered and trusted me with the development of an interdisciplinary, ambitious and exciting project, which was exactly what I was looking for, and from which I have learnt more that I could have imagined. You mentored me along the way, and always encouraged me to pursue my own ideas, which has made the whole process a thrilling rollercoaster, as I now think every PhD process should be.

I want to thank all the Arendt lab members that I have had the privilege to work with and learnt from. Thanks to Emily Savage, for putting your heart in taking the best possible care to the very best animal model: the people in the lab. Oh yes, and for feeding and breeding *Platynereis* too. Special thanks to Paola Bertucci. We both dived into the EvoDevo field together, so deep that I ended up walking you to the aisle to marry my best friend Manuel Eguren, to whom I am also very grateful for being the other half of an awesome bromance. Thanks to Kaia Achim, for your scientific help and your limoncello, specially for your limoncello. Thanks to Thibaut Brunet and Maria Tosches, for, between countless things, suggesting that it would be fun to target Dbx1 with Crispr/Cas9. Thanks to Nicole Springer for your help with countless rounds of *in situs*. Thanks to Oleg Simakov and Tomas Larsson your help with image registration and bioinformatics. Thanks to Elia Benito Gutierrez, for being a role model to me, and for showing me how awesome *Amphioxus* is, specially because you have to fish for them in the Maldives. Thanks to Pavel Vopalenski, Silvia Rohr, Mette Handberg-Thorsager, Thomas Chartier, Heather Marlow, Jacob Musser, Daniel Bucher, Antonella Lauri, Wiebke Duerichen and Marzia Sidri, for all your help during the last four years. Thanks to Lukas Anneser for translating the summary of this thesis into a language that I haven't had the courage (and the time, German takes too much time) to learn. Thanks to David Puga and Ryan Prestil, working with you in this project has been a enormous pleasure.

Thanks to the EMBL EM facility, Yannick Schwab and Pedro Machado, for achieving a beautiful fixation of *Platynereis* and preserving so much information. Thanks to the FMI EM facility, Rainer Friedrich, Christel Genoud, Benjamin Titze and Adrian Wanner, for your gigantic efforts in achieving the best possible data to generate a full connectome of the larvae. I am truly grateful to Gáspár Jékely and his team, specially Albina Asadulina, Sanja Jasek and Luis Bezares, for helping in making the *Platynereis* community a very happy place where resources can be shared.

Thanks to the EMBL ALMF and the Image Analysis EMBL club, specially Kota Miura and Christian Tischer for your constant technical and intellectual assistance. Thanks to Gustavo Glasner Quintas de Medeiros, for letting me play with your

awesome microscope. Thanks to Rob Meijers for your assistance and data regarding the structure of axon guidance molecules. Thanks to Pedro Martínez for your sharing your acoels data with me. Thanks to Jaime Huerta Cepas for giving us such powerful tool to study evolution. Thanks to John Marioni and Jean-Baptiste Petit for your help in bringing the atlas to the world wide web.

Special thanks to my TAC committee, Alexander Aulehla, Hannah Monyer and Cornelius Gross, that during the years used their precious time in providing me with extremely valuable input that has shaped this project into something that I am very proud of.

I want to acknowledge the intellectual and ludicrous assistance of St. Flaubens College: Alexandros Drainas, Niklas Hoffmann, Simone Li, Aleksandra Bebel, Sarah Brechbuehler, Ana Rita da Silva, and Florian, Michaela and Mathilda Schur. You guys are the best of Heidelberg.

Thanks to my family, for your unconditional support from a distance.

And for the big *finale*, and the last thing I write in this book, I am truly grateful to Gemma Estrada, for being the living proof that someone can survive reading this entire thesis, and more surprising that someone can survive me. Thank you for sharing your life with me, thank you for being always there, thank you for everything you are. I love you.

Contents

List of Figures

List of Tables

List of Abbreviations

AcTub	**Ac**etylated **Tub**ulin
aINs	**a**scending recurrent inhibitory **INs**
AP	**A**nterio-**P**osterior
BNS	**B**lastoporal **N**ervous **S**ystem
CDS	**C**oding **D**NA **S**equence
ChR2	**Ch**annel**R**hodopsin **2**
CiA	**Ci**rcumferencial **A**scending INs
CiD	**Ci**rcumferencial **D**escending INs
CNS	**C**entral **N**ervous **S**ystem
CoBL	**C**ommissural **B**ifurcating **L**ongitudinal
CoSA	**C**ommissural **S**econdary **A**scending
CPG	**C**entral **P**attern **G**enerator
CS	**C**ryptic **S**egment
CSF-cNs	**C**erebro**S**pinal **F**luid-**c**ontacting **N**euron**s**
DAPI	4,6-**dia**midino-2-**p**heny**l**indole
dpf	**d**ays **p**ost **f**ertilization
DV	**D**orso-**V**entral
EINs	**E**xcitatory **I**nter-**N**eurons
EvoDevo	**Evo**lution and **Dev**el**o**pment
FNSW	**F**iltered **N**atural **S**ea **W**ater
GUI	**G**raphical **U**ser **I**nterface
hpf	**h**ours **p**ost **f**ertilization
iCINs	**i**nhibitory **C**ommissural **INs**
IHC	**I**mmuno **H**isto**C**hemistry
INs	**I**nter **N**euron**s**
IPSPs	**I**nhibitory **P**ost**S**ynaptic **P**otential**s**
L-INs	ipsilaterally inhibitory **INs**
MCoD	**M**ultipolar **Co**mmissural **D**escending
ML	**M**edio-**L**ateral
MNs	**M**otor**N**euron**s**
Mya	**M**illion **y**ears **a**go
NSW	**N**atural **S**ea **W**ater
PCA	**P**rincipal **C**omponent **A**nalysis
PduLoco	*Platynereis **du**merilii* **Loco**motion monitoring tool
PNS	**P**eripheral **N**ervous **S**ystem

PrImR	**Pr**ofiling by **Im**age **R**egistration
ProSPr	**Pro**filing by **S**ignal **Pr**obability mapping
RCs	**R**enshaw **C**ell**s**
RT	**R**oom **T**emperature
SBFSEM	**S**erial **B**lock **F**ace **S**canning **E**lectron **M**icroscopy
SEM	**S**canning **E**lectron **M**icroscopy
SNPs	**S**ingle **N**ucleotide **P**olymorphism**s**
SPIM	**S**ingle **P**lane **I**llumination **M**icroscopy
TDE	2,2´-**t**hio**di**e**e**thanol
TEA	**T**ri**e**tanol**a**mine
TEM	**T**ransmission **E**lectron **M**icroscopy
TFs	**T**ransctiption **F**actor**s**
tSNE	**t**-distributed **S**tochastic **N**eighbour **E**mbedding
TSS	**T**ranscription **S**tart **S**ite
UCoD	**U**nipolar **Co**mmissural **D**escending
VeLD	**Ve**ntral **L**ateral **D**escending
VeMe	**Ve**ntral **Me**dial
VNC	**V**entral **N**erve **C**ord
WMISH	**W**hole **M**ount *In Situ* **H**ybridization

Chapter 1

Introduction and motivation to the project

In this initial chapter I outline the main scientific motivations that have driven the project presented in this dissertation. Similarly, I reason about the techniques and approaches selected to partially solve the scientific questions that first conceived this project, and the many more that arose during its development, with special emphasis on the benefits of *Platynereis dumerilii* as an evolutionary model species.

1.1 Nervous system evolutionary challenges: *Urbilateria*

To fully comprehend the functional significance of the nervous system there are many still unresolved questions regarding the reconstruction of its ancient states and its evolution in different metazoan lineages. Of special evolutionary interest is the reconstitution of the animal that gave rise to the bilaterian lineage, named *Urbilateria* (the common ancestor of all known extant animals except for sponges, placozoa, cnidarians and ctenophores). The structure and composition of the nervous system of *Urbilateria* is still a big topic of dispute in the EvoDevo community, and no general consensus has been achieved, specially regarding the homology or convergence of centralized nervous systems (Arendt et al., 2008; Holland et al., 2013; Strausfeld and Hirth, 2013; Holland, 2016). On the one hand, the molecular specification of neurogenic territories are strikingly similar across bilaterians (Denes et al., 2007; Arendt, Tosches, and Marlow, 2016). On the other hand, the remarkable plasticity of the nervous system has been a key driver of evolution given the importance of behaviour for adaptation, which has resulted in a wonderful diversity of neuronal-based structures and animal capabilities.

Given the developmental similarity and the structural and functional diversity between extant nervous systems, it is highly intriguing to think about the following questions. What kind of nervous system did the bilaterian ancestors possess? What were the functional roles of the different cell types they were composed of? Which

environmental signals could these ancestors detect, and more importantly, what were their behavioural capabilities to respond to them? How did these cell types evolve in the different bilaterian lineages and give rise to the extant brains and nerve cords? All of these uncertainties point to the following main question: **What kinds of cells arise from the similar neurogenic territories in different animals?** The analysis and comparison of nervous system cell types across bilaterians will most likely help to resolve these unanswered questions.

1.2 The cell type model: a conceptual working framework

Besides the insights gained by exploring the palaeontological record, the only approach to reconstruct the history of metazoans is by comparing the features of extant animals. Given that the evolutionary process happens mainly in a progressive manner, parsimonious reasoning is applied to infer the temporal relationship of animals and therefore imagine how their extinct ancestors looked and behaved. The reconstruction of these ancestors helps a great deal towards the understanding of the distinct evolutionary paths that metazoan life took, but the kind of information that can be obtained about them depends deeply on the features that are being compared.

In the beginning of evolutionary science, whole body-plans and organs were studied to infer homologies and homoplasies, and the first evolutionary trees and temporal splits between animals began to be elucidated. Access to the genetic code (DNA and proteins) and machine computation represented a boost of power to investigate phylogenies, and as of today, the community has reached a consensus on the position of the major animal phylums (with a few still debated exceptions) in the evolutionary tree. However, just the genetic content is not enough to draw strong phenotypic conclusions about the ancestors, as this information is not encapsulated into a functional context. Animals are biological systems composed of diverse cells, each of them providing a different functional context of the same genome.

The different functional units that compose every animal are usually referred to as cell types. Given that the function of a cell depends mainly on its protein complement, a particular cell type can be defined by the set of genes that it expresses[1], and as cell types are subjected to gradual evolutionary changes, it can then be considered that this molecular signature recapitulates their evolutionary history. This cellular perspective on evolution is regarded as the **cell type model** (Arendt, 2008).

[1]Note that the proteome (which reflects the functional tool-kit of the cell) is the product of the transcriptome and the subsequent post-transcriptional and post-translational modifications. Nevertheless, the set of transcripts that a cell expresses are a common measure for its identity as it is generally accurate and easier to access technologically.

Under this framework, the cell type complement of extant animals can be compared to infer their inheritances, losses and divergences in the different metazoan evolutionary paths, and ultimately rely on them to achieve a deeper and more accurate analysis of ancestral phenotypes and body plans.

1.3 System-level understanding of *Platynereis dumerilii*

In order to compare cell types within and across organisms, it is necessary to use the molecular information in a system-wide and unbiased manner. *Platynereis dumerilii* is an emergent marine laboratory animal model, very powerful for evolutionary studies (see section 2.1), that represents an ideal opportunity for the *in toto* analysis of its cell type complement. This is now possible thanks to the recent convergence of technologies that allow the investigation of different cellular properties in a system-level manner.

- The development of image manipulation algorithms (e.g. 3-dimensional registration) permit the generation of 3-D gene expression atlases that can render the spatial location of hundreds of genes to study gene co-expression *in silico* (Tomer et al., 2010; Asadulina et al., 2012).
- The increased sensitivity of sequencing technology, coupled with microfluidics platforms, can be used to collect single cells and sequence their transcriptome, achieving an almost-complete picture of gene expression for tens of individual cells in one experiment (Achim et al., 2015).
- Serial block face scanning electron microscopy (SBFSEM) can be used to reconstruct with high resolution the cellular morphology for the entire body of a *Platynereis* larvae and accurately trace all the neuronal connectivity (Randel et al., 2014; Randel et al., 2015).
- The application of the prokariote immune system Crispr-Cas9 to metazoans has boosted the capability (technological and economical) for genome editing, allowing genetic functional assessment in model and non-model organisms.
- Modern neuronal activity reporters (e.g. GCaMP) in conjunction with single-plane illumination microscopy (SPIM), permits the monitoring of whole-body neural activity.

In order to achieve a system-level understanding of *Platynereis dumerilii*, the application and integration of these technologies in this organism is the major technical goal of this project. Following the cell type model and the application of these methods, the focus of the analysis has been the molecular description of *Platynereis* post-mitotic ventral nerve cord (VNC), and given its importance for locomotion,

the well-known locomotor circuits in vertebrates have been used for comparison, with an emphasis on the functional study of commissural neurons.

1.4 The evolution of locomotor circuits

Across different phyla, Central Pattern Generators (CPGs) constitute the neural substrate for locomotion. CPGs are neural circuits that generate an autonomous and coordinated pattern and rhythm. The straight forward readout of the output of these circuits motivated their study in the field of neurobiology more than a hundred years ago. Initial studies were done in cats at the beginning of the 20th century. By the middle of the century the main research shifted towards simpler invertebrates, where the concepts of CPGs organisation started to emerge. Functional circuitry data (mainly achieved by electrophysiology) has been produced for locomotor CPGs in the locust and stick insect, crayfish and lobster, *Clione*, *Tritonia*, *Aplysia*, and the leech. As a vertebrate, lamprey is arguably the best understood system thanks to its limited number of neurons and easily recognisable segmentation of the spinal cord. Thanks to modern genetic manipulation techniques, CPGs began to be studied in more complex vertebrate models, and nowadays there is molecular and functional data for *Xenopus* tadpole, zebrafish embryo and adult, mouse and rat (see Goulding, 2009; Kiehn, 2006; Grillner, 2003 for reviews).

One of the main unresolved evolutionary topics of bilaterians is their origin, in particular the body plan and cell type complement of the common ancestor of protostomes and deuterostomes (*Urbilateria*). Due to the vast amount of information collected on the neural circuits of the locomotor CPGs across different bilaterians, they represent an attractive framework to apply the approach of cell type evolution (Arendt, 2008), in order to gain insight into the origins of centralised nervous systems (CNS).

Approaching the comparison between the locomotor CPGs of protostomes and deuterostomes is not trivial for one main reason: the different nature of the information collected for different cell types in both groups. In protostomes, the systems used for study are chosen (at least in part) for their simplicity in distinguishing and recognising cell types by their morphology. Indeed, much of the information concerning general neuronal function comes from the electrophysiology experiments performed in protostomes CPGs. However, except for the two main ecdysozoan laboratory models (*drosophila and C.elegans*), very little molecular and developmental data is available for this group of animals. On the other hand, the research within deuterostomes has described cell population behaviours rather than individual neurons, due to the vast amount of cells in the CNS of these organisms. During the last decades, vertebrates have been the focus of research because they benefit from the modern genetic techniques, rendering valuable information about

the molecular profile and function of CPGs cell populations, as well as schemes and knowledge about the major synaptic interactions between groups of neurons. This research has allowed the start of the comparison of CPG cell types within vertebrates (Goulding, 2009). There is however, the need to extend these comparisons to protostomes in order to reconstruct the cell type complement involved in locomotion in the ancestral *Urbilateria* and understand the functional evolution of the CNS. The use of the lophotrochozoan annelid *Platynereis* as a model organism is ideal for this task given its phylogenetic position, ancestral features (see Section 2.1), technical advantages (see Sections 1.3 and 2.1), and available information about the evolutionary relationship of the VNC with neuronal structures in vertebrates (see Section 2.1.4).

1.5 Commissures: a bilaterian innovation

Of the cell types that constitute the locomotor CPGs in bilateral animals, the cells that cross the midline (commissural neurons) are of special evolutionary interest. These neurons are instrumental in the coordination of both sides of the body so locomotion can be effectively directed in the anterio-posterior axis. Comparative analyses between the molecular patterning of the blastoporal region in cnidarians and the bilaterian neural plate have added support to the notion of common origin between these two structures (Arendt, Tosches, and Marlow, 2016). Following this scenario, the bilaterian nerve cords would have evolved from an ancestral nerve ring, making the appearance of commissures a truly bilaterian innovation. The question whether commissural neurons are homologous between protostomes and deuterostomes is key in the reconstruction of the body plan, nervous system organization, and behavioural capabilities of their common ancestor *Urbilateria*.

1.6 Specific objectives and major contributions of the project

Achieving a system-level understanding of such a complex biological structure like a nervous system requires not only the implementation of descriptive methods that can render information about the system's properties, but also the development of genetic tools to test the functionality of neurons and circuits. A main aim of my PhD work has been the development and implementation of tools to describe properties of nervous system and cell types, as well as methods to assess their functional roles. Those approaches that have been the major focus of the work are described in the Results chapter. The results I have obtained for the other parts of the project, which are still under development at the moment of writing this thesis, can be found in the Appendix section A.

1.6.1 Descriptive methods

Expression atlas for complex body plans

By the time developing *Platynereis* larvae are able to perform a coordinated locomotion indicative of a CPG, the body plan of the animal is considerably more complex than in earlier stages for which expression atlases are available. I investigated the applicability of different image registration methods to an older developmental stage, and developed a computational pipeline for the automatic generation of gene expression atlases for more complex body plans (see Results section 3.1).

Analysis of the molecular structure of *Platynereis* VNC at 6dpf

Using the gene expression atlas that I generated, I focus the analysis on the region containing the cell types that control the animal locomotion, the ventral nerve cord (VNC). I describe the findings of a molecular subdivision of this region, determined by unbiased and systematic methods (see Results section 3.2).

Reconstruction of *Platynereis* nervous circuits

The cellular molecular fingerprint is not enough information to fully understand the functional complexity of nervous circuits. Together with the EMBL EM facility, we investigated and optimised the experimental conditions to prepare *Platynereis* larvae for SBFSEM analysis, with the ultimate goal of achieving the full reconstruction of the neural system in this animal (see Appendix section A.1).

Monitoring and quantifying locomotory behaviour in the larvae

In order to study the behavioural output of neural activity such as locomotion, unbiased and automatic monitoring and tracking tools are necessary for efficiently processing the data. I have developed a computational tool to semi-automatically track and analyse the crawling movement sequences of *Platynereis* larvae, which I have used to quantitatively describe this behaviour (see Results section 3.3), as well as to assess mutant phenotypes (see Results section 3.4).

1.6.2 Functional methods

Functional investigation of locomotor cell types

Genetic editing is a crucial approach for investigating the functional role of genes and cell types, which until very recently has been a drawback in non-model animals. I have investigated the applicability of the Crispr/Cas9 technology to *Platynereis*, and used it to functionally investigate the role of a cell type (commissural neuron) defined by the transcription factor Dbx1 (see Results section 3.4).

Assessment of neuronal function in *Platynereis*

Neurons are characterised by the fast transmission of electric currents encoding information, and so both the observation and manipulation of this feature are necessary to functionally study neural circuits. I explored the possibilities to image the activity of the complete nervous system of *Platynereis* using genetically encoded calcium and voltage sensors coupled with SPIM microscopy (see Appendix section A.2). I also performed a proof-of-principle experiment to manipulate neural activity using the optogenetic tool Channelrhodopsin (ChR2) in this animal (see Appendix section A.4).

Chapter 2

Scientific background

2.1 The animal model *Platynereis dumerilii*

2.1.1 Phylogenetic significance

Platynereis dumerilii is a polychaete (bristle worm) that lives in shallow sea waters around the globe. As a member of the phylum Annelida, it belongs to the Lophotrochozoans, the one bilaterian super-phylum under-represented in molecular studies, given that the main model organisms belong to the other two (Ecdyzozoa and Deuterostomia, see Figure 2.1). Several studies have shown that *Platynereis* has retained more ancestral features (Tessmar-Raible and Arendt, 2003; Morris and Peel, 2008) and has a less derived cell type complement (Denes et al., 2007; Christodoulou et al., 2010; Marlow et al., 2014) and genomic architecture (Raible et al., 2005) than other protostome models (Raible and Arendt, 2004) (Protostomia includes Ecdyzozoa and Lophotrochozoa). Given its phylogenetic position and its slow evolutionary rate, *Platynereis* represents a very powerful system to unravel the evolution of bilaterians.

2.1.2 Technical advantages

Besides it's phylogenetic significance, *Platynereis* is an excellent laboratory model organism given it's technical advantages regarding the applicability of modern molecular approaches. Gene expression analyses can be done in a medium throughput manner, and the stereotypicity of the animals can be used to register the images creating virtual atlases (see Section 2.3) for many developmental stages (Tomer et al., 2010; Asadulina et al., 2012). The larvae are transparent during the first weeks of development, which facilitates whole-body imaging using modern microscopy and fluorescent probes, even *in vivo*. Zygote microinjection is an established technique routinely performed in the lab, allowing the implementation of functional approaches in this animal. These aproaches include morpholino loss of function assessments (Williams, Conzelmann, and Jékely, 2015), overexpression analyses and fluorescent labelling by mRNA delivery, Crispr-mediated knock-downs and

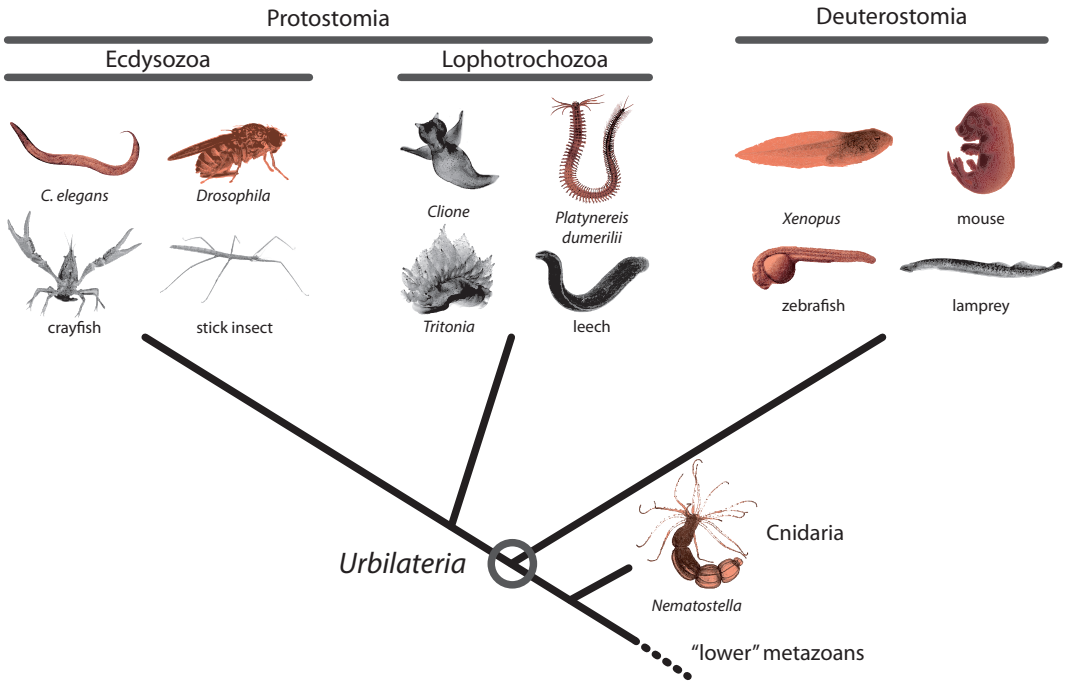

FIGURE 2.1: Main bilaterian super-phyla and its representative animal models. Phylogenetic tree showing the relationships between the three main bilaterian super-phyla. The animals represented in the tree have been selected based on the best understood models for the neurobiology of locomotion. Coloured in red, those model systems for which high-throughput molecular information is (and can be in the short term) available. Note that this tree is highly incomplete in terms of phyla representation, which also resembles the lack of molecular data for the major part of the bilaterian tree. Images modified from google.com.

knock-ins (this project), and neuronal activity monitoring using GCaMP and electrophysiology (Tosches et al., 2014), among many others. *Platynereis* genome is sequenced and assembled (unpublished data from Dr. Thomas Larsson and Dr. Oleg Simakov), there are several transcriptomes available, and transgenic lines have been established (Backfisch et al., 2013). Pharmacological treatments are easy to perform by just dissolving the drug in water, as the animals have a permeable cuticle (Denes et al., 2007; Tosches et al., 2014). The small size of the *Platynereis* larvae allows the reconstruction of neural circuits on the entire animal using electromicroscopy (Randel et al., 2014), and given that the central nervous system (CNS) of *Platynereis* larvae is composed of a small number of cells (few hundreds) at stages where complex behaviours are produced, it is possible to understand this system at the cellular level (see also section 1.3). These reasons and the easy daily access to thousand of developmentally synchronous larvae (see next section), makes *Platynereis* suited to be studied exhaustively in terms of gene expression, cell type complement, connectivity, and functional analysis.

2.1.3 Development and life cycle

Platynereis dumerilii has been cultured in the lab for more than 60 years, and therefore its life cycle and early development are very well characterized (Fischer, Henrich, and Arendt, 2010). The life cycle is biphasic, meaning there is a pelagic (free-swimming) stage and a benthic (floor settled) stage. Adults spawn the gametes in the ocean water where fertilization occurs. The fertilized oocytes sink to the bottom and extrude a gelatinous coat for protection and they start developing by spiralian cleavages. Fate mapping studies (Ackermann, Dorresteijn, and Fischer, 2005) and more recent live-imaging recordings (Dr. Maria Tosches, Dr. Pavel Vopalensky and Dr. Mette Thorsager, unpublished) have begun to trace the full cell lineage of the early development.

After one day [1], they develop into the first larval stage called trochophore. Trochophores are spherical and are composed of 2000 to 4000 cells, measuring around 160μm in diameter (Figure 2.2). A ciliary band that allows the animals to swim, divides the body into two halves: the episphere, which contains the brain, and the hyposphere or trunk, where the ventral plate, stomodeum, future gut and main musculature are located. During more or less five days, the trunk develops into three segments, each of them bearing two bilateral parapodia that the animal uses to crawl in the substrate. Up to this point, the energy that animals need for movement and development comes from lipid droplets inherited from the oocyte.

[1] Development is temperature dependent. Timing mentioned here refers to conditions of 18 degrees Celsius.

After 6 days of development the animals are able to start feeding, and at this point the development becomes asynchronous. 6 days post fertilization (dpf) corresponds to the nectochaete stage (Figure 2.2), and it is a transitory phase when the animals change from a ciliary-based swimming life style to an exploration of the substrate using periodic parapodial movements [2]. At this stage the animals are composed of more or less 8000 cells and their length is around 280μm.

The animals then start the maturation phase, when they start eating algae and progressively growing new segments from the posterior end of the body. Eventually they secrete a proteinaceus tube using the many specialized glands in their body (Daly, 1973), where they reside, leaving it only to feed. The maturation phase lasts between 2 and 5 months under lab culture conditions and most animals reach between 2 and 4 centimetres in length. Sexual maturation follows a lunar cycle, that can be replicated in the lab using a simple bulb. The lunar/bulb control of sexual maturation ensures a population synchronization that results in controlled spawning periods, achieving thousands of synchronously developing larvae.

2.1.4 Neurogenesis: Patterning of the trunk Nervous System

The neuroectoderm of *Platynereis* trochophore larvae, located in the trunk (see Figure 2.2), develops into a rope-ladder nervous system with discrete ganglia (similar to arthropods). This nervous system is also referred to as blastoporal nervous system (BNS), following the evolutionary hypothesis that it develops from tissue surrounding the blastopore and is therefore homologous to the oral ring of cnidarians (Arendt, Tosches, and Marlow, 2016). During the process of gastrulation, the ectodermal tissue fuses on the ventral side of the animal in a zipper-like manner from posterior to anterior creating a slit-blastopore between 24 and 48 hours of development (Denes et al., 2007). The initial waves of neurogenesis of the trunk neuroectoderm occur between 32 and 144 hours post fertilization (hpf), where the tissue couples cell intercalation and convergence-extension morphogenetic movements (Steinmetz et al., 2007) with cell proliferation (see Results section 3.1.1) to form three pairs of ganglia (one in each segment) surrounding two main longitudinal axonal bundles connected by commissures and forming a loop with the neuropil in the brain (see Figure 2.2). This structure (ganglia and nerve tracks) is located ventrally and referred to as ventral nerve cord (VNC). During neurogenesis, the process of cell proliferation and differentiation follows a ventro-dorsal (or apico-basal) direction creating three defined regions in the ventral plate (Denes et al., 2007; Simionato et al., 2008). Neuroblasts are located in the most ventral region, where they proliferate by asymmetric division to generate neuronal progenitors, that occupy the medial region, and divide into developing neurons that differentiate in the most basal layer, above the axonal tracks.

[2] Animals at 6dpf can still swim by cilia, and they very often do so as it is a more efficient way of travelling longer distances.

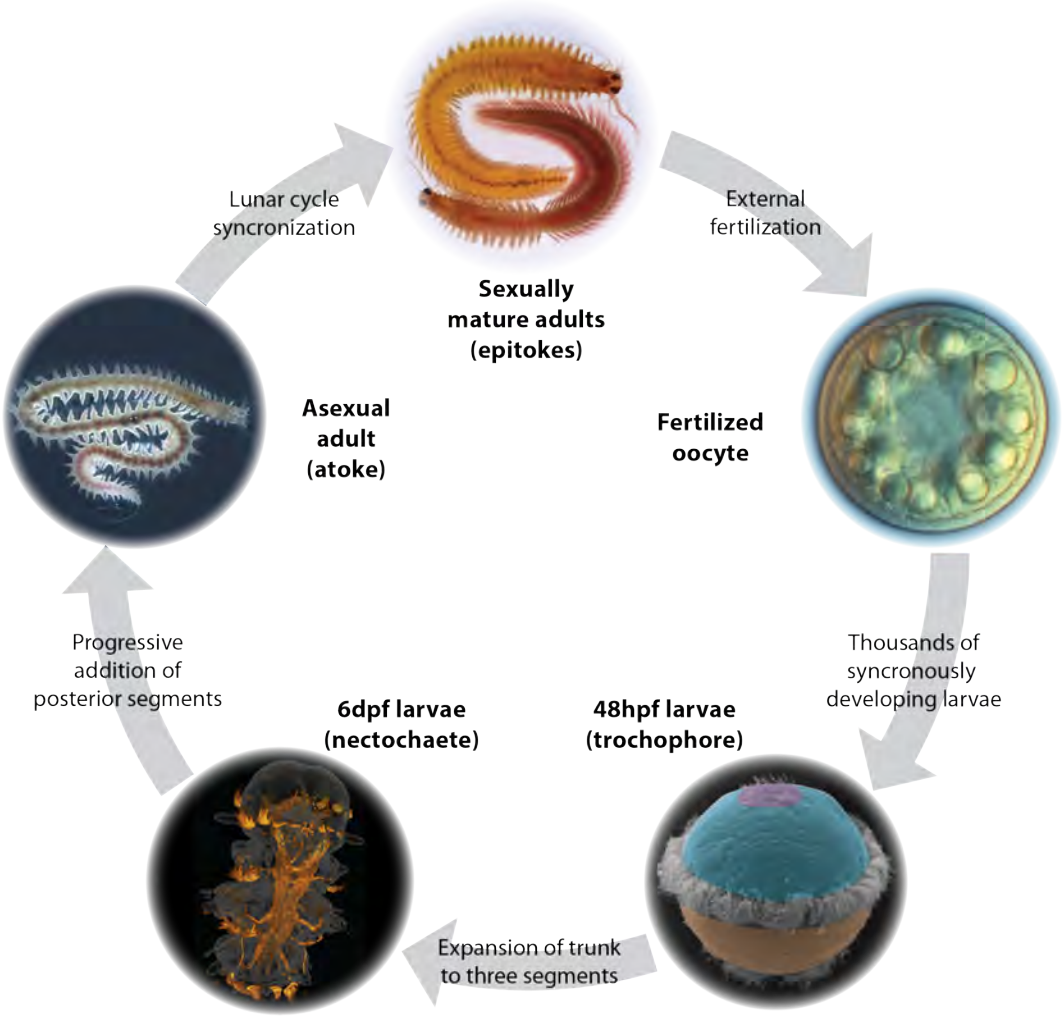

FIGURE 2.2: Life cycle of the annelid *Platynereis dumerilii*.
Main stages (especially relevant to this project) of the life cycle of *Platynereis*. From the top and clockwise: Sexually mature male (red colour) and female (yellow colour). Fertilized oocyte (note the transparency and the lipid droplets). Scanning electron micrograph of the 48hpf larvae, highlighting the apical organ (purple), the episphere (blue) and the hyposphere (orange), the last two divided by the ciliary band. 3D render of acetylated tubulin immunostaining in the 6dpf larvae showing the nerves in orange highlighting the rope-ladder morphology of the ventral nerve cord. Asexual adult showing the multiple parapodial-bearing segments. [Except for the 6dpf case, images modified from google.com].

Platynereis has been subject of many studies that have uncovered a high similarity with vertebrates regarding different aspects of this initial neurogenic program. Examples of these similarities include the function and expression of homeobox genes in body patterning and segment formation (Saudemont et al., 2008; Steinmetz et al., 2011; Kulakova et al., 2007; Pfeifer, Dorresteijn, and Fröbius, 2012), the columnar arrangement of homeobox genes during early neurogenesis (Denes et al., 2007), the control of neuronal proliferation and differentiation by the Wnt/B-catenin pathway (Demilly et al., 2013), and the role of bHLH factors on neuronal specification (Simionato et al., 2008; Demilly et al., 2011). Of special relevance for this project are the similarities regarding the medio-lateral regionalization of the trunk neuroectoderm by transcription factors (TFs), and the anterio-posterior specification of the VNC ganglia by Hox genes.

Medio-lateral patterning

In *Platynereis*, Paired, NK, Forkhead and LIM homeodomain TFs are arranged along the neuroectoderm medio-lateral axis, patterning it into discrete regions, each characterized by the developmental expression of a specific set of TFs (Denes, 2007; Denes et al., 2007). The most medial region corresponds to the midline, characterized by the expression of *foxA*, *nk2.2*, *nk6* and *sim*. As in the case of insects and vertebrates, it expresses axonal chemoattractants that guide the commissures to cross the midline, like *netrin* and *slit*. *nk2.2* forms a boundary of non-overlapping expression with *pax6* that defines a domain in the middle of the neuroectoderm, whereas *nk6* forms a more lateral boundary with *pax3/7*, whose expression extends more lateral than that of *pax6* and ovelaps with the even more lateral expression of *msx*. The patterning produced by these five genes generates five main mediolateral regions in *Platynereis* trunk:

- *nk2.2+/nk6+* domain, comprising both the midline and the adjacent cells. This is the region where the serotonergic cells arise, similarly to the situation in the vertebrate hindbrain.
- *nk6+/pax6+* domain, where differentiation markers such as *hb9, dbx1, chx10* and *evx* are expressed.
- *pax6+/pax3/7+* domain. This region corresponds to the most medial expression of *pax2/5/8* (that extends to the most lateral domain of the ventral plate).
- *pax3/7+/msx+* domain.
- *msx+* domain, where sensory neuron markers are expressed, such as *ath* and *lhx2/9*.

This combinatorial code of TFs expression and the subsequent regional specification of the neuroectoderm is remarkably similar to the patterning of the posterior

part of the vertebrate neural tube (hindbrain and spinal cord) (Arendt and Nübler-Jung, 1999; Denes et al., 2007) (see Figure 2.3 and Section 2.2.1).

Anterio-posterior patterning

The anterio-posterior (head to tail) axis of *Platynereis* also follows a sequential arrangement of TF expression. The PRD homeobox *otx* is expressed in the anterior end of the nervous system, limiting posteriorly with the expression of the ANTP-HOXL homeobox *gbx*, expressed anteriorly in the trunk. This expression dichotomy coincides with the anatomical separation of the head and the trunk in *Platynereis*, the Midbrain-Hindbrain boundary in vertebrates, and the Protocerebral-Tritocerebral boundary in arthropods (see Figure 2.4). The first chaetigerous segment (the anterior cryptic segment cephalizes during metamorphosis) of the trunk correlates with the expression of *hox1*, followed posteriorly by the expression of *hox4* in the second segment, and *hox5* expression in the third segment. The correlation between this sequential arrangement of TFs and body segmentation is very similar to the scenario described in arthropods (Steinmetz et al., 2011), which has been compared extensively with the tripartite brain organization of the vertebrate nervous system (Reichert, 2005; Hirth et al., 2003), (see Figure 2.4).

The anterio-posterior patterning described above is orthogonal to the previously mentioned medio-lateral patterning. Such a cartesian subdivision of the trunk neuroectoderm, which is initiated by morphogen signal and strongly defined by cell-fate cross-repression, provides positional information for the differentiation of neuronal types and the correct projection of axons. Given this similar arrangement of neurogenic territories during BNS developmental specification in bilaterians, what can be said about the cell types that arise from these territories? How do they compare between deuterostomes and protostomes? What can we infer about the cell type complement of *Urbilateria* BNS by this comparison?

2.2 Locomotor circuits in deuterostomia

Across the different phyla of *Bilateria*, the BNS comprises two main functions: digestion and animal body movement. This thesis is focussed on the study of the latter: the neural circuits controlling locomotion. These circuits are assembled into autonomously rhythmic networks termed Central Pattern Generators or CPGs (see 1.4). What is the evolutionary history of the cell types constituting the locomotor CPGs? In this section I will briefly describe the development and architecture of locomotor CPGs in Deuterostomes. This information will be used (see Discussion

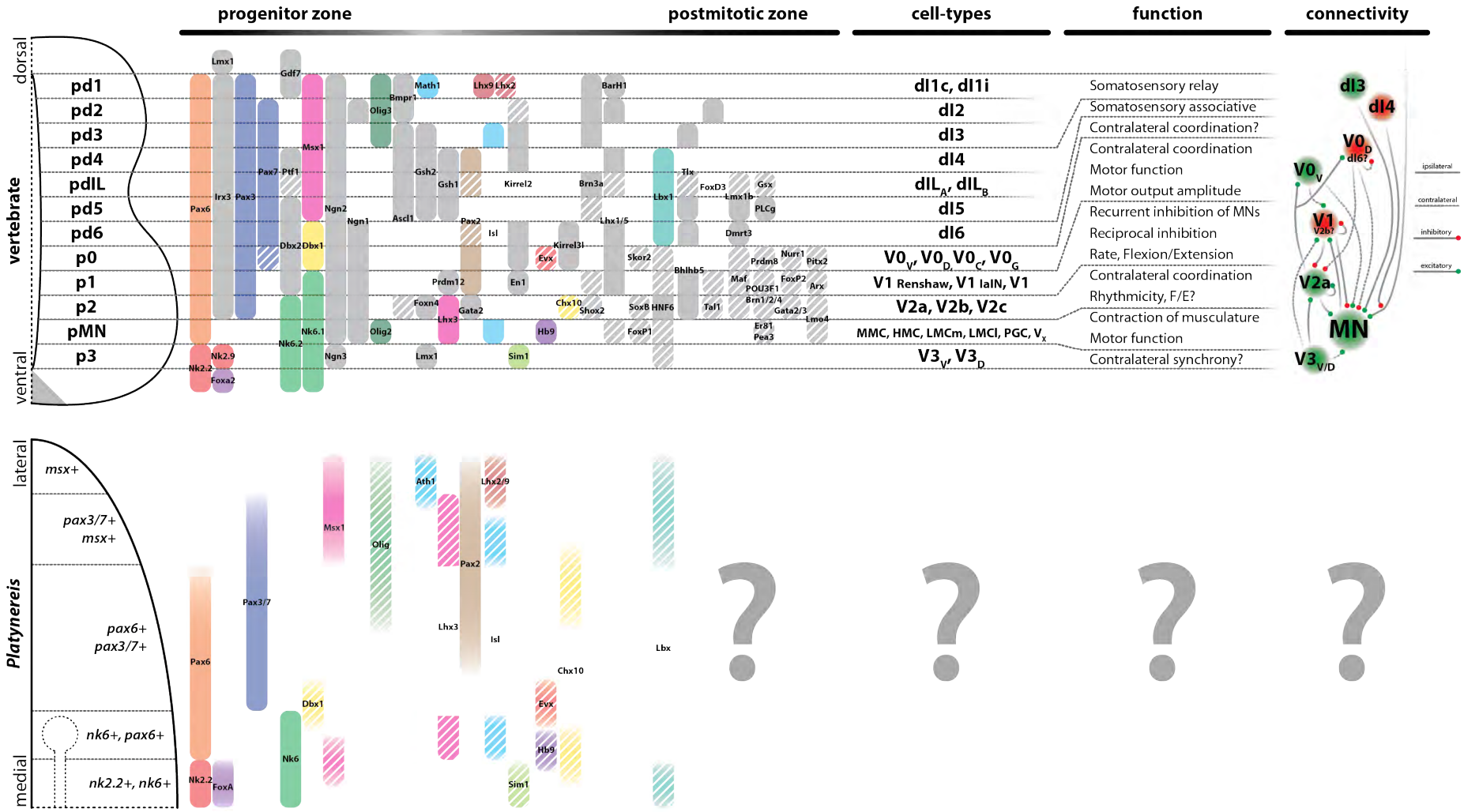

FIGURE 2.3: Comparison of gene patterning between the vertebrate spinal cord and *Platynereis* VNC.
Size of territories are not drawn to scale. Hatched bars in the vertebrate scheme indicate sub-population restricted expression. Hatched bars in the *Platynereis* scheme indicate that the gene is not continuous (along the anterio-posterior axis) in that area. Faded regions are used in the *Platynereis* scheme because I used the general anatomy of the animal to separate the domains and it can be unprecise. They are not faded when double *in situ* exists for those genes (e.g. Nk2.2 and Pax6). Connectivity diagram has been simplified and depicts the major inputs and outputs. [Sources mainly from Alaynick, Jessell, and Pfaff, 2011; Lu, Niu, and Alaynick, 2015; Denes et al., 2007; Kiehn, 2011; Denes, 2007, and complemented with various sources cited in this dissertation].

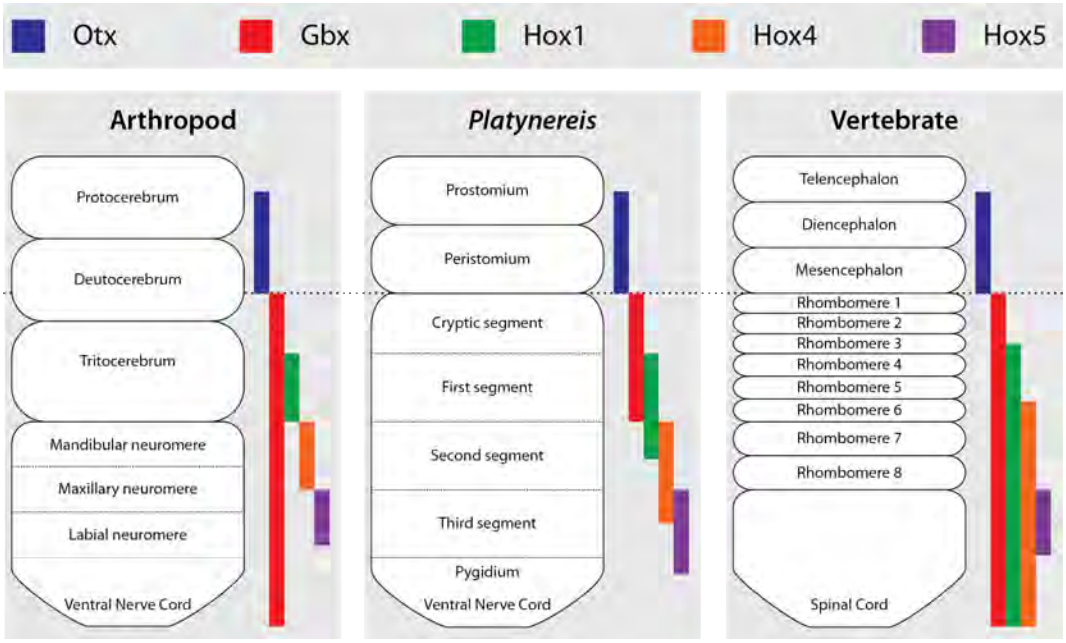

FIGURE 2.4: Conserved anterio-posterior patterning in bilaterians.
[Adapted from (Steinmetz et al., 2011; Reichert, 2005; Hirth et al., 2003)].

sections 4.2 and 4.4) to discuss the evolutionary relationships between the Protostomes and Deuterostomes BNS, in light of the main results found in this dissertation (see Results sections 3.1 and 3.4).

2.2.1 Vertebrate neurogenesis

Soon after gastrulation, the most dorsal portion of the neuroepithelium undergoes invagination, starting a process called neurulation. The folding of this tissue generates an anterior-posterior cylinder of epithelia that will give rise to the neural tube of the animal. The tube contains an apical or inner side that surrounds a central canal, and a basal or outer side. Analogous to the situation in Platynereis (ventro-dorsally differentiation zones), cell proliferation follows an inner to outer direction. Newborn neurons are first in contact with the central canal, and migrate to the basal part of the neural tube, expanding the layers of cells (Wentworth, 1984a; Wentworth, 1984b). Coupled to this cell proliferation stage, the genetic determination of neuronal types is shaped by their exposure to morphogens, which depends on their relative three-dimensional location in the neural tube, and is further specified by interaction with neighbouring territories.

The main source of cell type divergence in the neural tube occurs in the ventro-dorsal axis and is initiated by the antagonistic morphogens Shh and different Bmps and Wnts. Shh is initially produced in the notochord, an axial tissue derived from mesoderm that is located ventrally to the neural tube. This signal induces the most

ventral cells of the tube to become what is called the floor plate, that also starts expressing Shh. Bmps and Fgf signals are secreted from the opposite end of the tube, the most dorsally located cells, referred to as roof plate. Fgf signaling and retinoic acid concentration along the anterior-posterior (AP) axis, expressed from the neighbouring paraxial mesoderm, shape the specification and proliferative potential of these territories along the neural tube. Given that the embryonic program is not identical along the AP axis, the proportion of cell types, connections, and mode of cell specification can vary (Gouti, Metzis, and Briscoe, 2015).

In total, 14 domains can be identified along the DV axis (see Figure 2.3) based on a specific combination of Neural Progenitor Transcription Factors (NP-TFs) as a result of developmental cascades initiated by the above-mentioned morphogens and maintained by gene regulatory networks. In order to ensure proper boundaries in these progenitor domains, neighbouring territories are often defined by cross-repressive TFs (e.g. Nk6.2 and Dbx1) (Briscoe et al., 2000). Furthermore, the combination of NP-TFs that specify each territory ensures the blocking of the developmental program of each of the other territories (Kutejova et al., 2016). In addition to the domain where cells are born, migration has a big effect on the further functional properties of the neurons, as the dendrites will receive inputs from the neuropil nearby. This whole process is tightly regulated with proliferation and tissue growth to properly account for the scaling of the tissue (Kicheva and Briscoe, 2015). It is important to point out that these major subdivisions (14) of the neural tube are very general and based mainly on developmental specification. A recent study has reported more than 50 different cell types in just one of these domains (Bikoff et al., 2016).

2.2.2 Cell type specification in the vertebrate spinal cord

The main cell types composing the vertebrate locomotor CPGs are located in the ventral part of the spinal cord. From ventral to dorsal, these include 5 different classes: V3, MN, V2, V1 and V0 (see Figure 2.3). I will summarize what is known about their genetic specification, morphology, connectivity and functionality in different vertebrates. Figure 2.5 provides an illustration on how they are assembled into functional circuits for different locomotor adaptations. I will mention, for each population in the mice spinal cord, possible homologous cell types across different species, mainly lamprey, *Xenopus* and zebrafish (for a reference on the comparison between mice and zebrafish cell types, see Figure 2.6).

V3 population

V3 neurons are the ones located adjacently to the floor plate and represent the second major class of glutamatergic interneurons. 85% are commissural and project

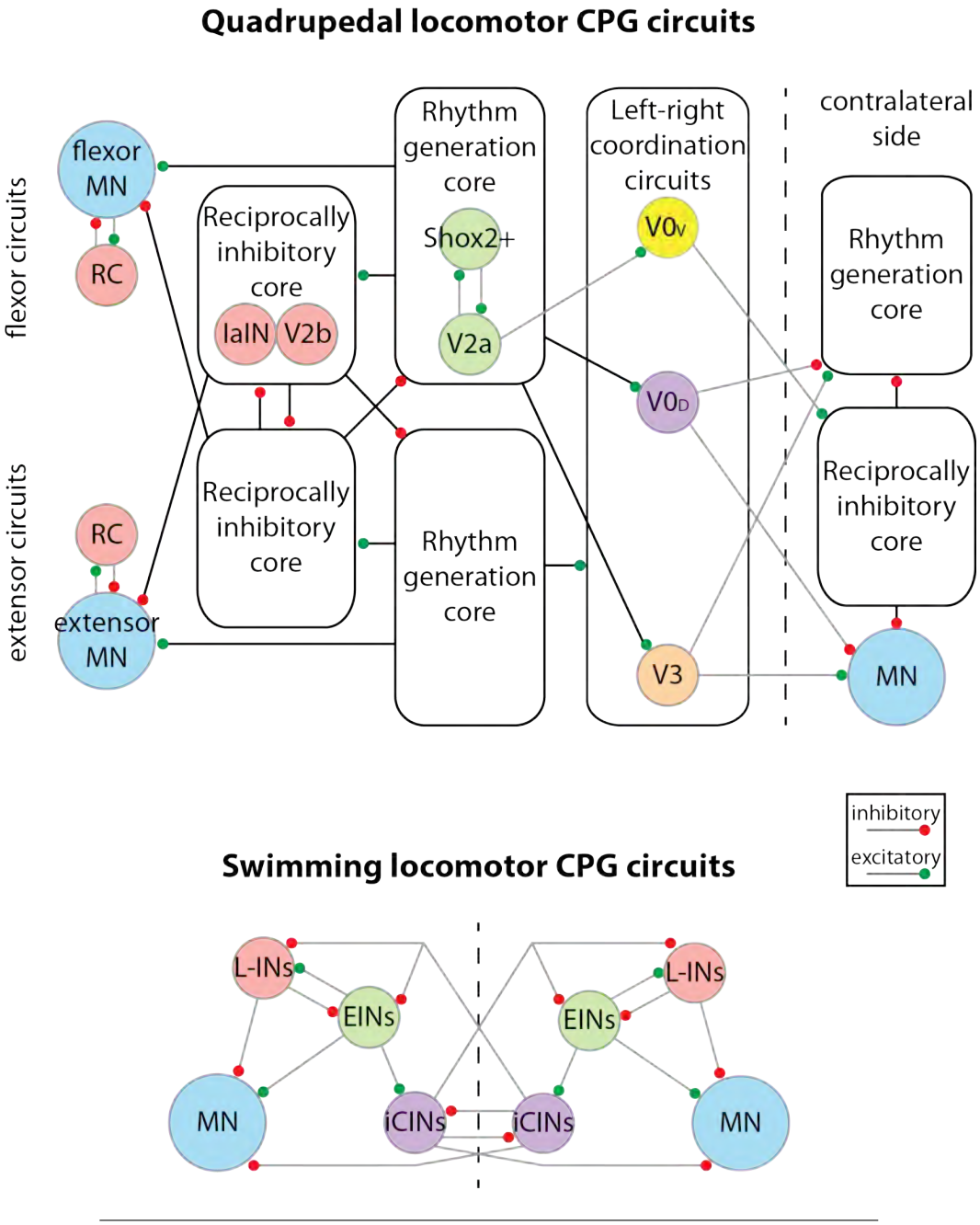

FIGURE 2.5: Comparison of locomotor CPGs in vertebrates.
[Adapted from (Goulding, 2009; Grillner and Jessell, 2009; Rybak, Dougherty, and Shevtsova, 2015; Dougherty et al., 2013; Kiehn, 2011)].

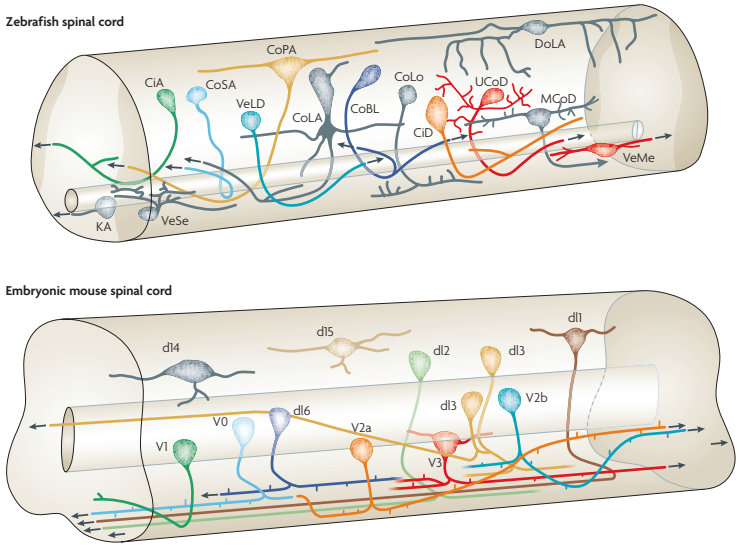

FIGURE 2.6: Comparison between the morphology and position of spinal cord cell types of zebrafish and mice.
[Reproduced with permission from (Goulding, 2009)].

caudally for more than 2 segments (Goulding, 2009). They form synapses on contralateral motorneurons (MNs), Renshaw Cells (RCs) and V1 IaINs (Zhang et al., 2014). 15% are ipsilateral and project to the V2 population (Garcia-Campmany, Stam, and Goulding, 2010). Ventral Medial (VeMe) and Unipolar Commissural Descending (UCoD) interneurons in zebrafish are homologues of V3 by location, projection and TF expression. They are defined by the expression of *sim1* (see Figure 2.3), arise from the Nk2.2+/Nk6+ column, and are excitatory. V3 specific ablation does not abolish the locomotion rhythm, and these cells are dispensable for generating proper left-right alternation, although this might be due to compensation by other cell types. They have been shown to be involved in stabilizing and balancing the rhythm by controlling burst robustness (Garcia-Campmany, Stam, and Goulding, 2010). Nevertheless, they are important for the production of a symmetrical left-right motor output, which indicates that they couple different oscillators in both sides of the body.

MN population

MN stands for motorneurons, and these cells are homologuized between different vertebrates and invertebrates by their definition (synapses onto muscles). There are many different classes of MNs based on specific musculature innervation and their specification by a correlated TF program, but they all arise from the same developmental domain (Olig2+/Nk6+/Lhx3+/Ngn+/Hb9+/Isl+) and all are cholinergic (excitatory). The sub-specification of different MN types has been tracked through

development and evolution in correlation to the appearance of new muscles and new ways of locomotion (Dasen et al., 2008).

V2 population

The V2 developmental domain (p2) is located dorsally to the MN one, and it has been identified by the specific expression of *gata2*. It comprises an ipsilateral population with a caudal projection of 2 segments (Goulding, 2009). Two major subpopulations arise from the p2 (Pax6+/Irx3+/Nk6+/Lhx3+/Pax3+/Ngn+/Foxn4+) domain: V2a and V2b (another minor population termed V2c has also been described but not much is known). V2a and V2b split in a terminal cell division that produces one excitatory cell (V2a) and one inhibitory cell (V2b) (Kimura, Satou, and Higashijima, 2008), in a process directed by Notch signalling (Batista, Jacobstein, and Lewis, 2008).

V2a are one of the two major classes of glutamatergic interneurons. They are homologous to excitatory interneurons (EINs) in lamprey, descending excitatory interneurons (dIN) in *Xenopus* and circumferencial descending INs (CiD) in zebrafish (Goulding, 2009). The specific marker for the V2a population is Chx10. V2a population connect monosynaptically to virtually all the other CPG components: to V0d and V0v (Garcia-Campmany, Stam, and Goulding, 2010), to MNs and to themselves (Kiehn, 2011). The scenario is similar in lamprey, zebrafish and tadpole (Fetcho and McLean, 2010; Garcia-Campmany, Stam, and Goulding, 2010). In the tadpole, dINs are electrically coupled to stabilize and make more reliable the rhythm during swimming (Roberts, Li, and Soffe, 2010). *Xenopus* dINs are the major source of ipsilateral rhythmic excitation via glutamate release. They co-release glutamate and acetylcholine which activates AMPA and nicotinic acetylcholine receptors (Roberts, Li, and Soffe, 2010). dINs fire long-duration potentials during swimming to drive head to tail firing of all other neurons during swimming. They have conditional pacemaker properties contributing to rhythm generation (like EINs in lamprey), and electrical coupling that leads to tightly synchronous firing, activating the NMDA receptors of the whole population (Roberts, Li, and Soffe, 2012). They fire a single long-duration action potential per cycle, and show post-inhibitory rebound firing to negatively augment pulses and inhibitory postsynaptic potentials (IPSPs) when they are depolarized, but rarely when they are at rest (Roberts, Li, and Soffe, 2010).

Glutamatergic Chx10+ cells have also been found to be the rhythmic CPG source in zebrafish (Ljunggren et al., 2014), providing on-cycle excitation to the CPG (Goulding, 2009). Furthermore, these cells (CiD) show a progressive recruitment, together with MNs, in a ventral to dorsal direction upon increase in speed (Fetcho and McLean, 2010). In tetrapods, the source of rhythm in the CPG has been harder to

disentangle genetically, although light-stimulation of glutamatergic neurons generates rhythm in mice spinal cords (Hägglund et al., 2013). Another neuronal population characterized by the expression of Shox2, which partially overlaps with V2a, has been found to be involved in the rhythm-generation circuitry (Dougherty et al., 2013). Ablation of Chx10 neurons in mice causes an uncoordinated alteration of the left and right limbs at high locomotor gaits, and the mutants show a hopping phenotype (Fetcho and McLean, 2010), supporting a differential recruitment of this population like in fish, as V2a also fire at low speeds (Kiehn, 2011). In these mutants, electrical stimulation of brainstem or Dorsal Root Ganglia (DRG) results in asynchronous and uncoordinated activity, while application of NMDA and 5-HT initiated normal motor bursting (Crone et al., 2008), showing that V2a might gate reticulospinal signals to mammalian CPG. This is consistent with their possible evolutionary relationship of reticulospinal neurons in the hindbrain (Kimura et al., 2013).

V2b comprises an inhibitory population molecularly homologous to Ventral Lateral Descending (VeLD) in zebrafish and ipsilaterally inhibitory INs (L-INs). They project to MNs and other members of the CPG (Zhang et al., 2014). By position, V2b are interspersed with part of the V1 population, but V2b project caudally and V1 project rostrally (Goulding, 2009). Ablation of both V1 and V2b cells in mice, affects the normal alteration firing of between flexor and extensor limb muscles (Zhang et al., 2014), suggesting that V2b might contribute to reciprocal inhibitory pathways forming part of a relatively redundant circuitry with at least part of the V1 population.

V1 population

V1 interneuron population is entirely inhibitory, ipsilateral and rostrally projecting (1-3 segments), and these cells are specifically defined by the expression of the TF *engrailed* (En1). Two major subpopulations, named Renshaw INs (RC) and IaINs, have been characterized, and comprise 25-30% of the total V1 population (Goulding, 2009). Nevertheless, a recent study has described more than 50 molecularly distinct cell types within the V1 population (Bikoff et al., 2016) in mice alone, although it is believed that in swimming animals this population is not as diverse, and that it has been greatly expanded in tetrapods alongside a more complex locomotion mode (Fetcho and McLean, 2010), given that in zebrafish the En1+ population seems more homogeneous. Homologous cell types in other animals include the L-INs in lamprey, Circunferencial Ascending (CiA) neurons in zebrafish, ascending recurrent inhibitory INs (aINs) in *Xenopus*, and they have molecular counterparts in chick as well. RC and IaINs connect monosynaptically to MNs (so do En1+ cells in zebrafish, tadpole and chick) and other interneurons like themselves

and the V2a (Fetcho and McLean, 2010; Goulding, 2009). In the tadpole, aINs connect to all the components of the CPG (Roberts, Li, and Soffe, 2012), and these cells also have a descending branch. RC receive excitation from the MN they innervate and inhibit these MN recurrently.

The V1 population develops from the Pax6+/Irx3+/Dbx2+/Nk6.2+/Pax3+/Ngn+ column, directly dorsal to the p2 domain. Their development is dependent on Pax6 and retinoic acid signalling (Kiehn, 2006). They are also characterized by the expression of *Prdm12*, which represses and is repressed by Dbx1 and Nk6 (Thelie et al., 2015). RC and IaINs are inhibitory and express GABA and glycine, RC express Calbindin (Garcia-Campmany, Stam, and Goulding, 2010) and nicotinic cholinergic receptor alpha 2 (Chrna2) (Kiehn, 2011), whereas IaINs express Parvalbumin (Garcia-Campmany, Stam, and Goulding, 2010). aINs are glycinergic and immunoreactive to GABA (Roberts, Li, and Soffe, 2012).

V1 cells are rhythmically active during tetrapodal locomotion. A subpopulation of IaINs show reciprocal inhibition (Kiehn, 2011) and are involved in the coordination of the motorneurons that control the flexor and extensor muscles so they alternate in their activity. IaINs receiving the same rhythmic excitation as the flexor circuitry project to the extensor MN, inhibiting their activity and therefore ensuring that the pool of muscles they innervate is not active when the flexors contract. The same scenario takes place in the extensor circuit. Therefore, IaINs fire out of phase with the MN population they innervate. In contrast, RC show in-phase firing with their innervated MN pool, as they are excited by the same MN they provide inhibition to, contributing to firing termination. Supporting the functional relevance of this circuitry, En1-/- mice show a reduced speed (slower rhythm / increased step cycle), and less abrupt termination of motor bursting. However, rhythmic inhibition to MNs and flexor-extensor coordination (even in the hemicord) is still present in these mice, meaning that V1 might be required only for fast movements (Goulding, 2009). It is likely that another cell population has an overlapping functional role, possibly V2b (see above).

These cells have also been investigated in swimming vertebrates. aINs and CiA are not responsible for the midcycle inhibition underlying left-right alternation (Kiehn, 2006), and *Xenopus* aINs are differentially recruited depending on the speed, being more active during faster movements, where fast inhibition might be functionally more relevant. During swimming, aINs inhibit rhythmically the active premotor INs and the MNs to limit the firing (Roberts, Li, and Soffe, 2012). CiA and aINs are candidates for the longitudinal inhibitory pathways that regulate the intersegmental delay and the swim cycle period. It is possible that part of V1 and V2b populations and circuitry were co-opted upstream of the fin adductor/abductor muscles, and that these circuits were elaborated into the control of the flexor/extensor musculature that permitted terrestrial locomotion (Zhang et al., 2014).

V0 population

V0 interneurons are specified in the p0 domain (Pax6+/Irx3+/Dbx1/2+/Pax3+/Ngn+), which is positioned in the most dorsal portion of the ventral part of the spinal cord (see Figure 2.3). The development of V0 is controlled by Dbx1, which is a specific marker of this population. Four subpopulations have been characterized within this domain (V0d, V0v, V0c and V0g), although two clear functional distinctions can be made between the inhibitory (VOd) and the excitatory (V0v, V0c and V0g) populations. They are mainly commissural and their main role is to control the activity of the motorneurons in the contralateral side of the body.

V0d compose two thirds of the V0 cells, and are homologous to inhibitory Commissural INs (iCINs) in lamprey and *Xenopus*, and Commissural Bifurcating Longitudinal (CoBL) and inhibitory commissural secondary ascending (CoSA) in zebrafish (Goulding, 2009). This population makes monosynaptic connections to contralateral MNs (Kiehn, 2006) and receive excitatory monosynaptic connections from ipsilateral V2a. Lamprey and *Xenopus* iCINs connect, in the other side, to excitatory INs, local inhibitory INs, commissural INs, and MNs. From lamprey to mammals, CINs are reciprocally connected along every segment of the spinal cord (Kiehn, 2006). *Xenopus* CINs have an unipolar somata in mid to dorsal positions, dendrites mid to ventral from initial ventral axon which crosses to ascend or branch (Roberts, Li, and Soffe, 2012). CINs are a highly variable population in terms of axonal trajectory after crossing, and activity pattern. V0d is a purely inhibitory population, expressing GABA and glycine. Zebrafish CoBL and *Xenopus* CINs are glycinergic (Goulding, 2009; Roberts, Li, and Soffe, 2012). V0d are very similar in their molecular profile, morphology, connections and functionality to the dI6 population, located adjacently dorsal in the spinal cord. V0d, as dI6, coexpress Pax7 mitotically, and Pax2 after differentiation, whereas V0v do not (Pierani et al., 2001; Talpalar et al., 2013).

V0d are directly involved in the control of left-right alternation (Kiehn, 2006). Dbx1-/- locomotor phenotype is characterized by periods of left-right synchrony (hopping movement), but normal alternation periods are also observed (Lanuza et al., 2004), suggesting that there might be redundancy of other differentially genetically-defined population (possibly dI6). Inhibitory CINs (lamprey and tadpole) are proposed to control left-right alternation through the inhibition of the contralateral rhythmic core (Roberts, Li, and Soffe, 2010; Kiehn, 2006), and the knock out of dbx1a/1b in zebrafish produces similar phenotypes as in mice.

V0v compose one third of the V0 population (V0c and V0g are derived from V0v and constitute around 5% of total V0). V0v have been compared to excitatory CINs (eCINs) in lamprey and to Multipolar Commissural Descending (MCoD), UCoD and excitatory CoSA neurons in zebrafish (Juárez-Morales et al., 2016; Goulding, 2009). No excitatory CIN (located in a position similar to V0) has been found in

Xenopus, but this could be because they arise in later embryonic stages (eCINs are more relevant for slow movements, and fast movements are the first to arise). V0d are mainly commissural, and V0c and V0g can also project ipsilaterally (Kiehn, 2011). Similar to V0d, these populations make monosynaptic connections to contralateral MNs and other INs and receive excitation from ipsilateral V2a. V0v is an excitatory population expressing glutamate. They are selectively controlled by Evx1/2, which differentiates them from the V0d cells, producing an excitatory phenotype at the expense of an inhibitory one (Juárez-Morales et al., 2016). MCoD are Evx1/2 positive and excitatory (Goulding, 2009; Fetcho and McLean, 2010). Mice lacking V0v do now show general left-right alternation defects (Lanuza et al., 2004), although this might be caused by the redundancy of direct and indirect inhibition of motorneurons (Kiehn, 2011). Nevertheless, V0v cells have been shown to affect the alteration of hindlimbs at higher speeds (Lanuza et al., 2004). V0c are important for sensory feedback and use acetylcholine as neurotransmiter, and are believed to be the exclusive source of the large cholinergic C-boutons that surround the MNs (Kiehn, 2011). MCoD are preferentially recruited during slow movements and dispensable for fast swimming left-right alternation (Goulding, 2009; Fetcho and McLean, 2010). They show rhythmic activation during slow swimming, that correlates with the slow movements of the pectoral fins, which are pressed against the body during fast swimming locomotion (Kiehn, 2011). The function of eCINs in lamprey is not known (Goulding, 2009).

2.3 Molecular atlases to profile cell types

The comparison of cell types across the animal kingdom is a promising strategy to gain a deeper understanding of metazoan evolution. However, the current lack of data on cell type molecular characterisation for non-model organisms restricts and biases these comparisons. Gene expression atlases represent a powerful tool for a thorough and systematic molecular description of animal body plans, even at cellular resolution, but the implementation of this methodology is limited to highly stereotypic developmental states (Tomer et al., 2010; Asadulina et al., 2012), or relies on species-specific characters (Heckscher et al., 2014).

2.3.1 PrImR in *Platynereis*

Platynereis is a powerful model organism for medium and high throughput systematic studies due to its highly stereotypic developmental period. This high temporal and structural similarity between the body plans of different individuals within and across batches has been exploited to build a single-cell resolution atlas of the trochophore brain (Tomer et al., 2010), and near cell-resolution atlases for older

stages (Asadulina et al., 2012). Initially, Tomer and colleagues used three dimensional image registration algorithms to align individual epispheres (brains) to a common reference using the nerve tracks structure to guide the transformation. Using just a handful of samples for each gene of interest they were able to achieve single-cell resolution for this highly-stereotipical organ. They called the method **PrImR**, which stands for Profiling by Image Registration.

However, the applicability of this method is limited, as animals need to be manually oriented in the microscope before imaging, the method does not cover the entire body plan, and the need to incorporate an antibody to the protocol to label the nerves makes the experimental workflow more difficult. Asadulina and colleagues introduced two modifications to this pipeline that greatly improved the applicability of the method. First, they changed the mounting medium from glycerol to 2,2'-thiodiethanol (TDE), making the animals transparent enough to be imaged completely (through the entire body plan) and in any direction. Second, they used the nuclear staining DAPI (4',6-diamidino-2-phenylindole) as a reference marker for the registration, which is easier to implement and represents a higher applicability to other species. This new protocol allowed them to create full-body atlases for 48hpf and 72hpf animals with near cell resolution. As the base of the approach is still the same, I will also refer to this method as PrImR.

PrImR is therefore a technique used to catalogue gene expression patterns in a tissue or an animal by mapping individual expression signals to a common three-dimensional reference. This allows for the comparison, *in silico*, of different gene expression patterns, and the measurement of gene co-expression levels at the resolution of tissues or even individual cells.

2.3.2 PrImR workflow

The main steps of PrImR principle are represented in Figure 2.7. I will base my descriptions on the resource built in Tomer et al., 2010, but the principles employed in Asadulina et al., 2012 is overall the same. As a first step, the nerve tract architecture of several individuals, stained using an antibody against acetylated tubulin (AcTub), is used to generate an average model called "reference", that is going to be used to register all individuals to the same three-dimensional space. In order to generate the reference, an iterative registration protocol is used. First, all the animals are aligned to a single individual, and an average is built from all the registered samples. Next, raw images are aligned again, this time to the previous average. This process is repeated with the updated averages until no significant change can be observed (Figure 2.7a).

In order to register individual animals to build each gene expression map, a three-dimensional transformation is calculated to optimize the overlap with the reference using the AcTub channel of the sample. First, a rigid transformation is applied to

the image, where just translation, rotation and scaling are applied. This is followed by affine and deformable registration, where parts of the image field can be deformed independently to fit the reference. The algorithm calculates the mutual information between the transformed image and the reference to optimize their similarity. The total transformation calculated for the AcTub channel is then applied to the signal channel, so this one is also aligned three-dimensionally relative to the reference (Figure 2.7b). For each genetic marker, the signal channels from 2 to 5 samples are normalized and averaged to build a canonical gene expression pattern. The registration of many markers to the same reference creates a gene expression atlas that can be used for *in silico* co-expression analysis of hundreds of genes, allowing a molecular categorization of cell types (Figure 2.7c).

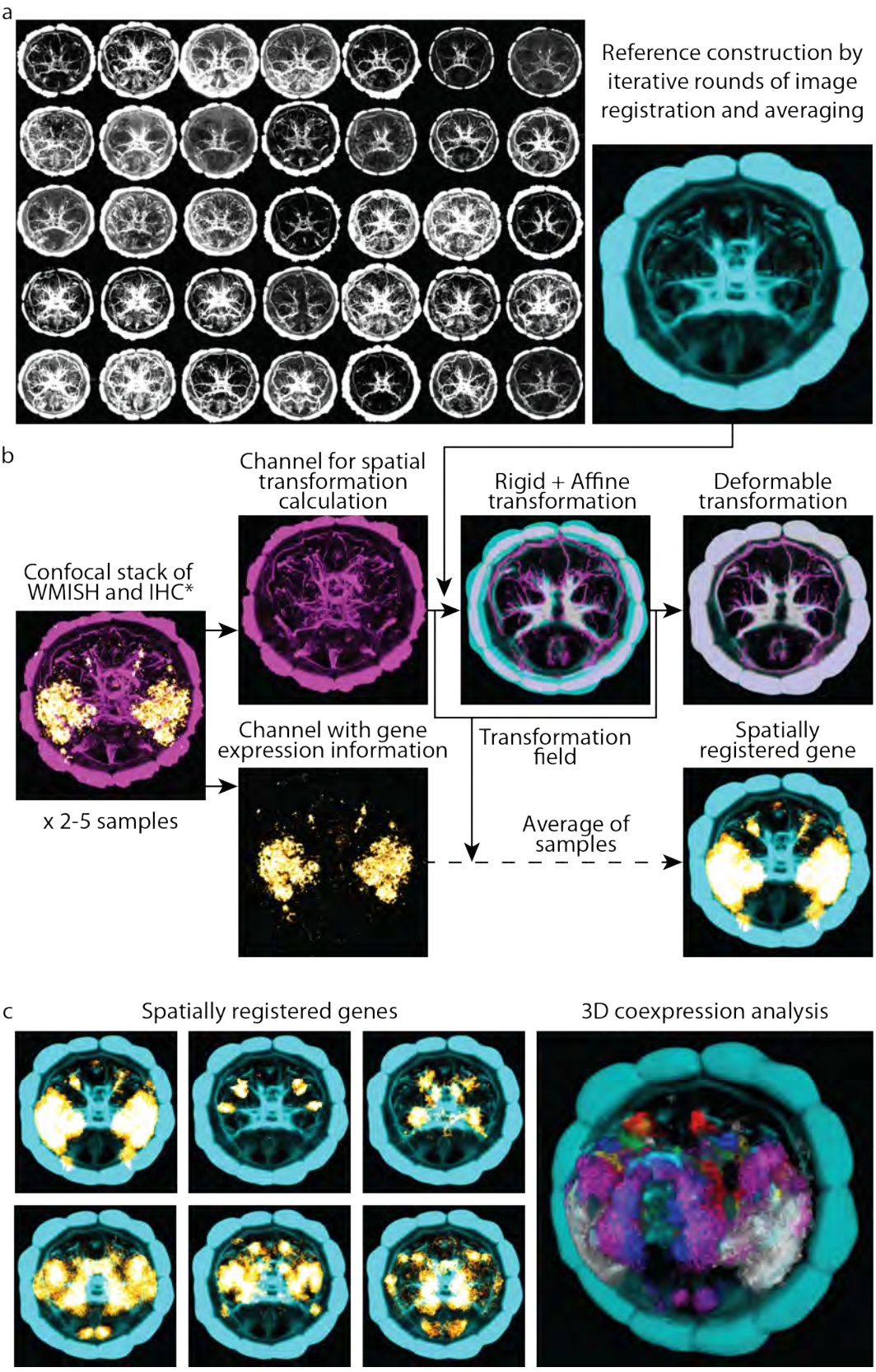

FIGURE 2.7: Workflow of PrImR.
(a) On the left, maximum projection panel of individual brains stained with AcTub, used to construct an average (right) by iterative registration. **(b)** Illustration of the samples 3D registration process. Description in the main text. **(c)** Co-expression of two or more genes can be analyzed *in silico*. WMISH, Whole mount *in situ* hybridization; IHC, Immuno histochemistry. Panel from **a** and transformation images from **b** are reproduced with permission from Tomer et al., 2010.

Chapter 3

Results

In this chapter I will describe the main results I have obtained on different aspects of the project. The first two sections are related to descriptive approaches in *Platynereis* nechtochaete (6 days post fertilization) larvae, and the third one describes the functional results of a knock-down strategy to assess the role of a specific neuronal population. These sections are as listed:

- Generation of gene expression atlases. Development of computational tools to build gene expression atlases for full body animals.
- Molecular profiling. Exploration and mining of expression atlases and analysis of the molecular structure of *Platynereis* VNC at 6dpf.
- Locomotion monitorization. Development of software to track the larvae body posture and monitor and quantify several locomotion parameters.
- Functional analysis of commissural neurons. Crispr/Cas9 mediated knock-down of *dbx1* and analysis of the phenotype.

3.1 *In-toto* molecular profiling of *Platynereis* nechtochaete larvae

Pushing PrImR (see Section 2.3) to older and more complex stages in *Platynereis* development while having single-cell resolution is of special importance for two main reasons. First, the locomotion movements that I am interested in describing are more frequent and robust at 6dpf (see Section 3.3), indicating a complete differentiation of the neuronal players and established neuronal circuits. Second, the part of the nervous system controlling this locomotor behaviour, the ventral nerve cord (VNC), is almost fully composed of post-mitotic neurons (see Section 3.1.1) at this stage. This makes the VNC of the 6dpf *Platynereis* an ideal system where to apply the cell type model (see Section 1.2), as it will allow for the analysis and comparison of the differentiated neurons that arise from similar developmental territories in vertebrates and annelids (see Section 1.1). Additionally, it is important for

the method to be systematic, easily applied and as automatized as possible while allowing easy supervision and modification.

To this end, I have extended the previous methodology into an automatic pipeline for the generation of full-body gene expression atlases with single-cell resolution that I have called **ProSPr**: **Pro**filing by **S**ignal **P**robability mapping. I have developed a new image analysis routine, so that gene expression atlases can be built in a highly automated manner for more complex and/or less stereotypic developmental stages of non-model organisms. We tested this method on *Platynereis dumerilii* nectochaete larvae, and found that we can successfully generate a whole body gene expression atlas achieving cellular resolution. We have produced an automatic workflow to generate gene expression maps, and to further subdivide them into expression domains for which quantitative information can be extracted to generate cellular models of whole body plans.

In this section I will explain the process and research I undertook in order to generate ProSPr. Given that this research is aimed to explore and generate novel methods, both methodology and results will be discussed.

3.1.1 *Platynereis* VNC is composed of post-mitotic neurons at 6dpf

I monitored the development of the VNC in the annelid *Platynereis dumerilii* and analyzed the expression of neuronal differentiation markers and mitotic activity. I found that at 144hpf (equivalent to 6dpf), the VNC is mainly composed of post-mitotic neurons, as shown by the markers ElaV and Synaptotagmin (Figure 3.1a). EdU stainings over development show that the VNC proliferates mainly between 48hpf and 72hpf, with a decrease in proliferation over time, and just a few cells proliferating between 120 and 144 hpf (Figure 3.1b). Interestingly, this is coincidental with the stage at where *Platynereis* larvae undergo pelago-benthic transition and start feeding (see Section 2.1.3). Also at this stage, larvae are able to crawl in the substrate performing a periodic alternation of anterio-posterior movements of the parapodia (see Section 3.3). This movements are entirely controlled by neurons located in the VNC of the animal, as decapitated animals can still perform them indistinguishably from normal animals (observations by Dr. Paola Bertucci and Hernando Martinez). The VNC of *Platynereis* at 6dpf therefore represents a good system to study the similarity between the molecular profile of cell types that compose this structure with those in the neural tube of vertebrates.

3.1.2 Assessment of PrImR performance in *Platynereis* 6dpf larvae

Measuring the resolution achieved by PrImR is a necessary step in order to be able to interpret the results in a meaningful way. By PrImR resolution, I mean the possible deviation of homologous points between animals after they are registered to

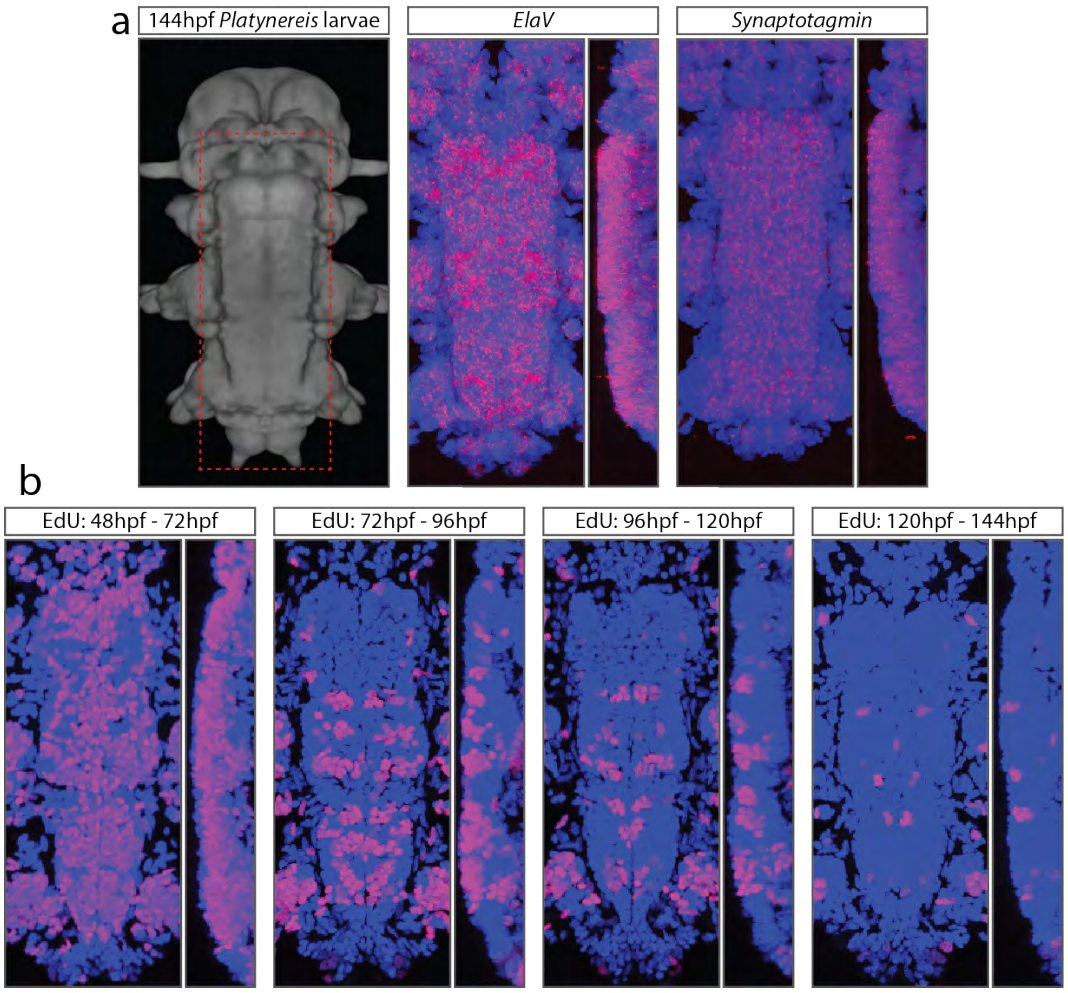

FIGURE 3.1: *Platynereis* VNC is composed of post-mitotic neurons at 6dpf.
Panels show the ventral and side projections of the ventral nerve cord (VNC, dashed red box) of *Platynereis* 144hpf (6dpf) larvae. **(a)** VNC location and RNA localization for post-mitotic pan-neuronal markers (ElaV and Synaptotagmin). **(b)** Cell age as determined by sequential 24 hour pulses of EdU, from 2 to 6 days post fertilization.

the reference. This resolution depends on various factors:

- Image Resolution. Microns per pixel on the original confocal images. In theory, the higher the resolution (the less microns per pixel), the more information one collects, and, if the registration algorithm performs fine on big data, the better the alignment would be. The higher the resolution the need of more imaging time, as well as more registration time or computational power.
- Registration Performance. The type of image registration method employed, and its characteristics and performance will affect the final resolution. Pixel-based, segmentation-based, and features to rely on (e.g. DAPI signal or axonal scaffold) are some of the options.
- Experimental deformations. Technical variability can be generated during the experimental procedure (e.g. *in situ* hybridisation technique), which can affect the registration performance.
- Animal stereotypicity. While technical errors can come from the three points mentioned above, we also need to consider biological variability. The developmental stereotypicity of the animal will be a factor that we cannot control fully (no more than keeping the environmental conditions stable), and that will impose a constraint on the final atlas resolution. By stereotypicity I mean how similar the individual animals are. Normally, this refers to the number of cells, the relative position of them on the body plan, and the identity (molecular profile in this case) of those cells.

Experimental design

Separating technical and biological variability is a hard task without a ground truth of biological variability. Nevertheless, an assessment of the general variability is important in order to know what type of information we can get from this analysis. In order to estimate PrImR resolution in *Platynereis* at 6dpf, one can select one cell that has a specific and unique expression on its surroundings, so it can be unequivocally identified in different animals. One can then calculate the position of that cell for many animals after these have been registered, and study the distribution of the signal of that specific gene in order to quantify the resolution and performance of the method.

For *Platynereis* at 6dpf, the expression of *GAD* provides an excellent example, as in each ganglion of the second segment, there is one cell specifically expressing *GAD*, so it can be easily identified across animals (see Figure 3.2a). As I aimed to set up the method for a new developmental stage, the same experiment was used to set up the technical parameters (such as imaging resolution needed). Results will be discussed in relation to both the final cellular resolution and the technical parameters. The steps to analyse PrImR resolution have been the following:

1. *GAD in situ* hybridisation was done for many animals at 6dpf.
2. Confocal stacks were taken at isotropic resolution of 334nm per pixel.
3. A reference was constructed using 36 animals and three iterations of affine and deformable registration.
4. 20 *GAD*-stained animals were registered to the reference.
5. The same *GAD* cell was identified and manually segmented for every registered animal, and the centroid position was calculated.
6. Original images were downsized to lower resolutions and steps 3 to 5 were repeated.

PrImR cannot achieve single-cell resolution at 6dpf

The experiment mentioned above resulted in a collection of the post-registration spatial position and volume of the same cell in 20 different samples. Moreover, this dataset was generated for five different pixel (imaging) resolutions (334nm , 440nm, 550nm, 660nm, 3000nm). In order to estimate the variability in the spatial distribution of the cell, the distances between the centroids of the cells was calculated for every imaging resolution (see Figure 3.2b'). The volume of the cell was estimated as the median value of the volumes calculated after manual segmentation using the higher resolution images (334nm). From this point, cells are considered as perfect spheres with a median volume, positioned in the calculated centroid for each one (see Figure 3.2c''). In order to achieve cellular resolution, the distances between the cells should be smaller than a cell radius, so one can always correlate cells across experiments based on proximity. As shown in Figure 3.2b', for every pixel resolution, the distances between the *GAD* cells are greater than a cell diameter, making it not possible for PrImR to generate an expression atlas with cellular resolution for *Platynereis* at 6dpf. This variation in spatial position is different depending on the axis, being bigger in the ventro-dorsal direction (Figure 3.2c'), possibly because this is the axis of cell proliferation (see Section 2.1.4).

3.1.3 Inference of cell distribution can be used to achieve single-cell resolution

Despite the lack of cellular resolution between individuals, I wondered which kind of distance distribution was obtained for that cell after registration. For that, I calculated the centroid of the cell distribution (centre of mass) as the median position between the 20 centroids, independently for every resolution. I then calculated the distances between each cell centroid and the centre of mass. As can be seen in Figure 3.2b'', and for every resolution, cells are located around a single centre of mass

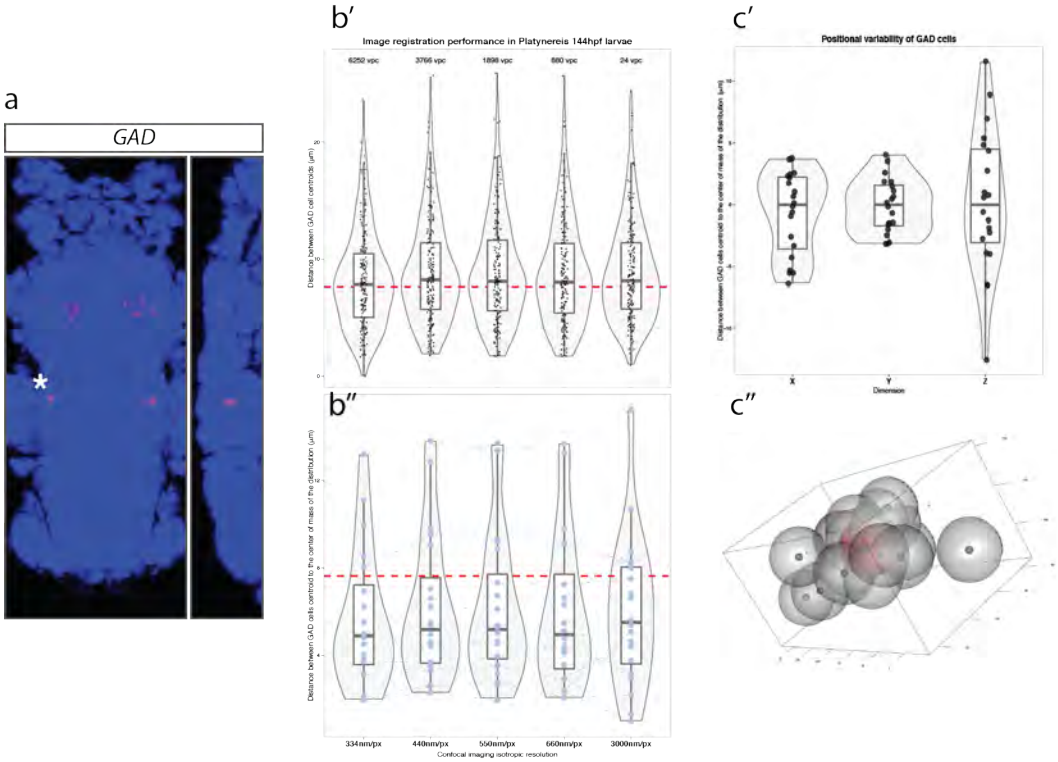

FIGURE 3.2: PrImR cannot achieve single-cell resolution at 6dpf.
(a) VNC location and RNA localization for glutamate decarboxylase (GAD). **(b)** Assessment of Image Registration of 144hpf Platynereis larvae for different imaging isotropic resolutions. GAD RNA localization was used to identify a single cell in the second ganglia of the VNC (asterisk in a) for 20 animals. For every pixel-resolution (x-axis), the center of the GAD cells was calculated. The dashed red horizontal line is the average diameter of a cell (calculated as the median of all the cells segmented at 334nm and assuming cells as spheres). The voxels per cell (vpc) are indicated on top, which show, for every imaging resolution, how many voxels (3D pixels) fit inside a cell. **(b′)** The individual points represent all the distances (20*19/2), in microns, between the centroids of the cells. Box plots and violin plots represent the distribution of these distances. **(b″)** Distance between the GAD cells centroids, and the center of mass of their positional distribution, calculated as the median position of the 20 cells. **(c′)** Variation of the position of the GAD cells in the three dimensions. **(c″)** 3D representation of the location of the GAD cells after the registration. They are represented as spheres with average cell diameter. In red, the centroid of the distribution.

forming a normal distribution with mean of approximately a cell radius (see also Figure 3.2c''). Even though the same cells don't register to the same three dimensional location, if we assume that each individual cell follows a unique distribution distinct to other cells, we could exploit the sample size to infer the centroid of that distribution and correlate between experiments.

For instance, let's assume we have one cell (named Joe), specifically expressing gene *A*, that is located in a morphologically distinct domain of the animal (e.g. one ganglion, which contains 20 cells). If the development of the animal is highly stereotypic, cell Joe will be located in the same relative position (to the other cells in the ganglion) in different animals, and, if the technical error of registration is zero, Joe will end up in the same three-dimensional position for all the animals registered to the reference. It will then be straight forward to assign any other gene to cell Joe. If we take an animal showing the expression of gene *B*, and when registered to the reference, the signal of gene *B* is present in the coordinates of Joe, and given that the position of cell Joe is invariant, we can conclude that genes *A* and *B* are co-expressed in cell Joe. This will require the registration of only one animal with the pattern for gene *B*. The opposite scenario would be that cell Joe is randomly distributed through the ganglion. Assuming again that the technical error of the registration is zero, cell Joe would end up in a different position for each animal registered. If we register several animals showing the expression of gene *A* (only showed by cell Joe in that ganglion), we could extract probabilities of expression for gene *A*, and we would get that it is 1/20th for every position in the ganglion.

In intermediate scenarios, however, with enough samples registered, one would be able to estimate the distribution of cell Joe, and to define a "centre of mass" for the expression, and an accurate "expression probability map". The more stereotypic the system is, the smaller the distribution of cell Joe will get, the less variability, and the less samples needed to get a probability map for the expression of gene *A*.

In order to estimate the number of samples needed to approach the centroid of the distribution, I calculated, for random subsets of cells, the distance of their centre of mass to the "true" centre of mass (calculated from all the samples). As seen in Figure 3.3, with just a few segmented samples, it is possible to accurately reconstruct the position of the cell and have cellular resolution. Surprisingly, the resolution at which the images are taken (in the range assayed here) does not make a difference to the registration of individuals (Figure 3.2b) and the inference of the centroid (Figure 3.3). In the case of the lowest resolution, each cell is still covered by 24 voxels, and this seems to be sufficient for the registration to work relatively fine, although the exact relative position of each cell does not correlate completely with the other resolutions (see Figure 3.2b'').

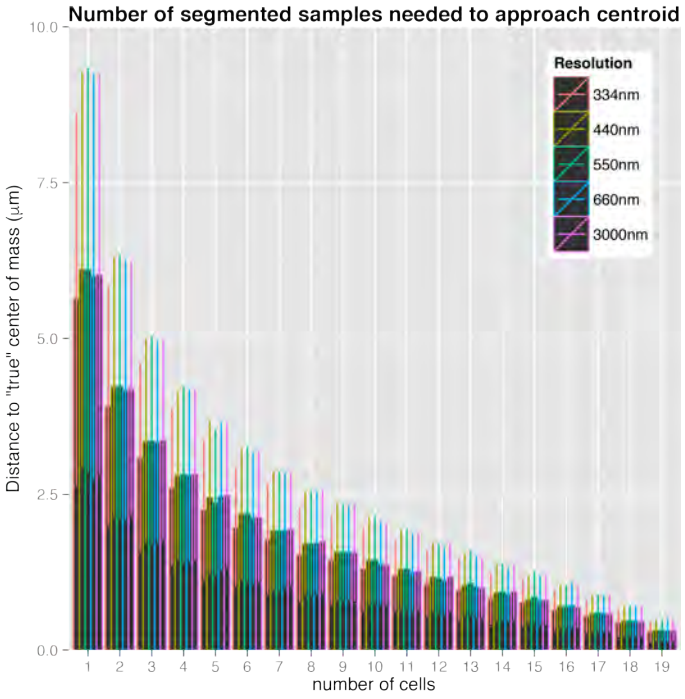

FIGURE 3.3: Few animals are needed to calculate cell distributions. Barplot showing, for each imaging resolution, the distribution of distances between the 'true centre of mass' of the *GAD* cell (obtained using 20 samples) and the centre of mass calculated using a determined number of random samples. 1000 repetitions for calculating each distribution.

Distributions can be calculated from Signal Probability Maps

Even though this is a promising strategy to achieve a cellular resolution expression atlas, manual segmentation of individual images is not practical, as it is highly time consuming and is not feasible for high-throughput approaches, so we have to rely on another strategy to retrieve the information. We need also to consider that *in situ* hybridisation signal is highly dependent on multiple parameters (concentration of probe, washes, temperature, acetylation step, concentration of antibody and NBT/BCIP, time of development, mounting, imaging conditions etc). Quantitative readouts of *in situ* signal are therefore very difficult to achieve in these experiments, and it is more practical to consider the result a qualitative (or binary) measure of expression. The investigator can then, using a convenient criteria (e.g.: educated decisions based on knowledge about the animal or automated methods), select a threshold for the individual images and binarize the registered data. With many samples for each experiment, one could overlap the samples to construct a Signal Probability Map (SPM), where each pixel will contain the information of the percentage of animals showing signal in its location. Figure 3.4 shows a schematic of the principle of SPMs.

Can we then "reconstruct" the cells (centroid and volume) from the SPMs? If we select, for instance, from the SPM of the 550nm/pixel resolution case (Figure 3.5), the signal that is present in at least 5 animals, we can calculate the centroid and volume of the pixels located in the position of the *GAD* cell. In this case, the recovered volume is 2058 voxels, similar to the size of a cell (see Figure 3.2b). The distance between the centroid of these pixels and the centroid calculated by segmentation is 1.49481 microns, with the radius of a cell (calculated as a median of segmented

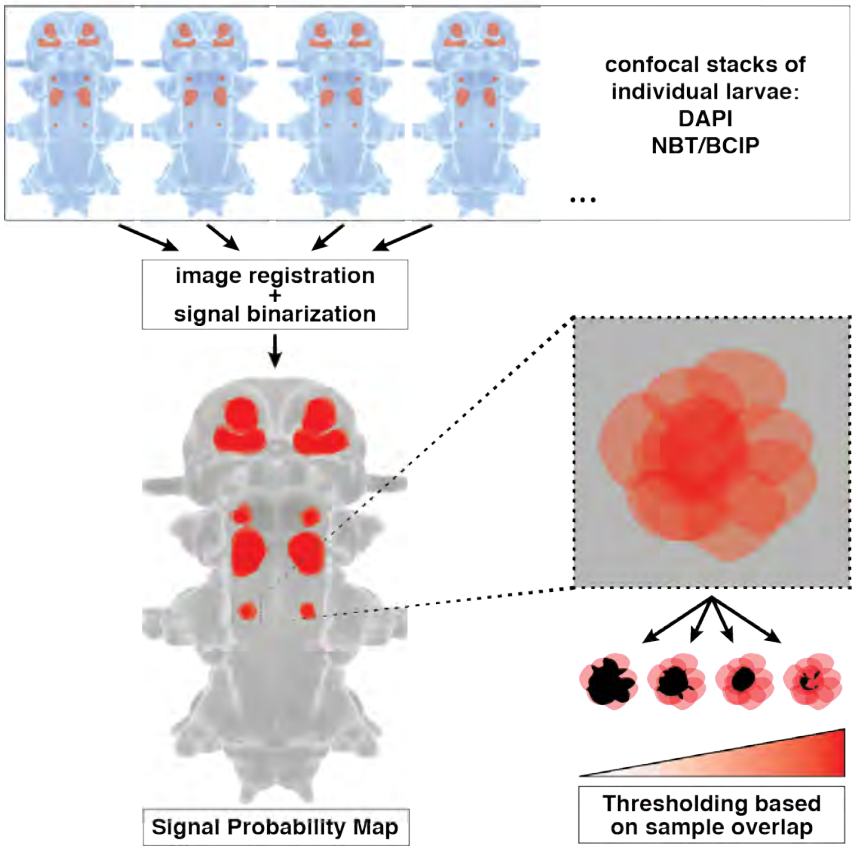

FIGURE 3.4: Generation of Signal Probability Maps.
Scheme representing the signal reconstruction process, generating Signal Probability Maps (SPM). Individual animals (samples) are registered to the 3D reference. The signal channel is binarized after registration, and the signal of different samples is overlapped to generate a probability map for the distribution of the signal (SPM). A threshold, dependent on the sample agreement, can be applied to the SPM to binarize the map.

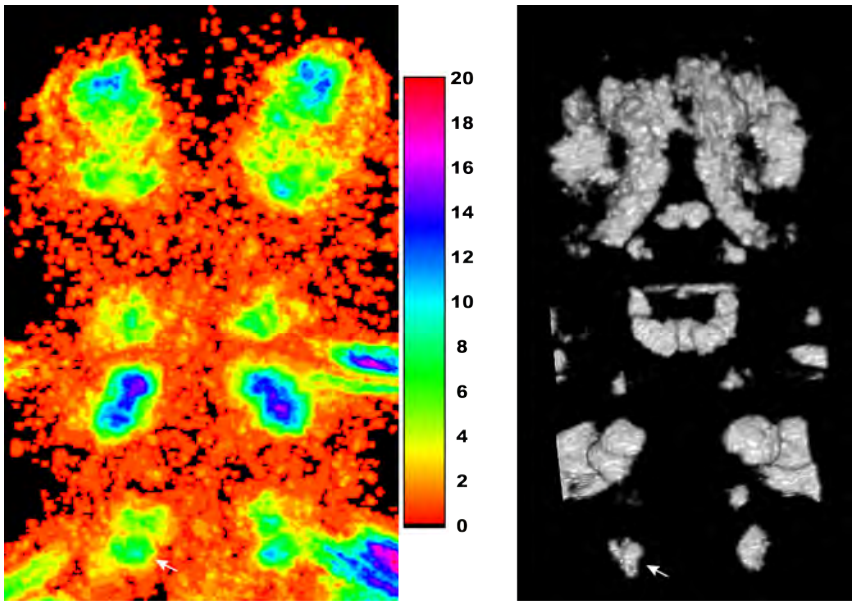

FIGURE 3.5: SPM for 550nm/pixel *GAD* WMISH.
On the left, maximum projection (of the most ventral part of the animal) of the Signal Probability Map for 20 animals imaged at 550nm/pixel (cropped to show the head and the two most anterior segments). The sample agreement is colour-coded with the gradient shown next the image. On the right, 3D Snapshot of the SPM thresholded to 5 animals agreement (25%).

GAD cells) being 3.81701 microns. This principle can therefore achieve single cell resolution for gene expression atlases.

In order to test the performance of this approach and the effects of different parameters in our example above, I used the registered binarized data to build, for each imaging resolution, a SPM in the area where the *GAD* cell is located. This procedure was done for different combinations of samples, varying the total number of samples and the binarization percentage. Figure 3.6 shows the result for the 550nm/pixel resolution case. As can be seen, both the number of samples and the binarization threshold have an impact on the result, although this is of different nature. Generally and without a surprise, the more samples that are included, the better the final cellular resolution. Very low threshold percentages (maintaining more signal) can achieve cellular resolution. As the threshold percentage is increased, the resulting volume changes dramatically (points move to the left in the graph), but the resolution tends to have an exponential decay, with little change at the beginning and the loss of resolution if the threshold is too high (see also how the sample agreement changes in the two dimensional representation of the SPM for that cell in Figure 3.5).

It is interesting to observe that the limit of this threshold inflexion point coincides with the real volume of the object (in this case the cell). This is very useful, as we can use at the same time both position and volume to approximate a canonical

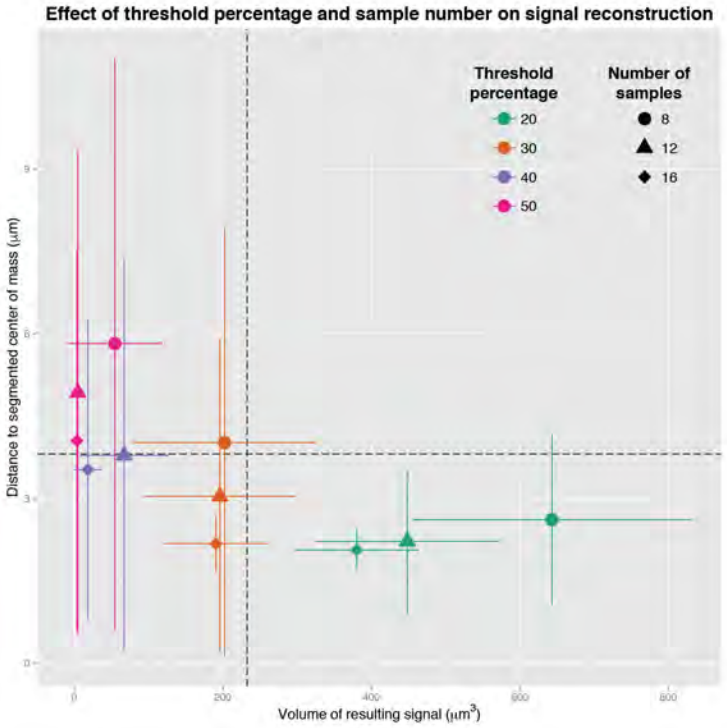

FIGURE 3.6: SPMs can render single-cell resolution.
Effect of the number of samples and the thresholding percentage on the performance of the signal reconstruction process, scored as the distance of the resulting 3D object centroid to the segmented center of mass of the total *GAD* cell distribution, and the volume of the 3D object. From the initial dataset of 20 manually segmented *GAD* cells, random picks of cells were made for every imaging resolution. Each point represents the distribution of centroid-center distance and volume (mean and standard deviation) of 1000 random iterations. Horizontal dotted line represents the average cell radius. Vertical dotted line represents the average cell volume. This figure corresponds to the data for 550nm/pixel imaging resolution.

representation in the atlas, and it will be exploited to calculate the threshold for the SPM (see Section 3.1.5).

We can therefore conclude that following the SPM approach, and using the correct parameters, it is possible to generate a cellular resolution gene expression atlas for *Platynereis* at 6dpf. I called this new method **Profiling by Signal Probability mapping (ProSPr)**.

3.1.4 Determination of experimental conditions and biases of the method

Imaging resolution has a limit

In order to determine the experimental conditions for the generation of the atlas using ProSPr, I extended this analysis to different imaging resolutions (Figure 3.7). In this case, it is evident that the lowest resolution (3000nm), despite behaving well with segmented measurements, is not a reliable source to be used for ProSPr, as it

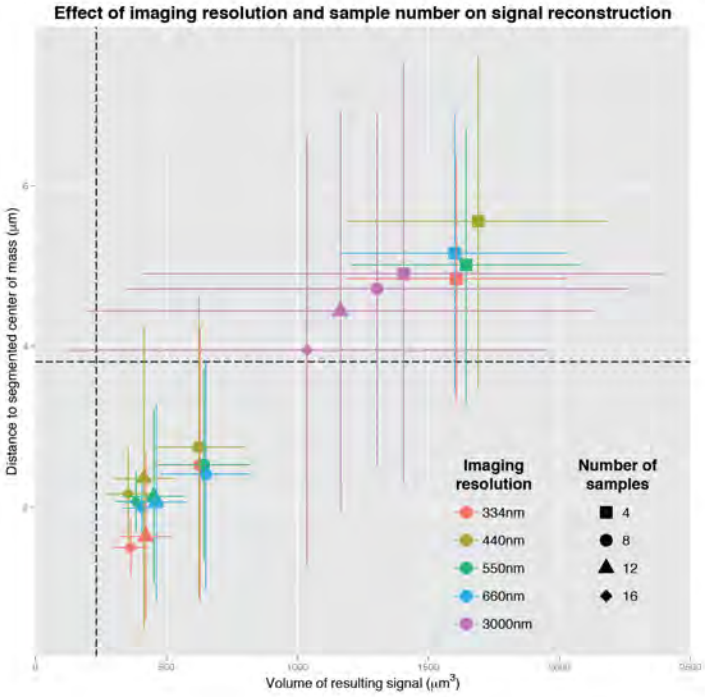

FIGURE 3.7: Resolution gain is insignificant above a certain threshold.
Effect of the imaging resolution and the number of samples on the performance of the signal reconstruction process. See Figure 3.6 for procedure and plot details. SPMs thresholded to 25%.

cannot achieve cellular resolution even when many samples are used. For the other cases, the main source of difference is the number of samples used and not the initial image resolution. With resolutions below 660nm we are oversampling the data, as no statistical difference is observed. As discussed before, a small sample size (in this case four) is not sufficient to generate cellular resolution. With eight samples the distance is recovered below the cell radius, but the volume is too big for it to be used for a cellular model (see previous section). The results do not vary much between 12 and 16 samples. To build the *Platynereis* 6pdf atlas using ProSPr, I decided to use images with isotropic resolution of 550nm per pixel as this ensures that every pixel is scanned twice in the confocal, and to include more than 12 samples for the generation of every gene SPM.

Cellular resolution measured with independent subsets

For the analysis above, the same samples are used to generate the "true" centroid, as well as the SPMs that were then analysed to calculate the distances to that centroid. In order to have a non-biased approach, we can divide the set of 20 samples in two subsets of 10 (and do this randomly 1000 times), and use one of the subsets to get the centroid from segmented data, and the other to recover the information from the SPM and calculate the distance to that centroid. Figure 3.8, shows that even with just a sample size of 10, cellular resolution can be achieved using different cohorts of cells.

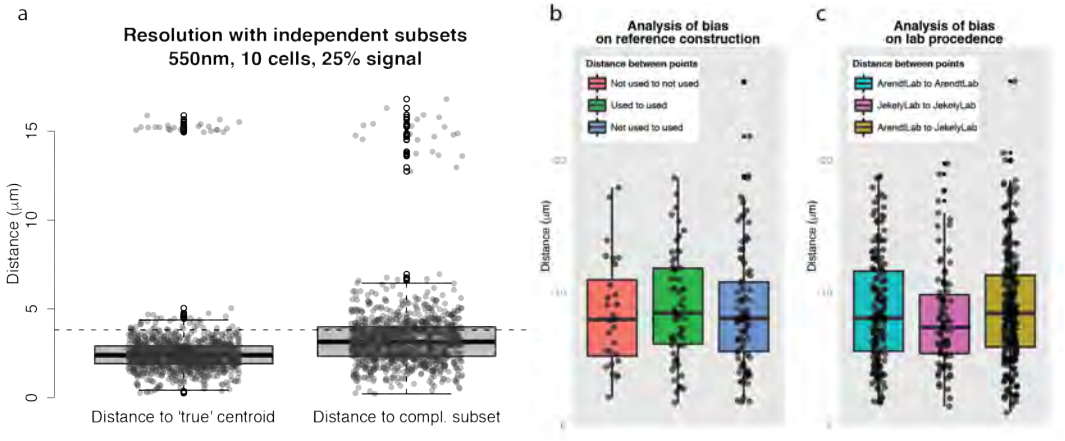

FIGURE 3.8: Analysis of SPM biases.
(a) Random sampling of 10 samples from the set of 20. For those 10 cells, the "signal probability map" was constructed, and for the signal present in at least 3 samples (25%), the centroid was calculated. The box plots show the distance from this centroid to the "true" centroid (computed from the 20 segmented samples), and the distance to the centroid computed from the complementary subset (also segmented data). Horizontal line shows the cells radius. 1000 repetitions. **(b)** Using the segmented-cell centroids, the distances between these 3-dimensional points was calculated for those belonging to the animals used to construct the reference, for those coming from animals not used for the reference, as well as the distances between the points of the two sets. **(c)** Similarly, point distances have been measured between the 20 centroids used for the analysis and 14 centroids calculated, using the same procedure, for samples from a different laboratory.

No biases detected for different animal subsets

Given that just a subset of the samples in the experiment described above were used to build the reference, we can also assess whether there is a bias on the registration results obtained for samples that were used to construct the reference, and samples that were not. Similarly, as this resource is meant to be used by the *Platynereis* community, it is interesting to see if the origin of the samples made a significant difference in the registration quality. Following the same principle, we can assess whether there is a bias or not on samples (animals) taken from other laboratories. I analysed 14 samples from a different lab as described above in order to compare the results. For that, I want to acknowledge the Jékely lab for sending some of their 6dpf animals for testing. Figure 3.8 shows that we did not find a bias either on the animals used for building the reference or in the origin of the samples. It is worth mentioning that in the case of lab origin, the *in situ* experiment was performed side by side using the same protocol from the same lab, and the difference that can arise from different protocols is not assessed here.

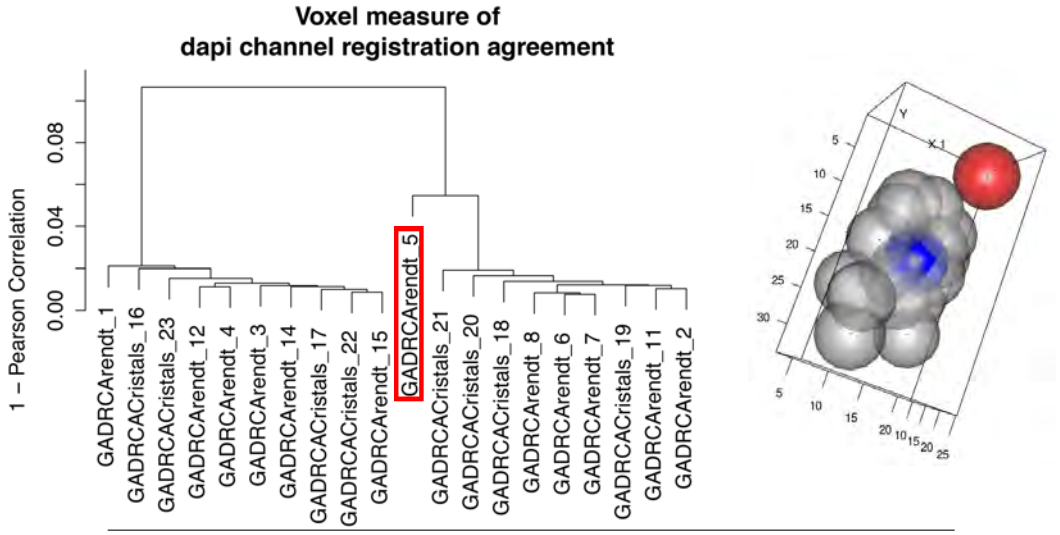

FIGURE 3.9: Automatic analysis of outliers.
On the left, the dendrogram representing the clustering of the super-voxel intensities for the registered DAPI channels of the 20 animals used in the analysis. Left cluster shows animals imaged from ventral side. Right cluster shows animals imaged from dorsal side. Outlier highlighted in red. On the right, virtual 3D representation of the segmented cells, as spheres centred on the centroid and with an averaged cell radius. Samples are represented in grey. In blue, a cell representation of the 'true' centroid, calculated as the median point of the 20 centroids. The virtual cell corresponding to the outlier is highlighted in red.

Outliers can be automatically identified

Another interesting aspect to analyse is whether we can identify, by image analysis, those samples (animals) that have not been registered properly (this happens mainly because the animal is misdeveloped or damaged). We can, for example, convert the DAPI channel to a super-voxel space (3D grid of groups of voxels), measure the intensity of each super-voxel and compare these measures for different samples. A simple pearson correlation analysis (figure 3.9) can already distinguish between the animals that have been imaged from the dorsal side from those imaged from the ventral side (the DAPI signal is more intense on those tissues that are closer to the objective during confocal microscopy imaging). This simple analysis can also find misaligned animals, like in the case of animal "GADRCArendt_5" (misalignment corroborated by manual inspection of aligned images). This animal, which shows the highest distance (based on DAPI signal analysis) to its cluster, is the sample for which the *GAD* cell analysed in this study shows the biggest discrepancy (Figure 3.9). Note that this outlier has not been removed during the analysis. Doing this kind of image analysis for the aligned animals can potentially improve the SPM results by identifying outliers. Note though that manual inspection of the registered images, although more time consuming, is still the best method to spot misalignments.

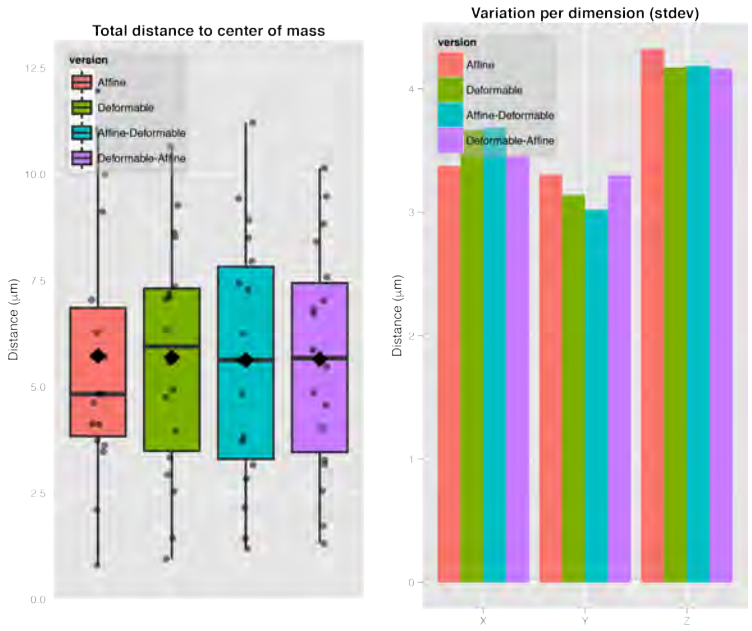

FIGURE 3.10: Analysis of image registration methods.
On the left, boxplot showing the distances between the cells centroids and the centroid of the spatial distribution for the different combinations of registration methods. Diamond-shape shows the mean for each distribution. On the right, standard deviation for the distributions of these distances in each of the axis. For this particular measure, the outliers were removed.

Registration methods do not significantly influence the final resolution

I also investigated the effect of the registration procedure on the final cellular resolution. Both of the previous implementations of PrImR (see Section 2.3) used affine (rigid) registration followed by deformable (non-rigid) registration of the samples to the reference. I saw no significant effect on the final spatial position of the *GAD* cells regardless of which of the two (or possible combination of the two) registration methods were employed (Figure 3.10). Given this observation, ProSPr is also based on a rigid transformation followed by a non-rigid one.

3.1.5 Image analysis and processing of the signal distributions

SPMs are highly heterogeneous

Once the conditions to generate the gene expression maps with cellular resolution were settled, I selected several genetic markers of interest, performed whole mount *in situ* hybridization (WMISH) on 6dpf larvae, and acquired stacks of several samples for each gene. I generated a reference using a subset of the samples, registered the animals to it, and built SPMs for all of them (see Section 3.1.6). I implemented a visualization panel for the SPMs and quantified their performance in a gene by

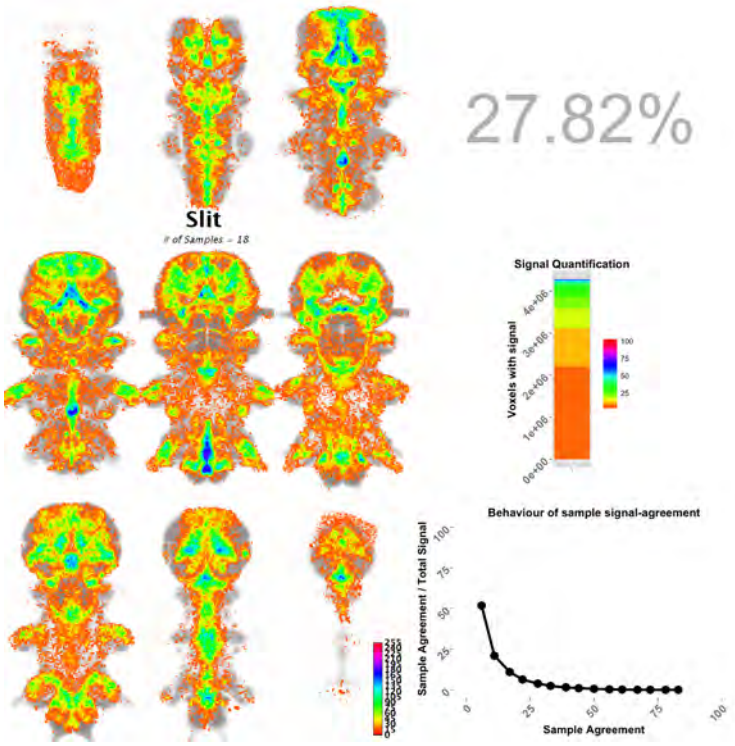

FIGURE 3.11: SPM analysis example.
On the left, visualization panel of the SPM for the gene *slit*. The nine images correspond to maximum projections of a 3D representation of the SPM overlapped to the reference (gray colour). From left to right and top to bottom, images correspond to ventro-dorsal portions of the body plan of *Platynereis* at 6dpf. SPM is colour-coded according to the signal agreement (scale in bottom right of the panel). Number of samples is indicated below the gene name. This panel is automatically generated for every SPM. On the right and from top to bottom, total quantification of signal in the body, stacked barplot showing the total amount of signal agreement in pixels, and behaviour of the signal decay as thresholding increases.

gene manner, in order to analyze all the data in an efficient way. Figure 3.11 shows an example of such an analysis for the gene *slit*.

A quick glance through the data already suggested that a common threshold or general binarization technique would not perform well across the different SPMs. This is because the SPMs differ in a gene to gene manner, possibly due to several reasons, like mRNA abundance, mRNA subcellular localization, signal location in the body plan, time of WMISH staining, signal to noise ratio, and other protocol uncontrolled variations. Figure 3.12 shows general quantifications of different parameters of the SPM signal for a hundred markers, as an illustration of the heterogeneous nature of the SPMs.

Given these differences, I considered applying a manual threshold for each genetic marker, as done in the previous implementations of PrImR, although this requires the user to be highly familiarized with the system and the particular gene. Nevertheless, careful analysis of individual SPMs showed that even for the same gene,

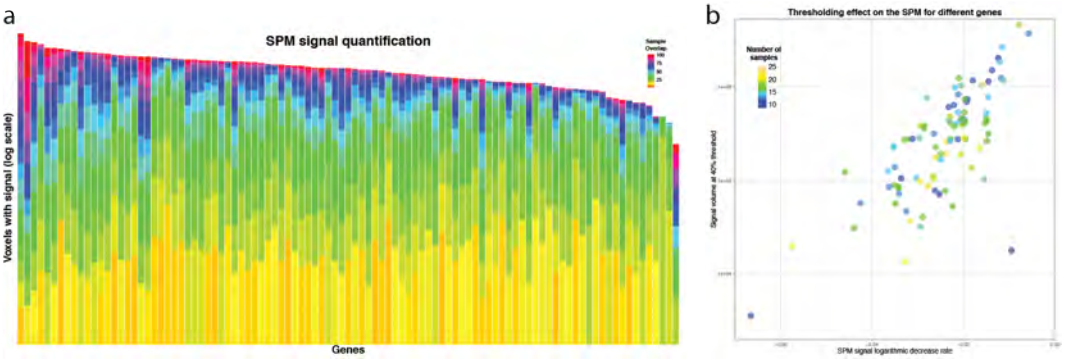

FIGURE 3.12: SPMs are highly heterogeneous.
Quantification of different parameters of the SPM signal for approximately a hundred markers. **(a)** stacked barplots showing the total signal and the sample agreement of each of this pixels (each bar corresponds to one genetic marker). **(b)** dot plot showing the low correlation between the number of samples in the SPM, the coverage of the signal, and the signal agreement decrease rate (as the threshold increases).

different regions would need a specific threshold. This is usually caused by different cells in the body expressing the gene at different levels. A different method than manual binarization was therefore required.

Minimal Expression Domains (MEDs) to threshold SPMs

Given the observations from the experiment on Section 3.1.3 and Figure 3.6, I aimed to threshold the SPMs so that the final volume of the resulting object approached the real one, as measured in the raw data. For that, I developed an image analysis routine to dissect the SPM into distinct three-dimensional domains. Each of these domains is then analyzed separately. Taking advantage that the animals are all registered to the same reference, the 3D space for each domain was measured in each of the samples composing the SPM. The total amount of signal was quantified for every sample, and the threshold that caused the SPM volume to better approach the real value was selected to binarize the SPM. This procedure can cause 3D domains composed by non-overlapping cells (or sub-domains) to separate in 3D, which I used to iteratively apply the same method to each, and therefore threshold every spatially distinct domain in a highly specific manner to approach the real signal, achieving cellular resolution and making the thresholding of the SPM completely automated. I call the resulting 3D objects Minimal Expression Domains or MEDs. The MED principle is shown in the schematic of Figure 3.13.

3.1.6 Generation of the dataset

The relatively high number of samples needed to calculate the spatial expression pattern for each gene, and the high number of genes that we aimed to incorporate

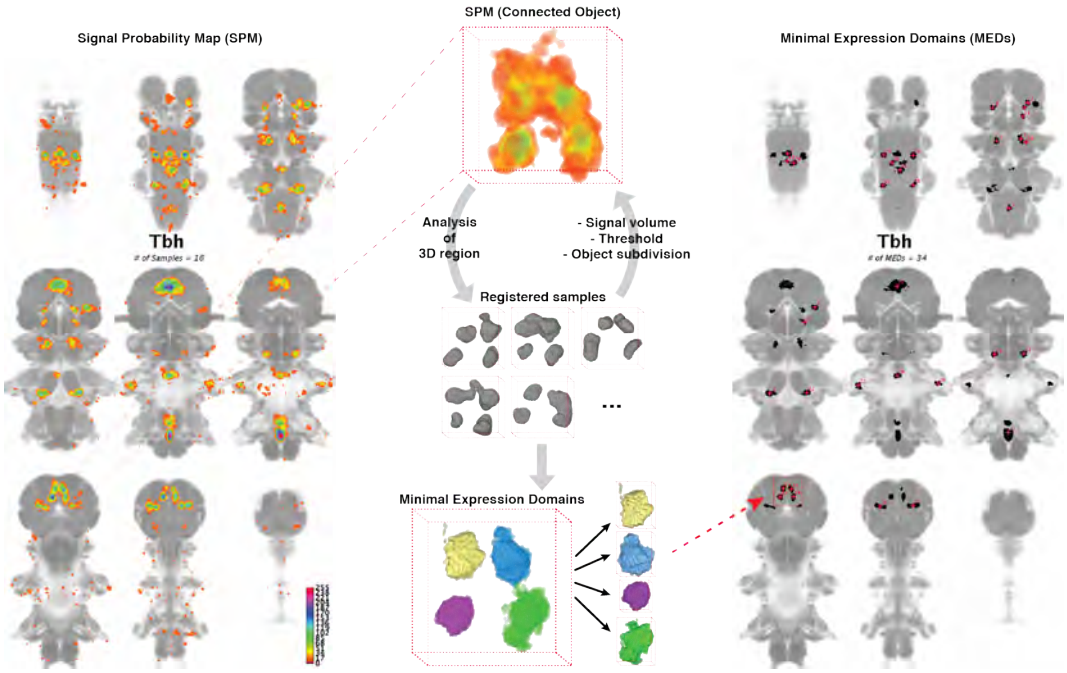

FIGURE 3.13: MEDs to dissect SPMs.
Schematic representation of the process to generate Minimal Expression Domains (MEDs), as an efficient and useful approach to threshold and dissect SPMs. On the left, a panel for the visualization of the SPM for the gene *tbh*. In the middle, representation of the algorithm principle to iteratively dissect 3D domains (in this case one composed by 4 cells in the middle of the head) into separate expression domains, by iteratively thresholding the object so that the resulting volume approaches the one measured in the registered samples. On the right, the final result of the SPM binarization, with a representation of the resulting MEDs, each of them having a unique ID, so that their specific information can be retrieved.

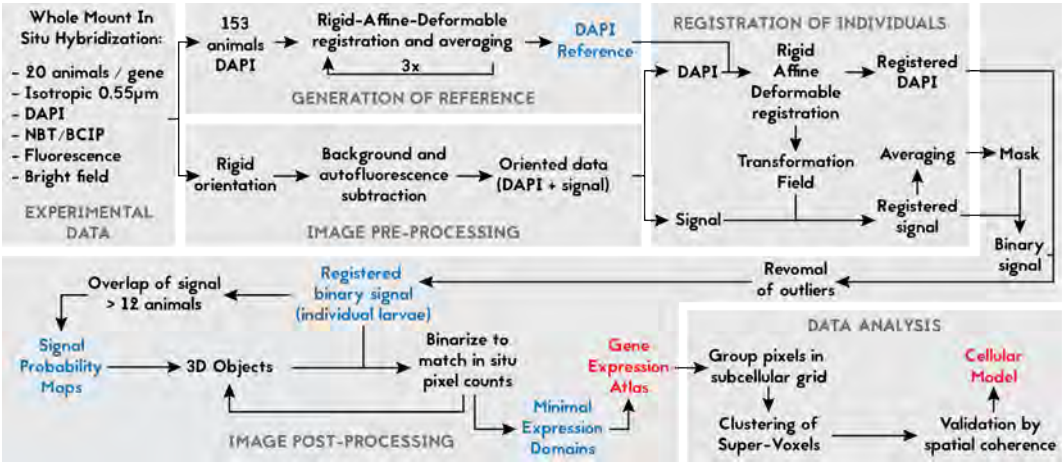

FIGURE 3.14: Flowchart of the method ProSPr.
In blue and red, the most important steps and results, respectively, of the process.

into such a study, require a highly automated image analysis workflow that ensures as little as possible and as easy as possible human intervention. Each of the workflow components (steps or modules of the full pipeline) has been automated, and optimized to work in a computer cluster so that computation can be paralellized for a time-efficient process. Each component produces comprehensive and easy to interpret visual outputs so that the process and results of the workflow can be inspected quickly, easily and effectively by the researcher in order to identify outliers or general errors. Algorithms have been coded in free software (Fiji, Python and R) to make the pipeline available to anyone, and it has been structured in such a way that the most relevant parameters can be adjusted easily if needed. Figure 3.14 shows a flowchart of ProSPr.

In the two previous sections I have already explained the two most important steps of the post-processing part (SPMs and MEDs). The experimental data details can be found in the Materials and Methods (Section 6.3). The image registration processing is identical to previous implementations of PrImR (for details see 2.3). The data analysis section is explained in the following sections as results are being presented. In this section I will present in more detail the construction and results of the reference, the steps for processing the images before registration, and the initial and final post-registration processing (*in situ* signal binarization and MEDs curation).

Reference

In order to align the animals to a common three-dimensional framework, a reference was generated using the DAPI signal of 153 individual larvae. The procedure used for the generation of a cannonical reference (average three-dimensional body plan), is based on a suggestion by Dr. Albina Asadulina and Dr. Stephan Saalfeld,

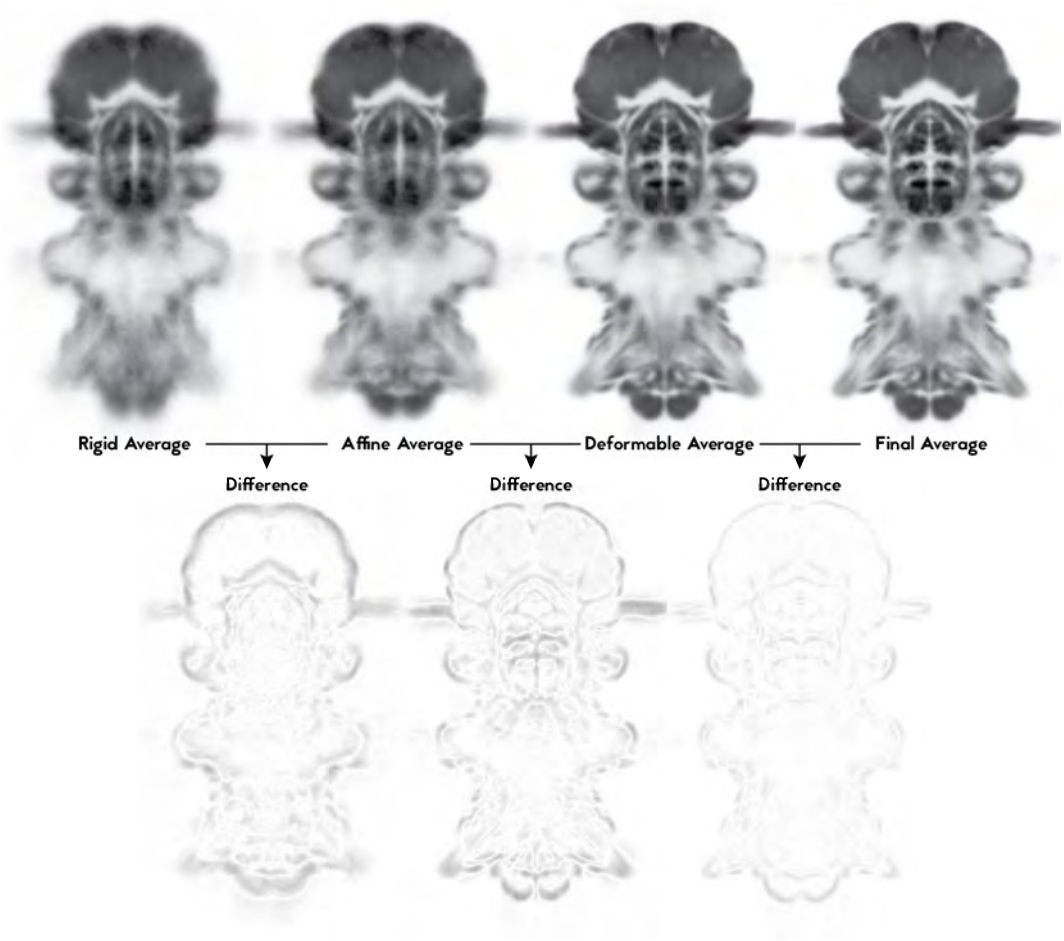

FIGURE 3.15: Refinement of the reference.
Central slices of the average results after registration of the 153 individuals (DAPI signal). In the top row from left to right, the averages resulting from the rigid registration, affine registration, and first and second affine and deformable registration. In the bottom row, the difference between the averages.

with some modifications. Put simply, it is built by rigidly aligning the animals (DAPI signal) between them and averaging the result. Raw DAPI signal of all animals is aligned using affine transformation to the rigid average, and an average of the resulting images is created. A round of affine and deformable registration of the original images is performed to the affine average, again averaging the result. A final round of affine and deformable registration is performed to create the final average that will become the reference. Figure 3.15 shows the results of the refinement process, where the increase in definition of the body plan structures (as grouping of cells) can be appreciated. Further rounds of registration and averaging did not improve further the quality of the reference (data not shown).

The process of iterative refinement resulted in a very high quality reference. This was somehow surprising, taking into account the differences between individuals (size, relative appendage positions, imaging orientations, and other morphological artefacts arising during the WMISH procedure). Figure 3.16 shows several virtual

single slices of the reference to illustrate the sharp tissue boundaries and high definition of cellular structures along the body plan. For example, the location of the muscles and the neuropil (devoid of DAPI signal) is very well preserved, even in the case of small nerves (see the head in Figure 3.16 c and d). The impossibility to discern individual nuclei within the neural tissues like the brain and VNC was expected, as this is not even possible in individual larvae after the WMISH procedure (tissue shrinks and these tissues are highly packed with cells). Nevertheless, the structure of the tissue is very well defined, and individual ganglia can be easily identified (see panel a in Figure 3.16). Other anatomical regions show an even higher contrast, like the stomodeum (bigger cells and highly stereotypic), or even the anal canal (Figure 3.16 c and h respectively). I only observed the gut to be less defined (see panels b,c,d and g in Figure 3.16) than other parts of the animal. This is possibly because the position of the cells composing this tissue varies considerably with respect to the surrounding areas, as the tissue as a whole appears more loose (attached in the anterior and posterior ends but somehow "floating" on the body cavity). Nevertheless, the neural tissue, which is the main focus of the project, is really well defined in the reference.

These observations, together with the high level of symmetry that the reference displays, indicates that the registration works very well at this stage of development, and that animals have a stereotypic body plan.

Image pre-processing

The acquisition of images in the confocal microscope has been designed to generate individual files (.lif extension in our case) for each genetic marker. Each file contains information for 4 channels (bright-field, DAPI, autofluorescence, and NBT/BCIP reflection) for several individual samples, ideally more than 20 (some animals might be discarded for various reasons during the process). For details on how to acquire the images please refer to Section 6.3.

The first part of ProSPr pipeline is to reorganize the data in the .lif file, splitting each channel to use the information contained in each one for a distinct purpose. The bright channel information is exclusively used to create a panel (see Figure 3.17a) where the user can see the images for each individual animal and immediately recognize any potential problem (e.g. sample out of field of view or not completely contained within). It helps navigate the data, but can be skipped during the acquisition. The autofluorescence channel is used to subtract, from the signal (NBT/BCIP reflection) channel, the intensity that is caused by animal autofluorescence (in the case of *Platynereis* at 6pdf, the glands or the chaete for instance). This subtraction greatly improves the signal to noise ratio, although is not strictly necessary to acquire (see Figure 3.17b).

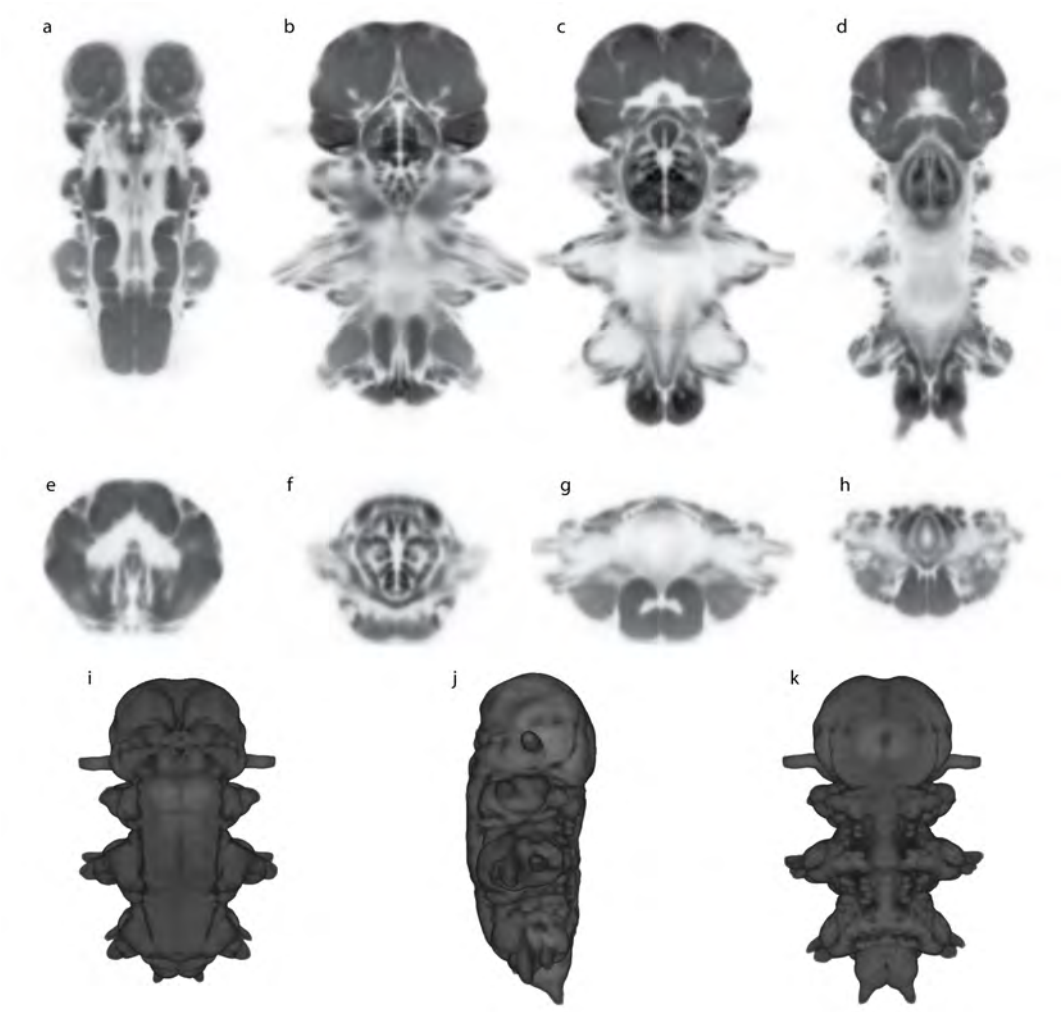

FIGURE 3.16: Reference for *Platynereis* 6dpf larvae.
(a-d) Coronal sections from ventral to dorsal (anterior up). **(a)** At the level of the ventral nerve cord (VNC) neuropil. **(b)** Dorsal to the VNC, at the level of the beginning of the stomodeum. **(c)** At the mid point of the stomodeum, gut, and brain neuropil. **(d)** At the level of the posterior part of the brain neuropil and end of the gut. **(e-h)** Transverse sections from anterior to posterior (ventral down). **(e)** At the level of the brain neuropil. **(f)** At the level of the stomodeum. **(g)** At the level of the second VNC ganglion. **(h)** At the level of the pygidium. **(i-k)** 3D view of the reference from the front, left side and back, respectively.

The two most important and completely necessary channels to acquire are the DAPI signal and the NBT/BCIP reflection, containing the information used for the three-dimensional transformation and the signal to generate the SPMs and the MEDs, respectively. The DAPI channel (whose intensity is corrected to normalize the effect of tissue penetration in the confocal) is used to orient the animal so that its position is the same as the reference (in this case, anterior up and ventral front). To this end, the anterio-posterior (A-P) axis is identified by the longer axis of the maximum projection of the image. The head is later identified by the fact that is the part of the animal whose "center" is further away from the outside (a progressive erosion procedure achieves this purpose). This part of the pipeline positions the animal with the head up. For calculating the dorso-ventral (D-V) orientation of the body, a maximum projection of the side view is generated and the longer (A-P) axis is positioned vertically along this axis. As the animals are bent towards the back (natural position of the body after WMISH), an ellipse is fitted to the head and its main axis angle is used to identify whether the animal was imaged from the ventral or the dorsal side. This part is used to position the animals ventro-dorsally. Note that this procedure relies on the fact that, due to the parapodia and chaete protruding sideways from the trunk, during the mounting of the slide for the imaging, the animals fall either on their belly or their back, and no left-right orientation correction is needed. All the parameters of this rigid transformation are applied both to the DAPI and the enhanced signal. Figure 3.17b illustrates this entire procedure.

The result is an automatic rotation of both the DAPI and the enhanced signal channels so that they are positioned in the same orientation as the reference, prior to the start of the affine and deformable registration step. For an easy inspection of the pre-registration step, a panel illustrating the oriented enhanced signal for every sample is automatically generated (Figure 3.17c).

Image post-processing: *in situ* signal binarization

The output of the registration process renders 3D aligned (to the common reference) DAPI and signal channels. As a requisite to build the SPM (see Section 3.1.3), the signal channel needs to be binarized, ideally in such a way that the signal to noise ratio is maximized. It is important to keep in mind that determining this ratio, even by eye, is not a trivial problem, given that both the signal and the noise (or background) are affected by many factors like gene expression strength, probe efficiency, staining time, and imaging laser power. The staining time specifically constitutes an arbitrary parameter in the pipeline, as it is completely up to the experimenter to determine when to stop the colorimetric reaction. Depending on this, the signal in different cells of the individual might vary in strength, and a compromise between what constitutes expression and what does not, needs to be found.

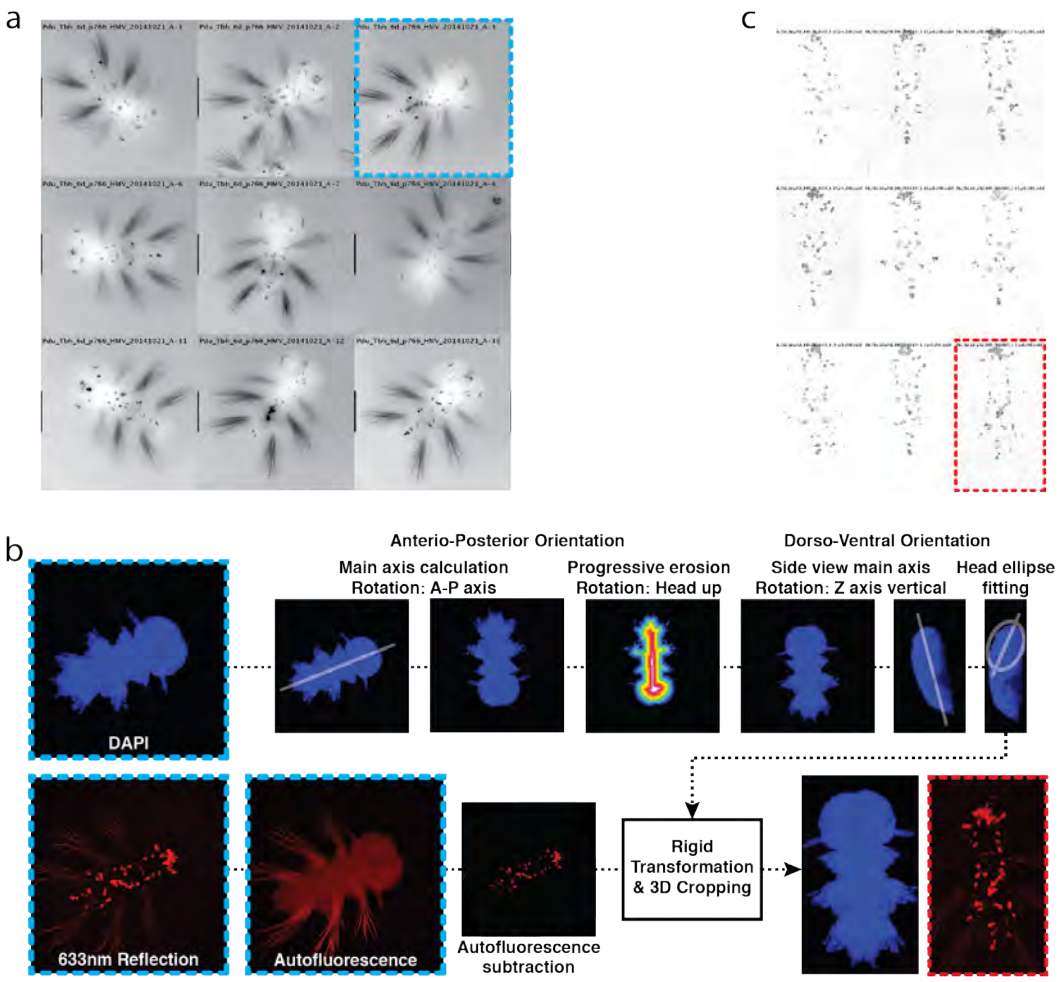

FIGURE 3.17: Pre-Registration processing of microscopy images.
Illustration of the steps and outputs of the image analysis pipeline prior to registration. **(a)** Bright-field panel of maximum projections of every individual in the microscopy file, indicating the ID for each individual. **(b)** Illustrative pipeline of the algorithms, explained in the main text. Note that the autofluorescence channel brightness has been enhanced for an easier appreciation. **(c)** Panel of maximum projections of the enhanced signal for every individual with their respective IDs. Note that only 9 animals have been selected for the panels a and c as illustrative examples, but these panels are typically composed of more than 20 individuals.

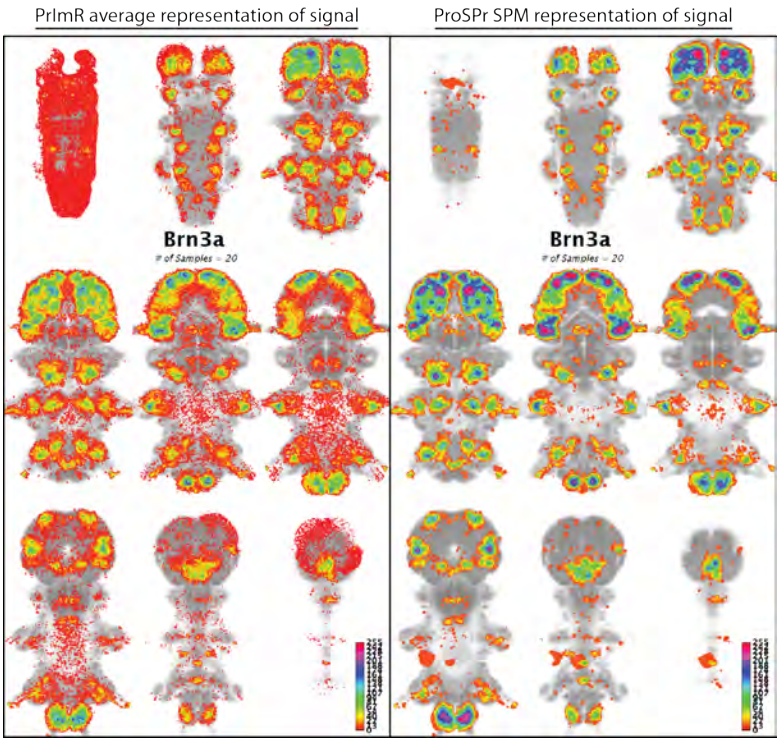

FIGURE 3.18: Comparison between PrImR average method and ProSPr SPM. Expression panels for gene *brn3a* as an example of the performance of the Signal Probability Maps of ProSPr and the regular average of PrImR to represent gene expression.

In previous implementations of PrImR, the individual registered signal channels are normalized and an average picture is calculated. Then the user manually thresholds that average (applying a common threshold value to the full image), based on previous knowledge about the expression pattern. The approach I followed in ProSPr builds upon this average, and represents a more systematic method that leads to a better representation of the signal in every case I have worked with (Figure 3.18).

As a first step, the average (same generation method as in PrImR) of the normalized registered samples is thresholded to create a mask to cover the expression domains. This mask is then applied to the individual registered samples, and for each one, the threshold is calculated based on the maximum difference between the inside and the outside of the mask (see Figure 3.19 a and b). This is based on the rationale that as the threshold is being increased, the noise signal is reduced at a higher rate than the real signal. Therefore, the pixel quantification difference (between the inside and outside of the mask) function keeps increasing. When the pixels inside the mask (containing the real signal), start being removed due to the increase in threshold, the difference function starts to decrease (see Figure 3.19b). The algorithm finds the maximum value of the difference function and applies this threshold to binarize the individual sample. The outliers are identified with the automatic generation of a maximum projection panel containing the DAPI and the binarized NBT/BCIP signals for each individual (Figure 3.19c). The binarized NBT/BCIP signals are overlapped to generate the SPM.

The algorithm generates three different masks using three automatic thresholding

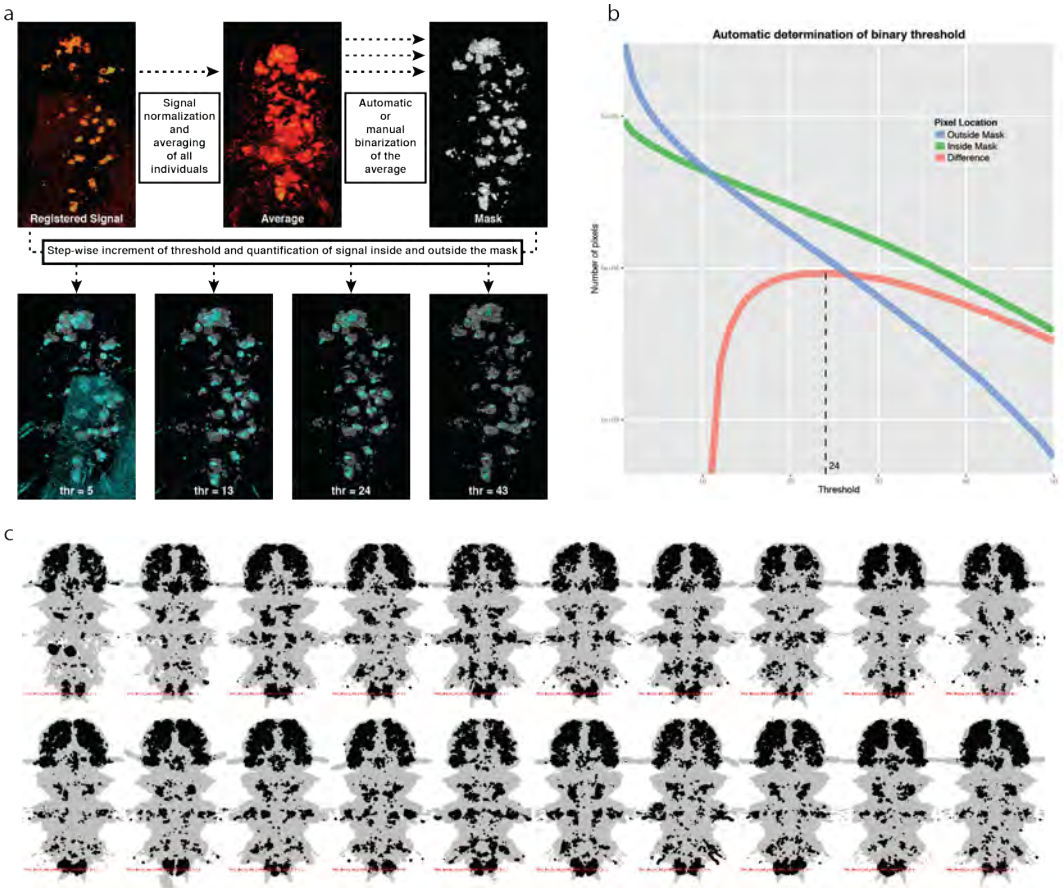

FIGURE 3.19: Post-registration process to binarize *in situ* signal.
(a) Illustration of the process of masking the samples. Registered individuals are normalized and averaged (PrImR method). This average is then thresholded (automatically and/or manually) to create a mask. The mask is applied to the individual samples, and the threshold value keeps increasing from 1 to 255. Images show snapshots of 3D representations for the gene *tbh*. **(b)** Quantification of the signals and difference between them. For each threshold, the signal inside and outside the mask is measured, and the difference between these values is computed. The algorithm identifies the first maxima of the difference function to assign a threshold value to the sample. **(c)** Panel showing the result of the algorithm with maximum projections of the binarized expression signal and the DAPI for each individual indicating their IDs (example for gene *brn3*).

methods (low, medium and high), and results are generated for each of them. The user can then survey the results and select the one best reflecting the true signal. I have also included a Graphical User Interface (GUI) to create the mask from the average selecting a manual threshold in case the automatic generation of the masks is not good. I have also coded another GUI to manually threshold each individual for those cases where the automatic binarization algorithm does not perform well (for this dataset this happened only once in more than a hundred cases). Although the arbitrarity of the system is partially corrected with the SPMs and the incorporation of the MEDs, the thresholding of the average to create the mask affects the binarization of the individual samples and can still influence the final result.

Image post-processing: MEDs curation

The final step of the image analysis process is the supervision of the final result: the MEDs that are going to be incorporated in the final atlas as three-dimensional coordinates of gene expression. Incorporating this step into the pipeline is necessary because the algorithm that dissects SPMs into MEDs is not able to fully identify the noise (and produces "noise MEDs"), and also because a final supervision by the experimenter is encouraged. I noticed that the "noise MEDs" were usually small, without a cellular shape, not symmetric, and located preferentially in parts of the body where unspecific signal usually appears. I therefore manually classified over a hundred small MEDs into two categories: real signal and noise. For each of the MEDs, I measured several parameters related to size, shape, location, and SPM signal structure. To generate an automated method to remove the noise MEDs, I asked whether principal component analysis (PCA) using these parameters could classify them into distinct groups. However, although there is some separation between the classes, PCA cannot differentiate clearly between real MEDs and "noise MEDs" (Figure 3.20a).

As I did not find an automated method to identify "noise MEDs", I implemented a GUI to curate the raw results of the MEDs in an efficient way (Figure 3.20b). The SPM panel and the resulting MED panel are automatically shown when the user clicks the gene, and the latter is updated as changes are being made. General parameters can be changed in order to remove MEDs automatically. Individual MEDs can be also classified as noise and removed by clicking on their identifier. The threshold for individual MEDs can also be adjusted if needed. It usually takes between 1 to 5 minutes to edit each result, depending on how noisy it is.

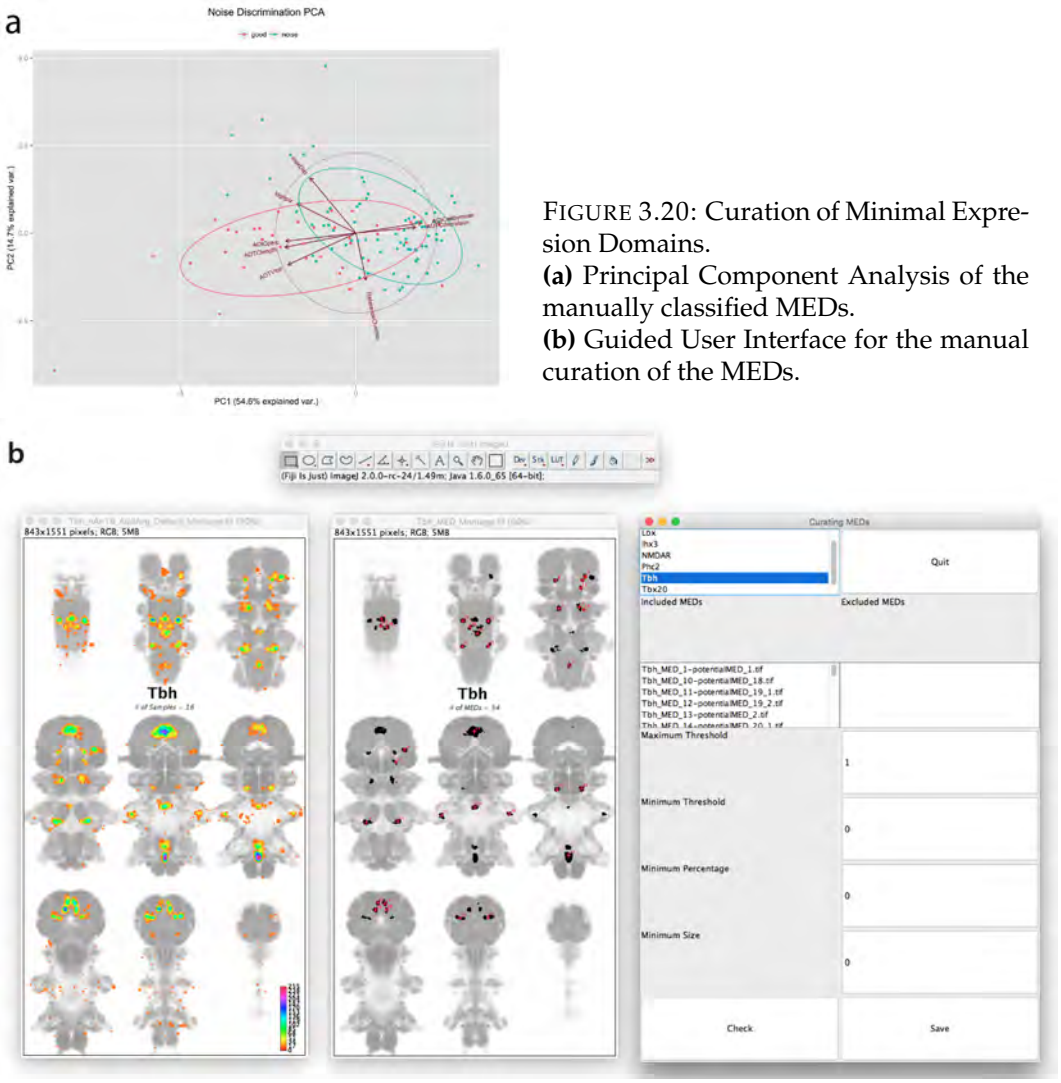

FIGURE 3.20: Curation of Minimal Expresion Domains.
(a) Principal Component Analysis of the manually classified MEDs.
(b) Guided User Interface for the manual curation of the MEDs.

3.2 Systematic analysis of *Platynereis* 6dpf expression atlas

In this section I explain the composition of the expression atlas that I have built for the 6dpf larvae, introduce the methods to reconstruct cellular information from the pixels and to mine the dataset, and focus on the analysis of the ventral nerve cord, which contains the cell types that are relevant for the topic of this dissertation.

3.2.1 Genetic markers included

Using the ProSPr pipeline described in previous sections, I have created an expression atlas for *Platynereis* at 6dpf that is composed of 138 genetic markers (134 distinct transcripts, 3 proliferation markers and 1 antibody[1]). 97 of the transcripts have been selected to represent the major TF families involved in neuronal specification (ANTP-HB, PRD-HB, LIM-HB, POU-HB, SINE-HB, TALE-HB, CUT-HB, PROS-HB, LIM-ZF, GATA-ZF, ZF-C2H2, bHLHs and FOX), as well as neurotransmitter-identity and differentiation markers. The other 37 transcripts were identified by Dr. Oleg Simakov in a transcriptomics study as genes preferentially expressed in *Platynereis* gut.

Table 3.1 shows the most common name for the markers included in the atlas. Figures 3.21, 3.22, 3.23, 3.24 and 3.25 show panels with ventral and lateral projections for all of the markers. I have cloned and performed WMISH for many other genetic markers, including all the ones depicted in Figure 2.3, but I did not detect any specific signal for them. Nevertheless, this does not imply that they are not expressed at 6dpf. As WMISH can be very hard at this stage, it is very possible that we need to try different protocol parameters in order to incorporate these genes into the atlas.

3.2.2 Creating a cellular model from the atlas

In order to make sense of the information contained in the atlas, I have developed a pipeline to reconstruct the cellular information, based on the expression and spatial coherence of the voxels constituting the atlas. This workflow can be described as containing three main steps:

- Build a super-voxel grid to group pixels.
- Cluster super-voxels based on molecular fingerprint (gene expression) coherence.
- Validate the clusters by the spatial coherence of their constituent super-voxels.

[1]The atlas contains an antibody staining against AcTub that is not included in subsequent gene expression analysis

AChE	ENR19-FboxLike	**FoxN**	Ngn	Sert
AcTub	ENR2-RYR2	FVRI	Nk2.1	Sim1
Arx	ENR20-PPIB	GAD	Nk2.2	Six12
BarH1	ENR22-CO1A1	Gata1/2/3	Nk6	Six4
Beta3	ENR25-Synaptopodin	Glt1	NMDAR	Slit
Brn1/2/4	ENR29-USP9X	GlyT	NOV1-globin-like	**Sox2**
Brn3	ENR3-JKIP3	GnRH	NOV15-OLM2A	Sox4
Bsx	ENR30-CCG5	GnrhR	NOV18-CEPU1	SoxB
Cal2	ENR31-Leucin-rich	Gpb	NOV2-unknown	**Sp8**
Calmodulin	ENR32-GRIK3	Hand	NOV29-unknown	Syt
ChAT	ENR34-MTHFSD	Hb9	NOV45-TMTC3	Syt12
Chx10	ENR39-RPC2	**HNF6**	NOV50-GDPD1	Syt7
Cnga	ENR4-SND1	Hox4	NOV52-KANL3	Syta
COE	ENR46-Calexcitin2	**Irx2/5/4/6**	NOV6-Stathmin	Tal1
Dach	ENR54-Boule-like	Isl	Olig	**Tal2**
DBH	ENR57-Junctophilin1	Lhx1/5	Otp	Tbh
Dbx1	ENR6-Nucleolin	Lhx2	Paraxis	Tbx20
Dlx	ENR62-NB5R3	Lhx3	Pax2/5/8	TH
edu3to4at6dpf	ENR64-PSMF1	Lhx6	Pax6	TrpH
edu4to5at6dpf	ENR69-BCA1	**Lmo4**	Pea3	TrpV
edu5to6at6dpf	ENR71-Patched	Lmx1	Phox2	Twist
ElaV	ENR8-NmMHC	**Maf**	Pikachurin	**Unc5**
En1	ENR9-NOE1	MHCL4	Pitx2	VAChT
ENR1-PGAM	Eve	Mitf	**PLCg**	VegfR
ENR10-UPP	Eya	MRLC2	Prox	VGluT
ENR12-UNC22	Fezf	Msx	Ptf1	Wnt16
ENR13-NDUS1	FOR	nAchR	**Robo**	
ENR16-ODO2	**FoxD**	NeuroD	**Scn8aa**	
Figure 3.21	Figure 3.22	Figure 3.23	Figure 3.24	Figure 3.25

TABLE 3.1: Genetic markers included in 6dpf ProSPr.
In bold, genes cloned by me. ENR and NOV label the genes identified in the transcriptomic study. The other markers were already published, or cloned by other people in the lab (Dr. Simakov, Dr. Achim, Dr. Bertucci, Dr. Brunet, Dr. Tosches and Dr. Vopalenski). The figure containing ventral and lateral projections for the genes within each column is indicated below the gene names.

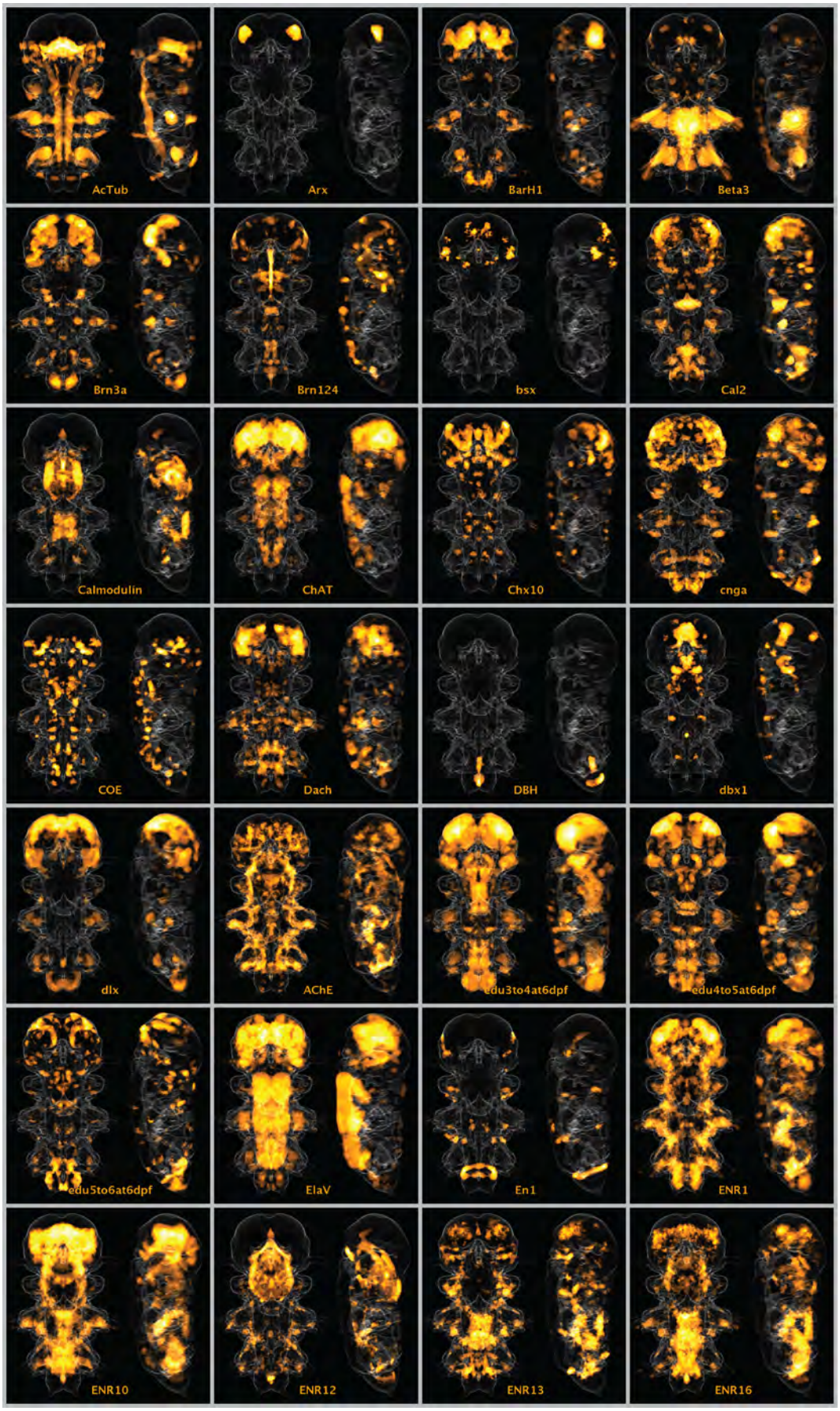

FIGURE 3.21: Genetic markers included in 6dpf ProSPr (I).

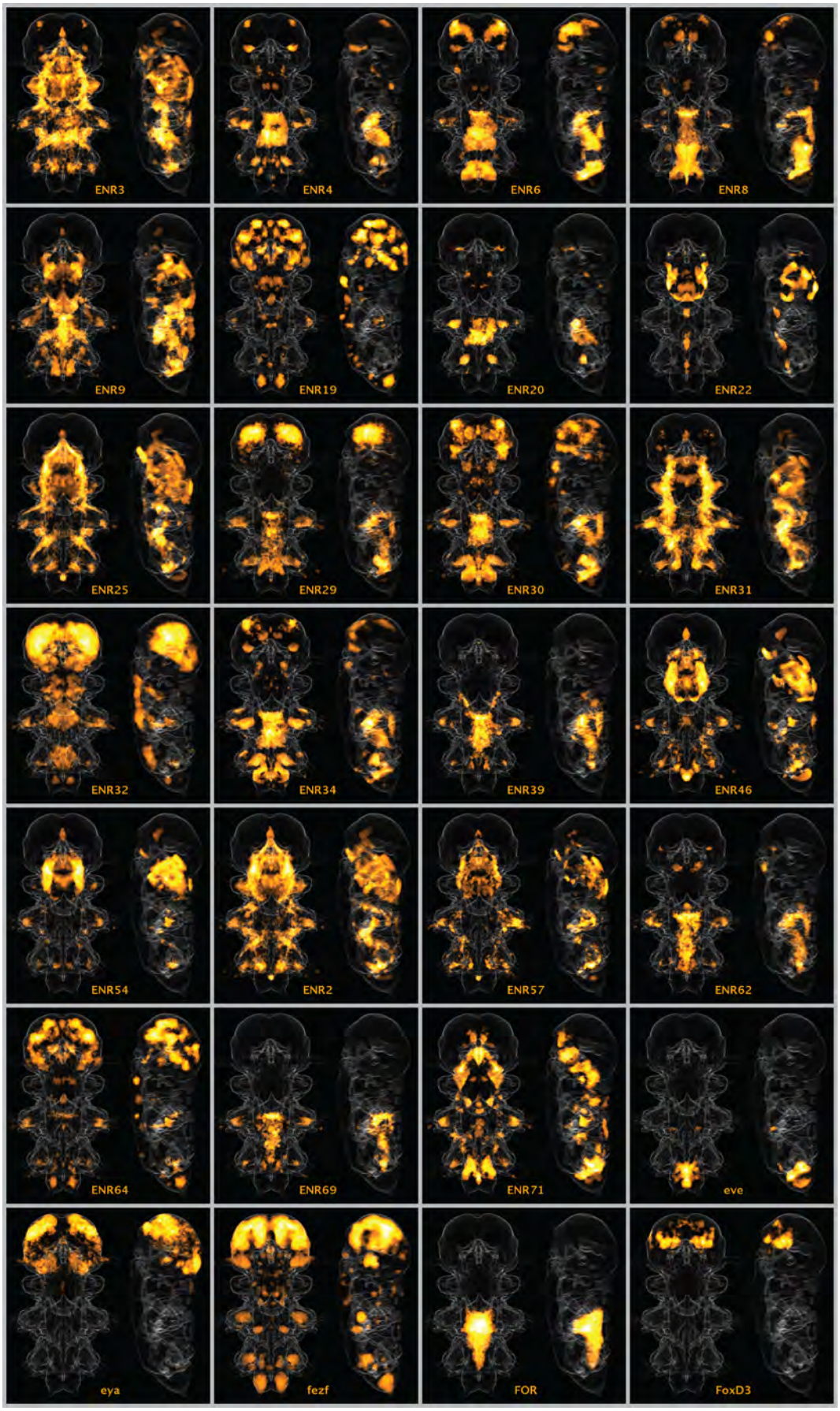

FIGURE 3.22: Genetic markers included in 6dpf ProSPr (II).

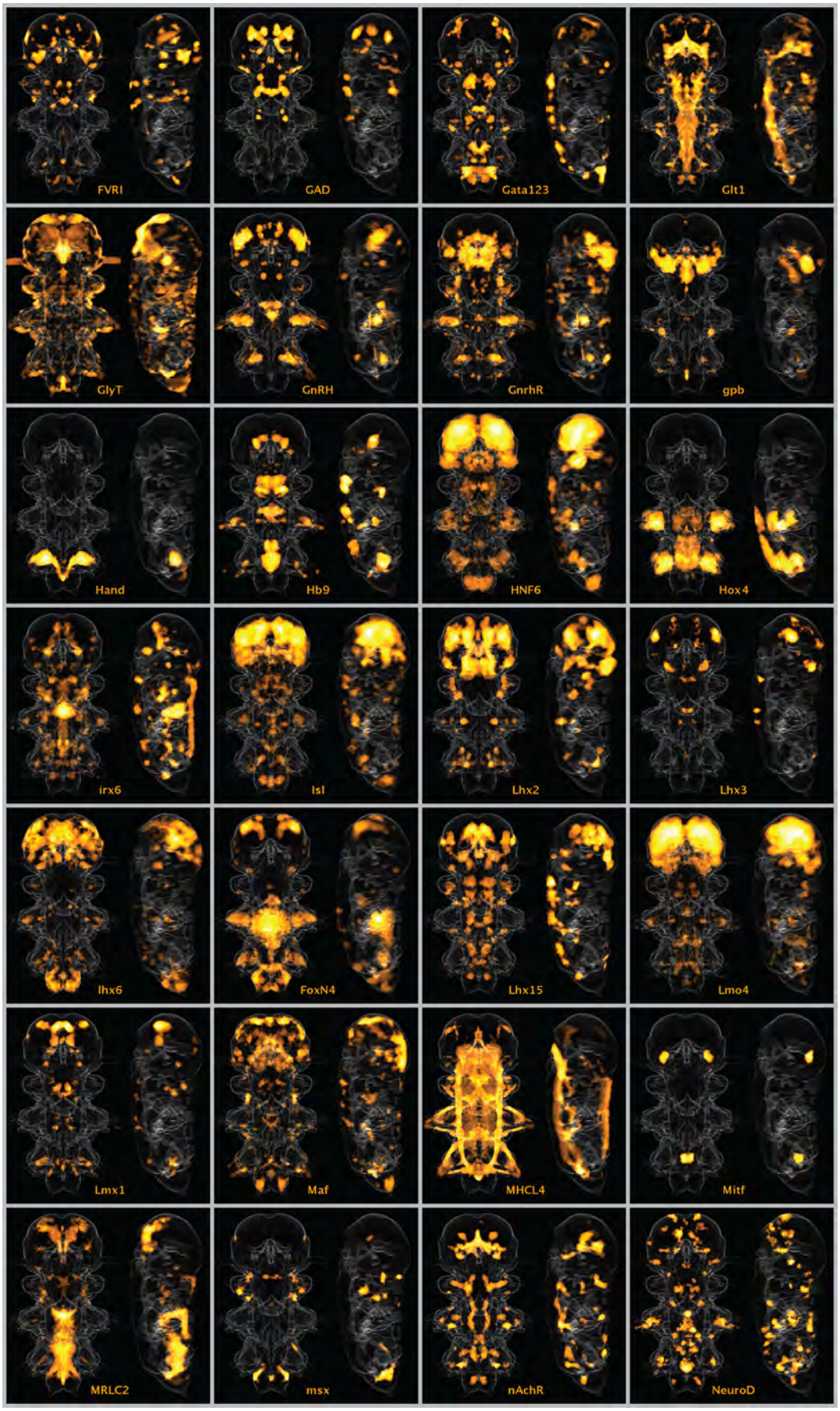

FIGURE 3.23: Genetic markers included in 6dpf ProSPr (III).

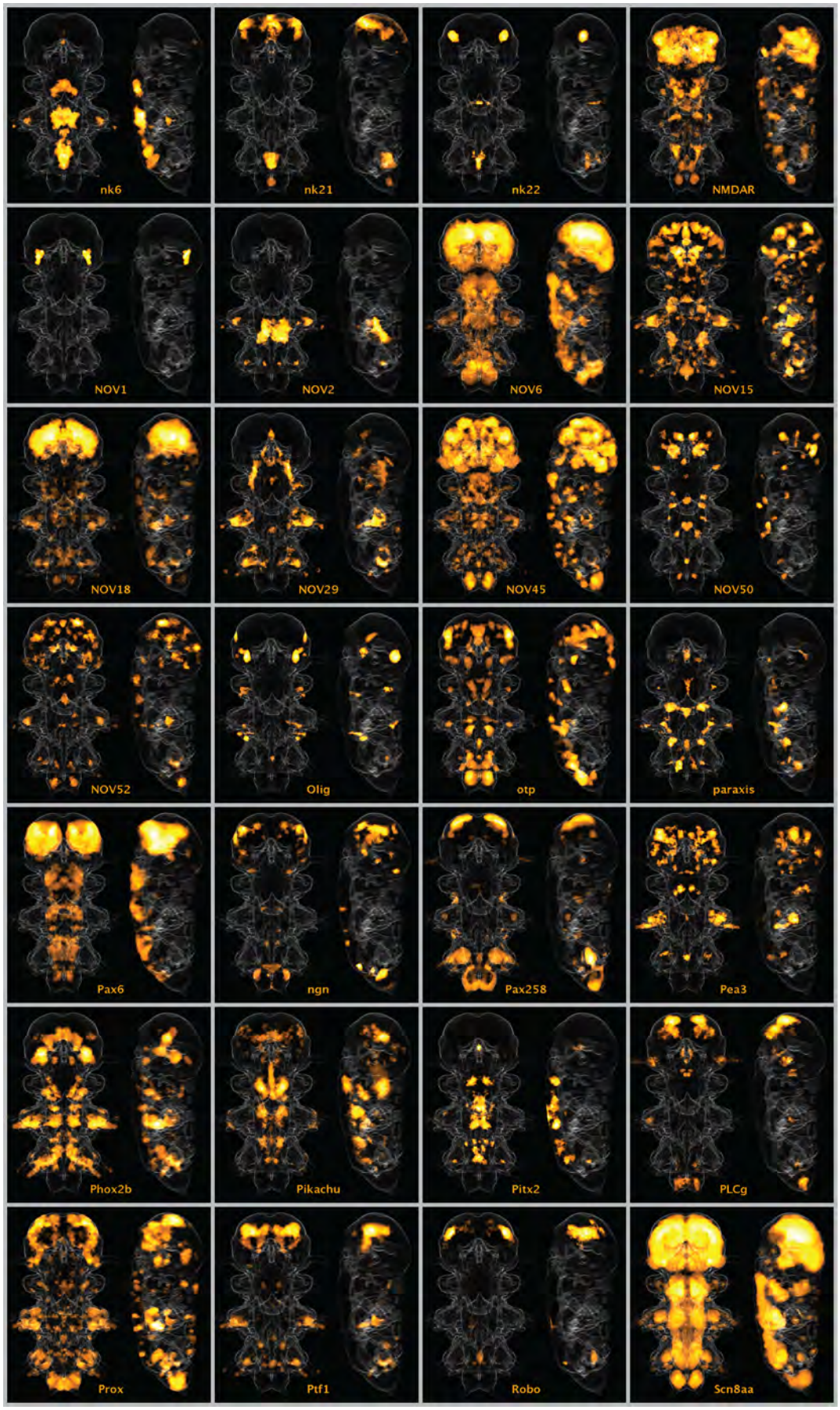

FIGURE 3.24: Genetic markers included in 6dpf ProSPr (IV).

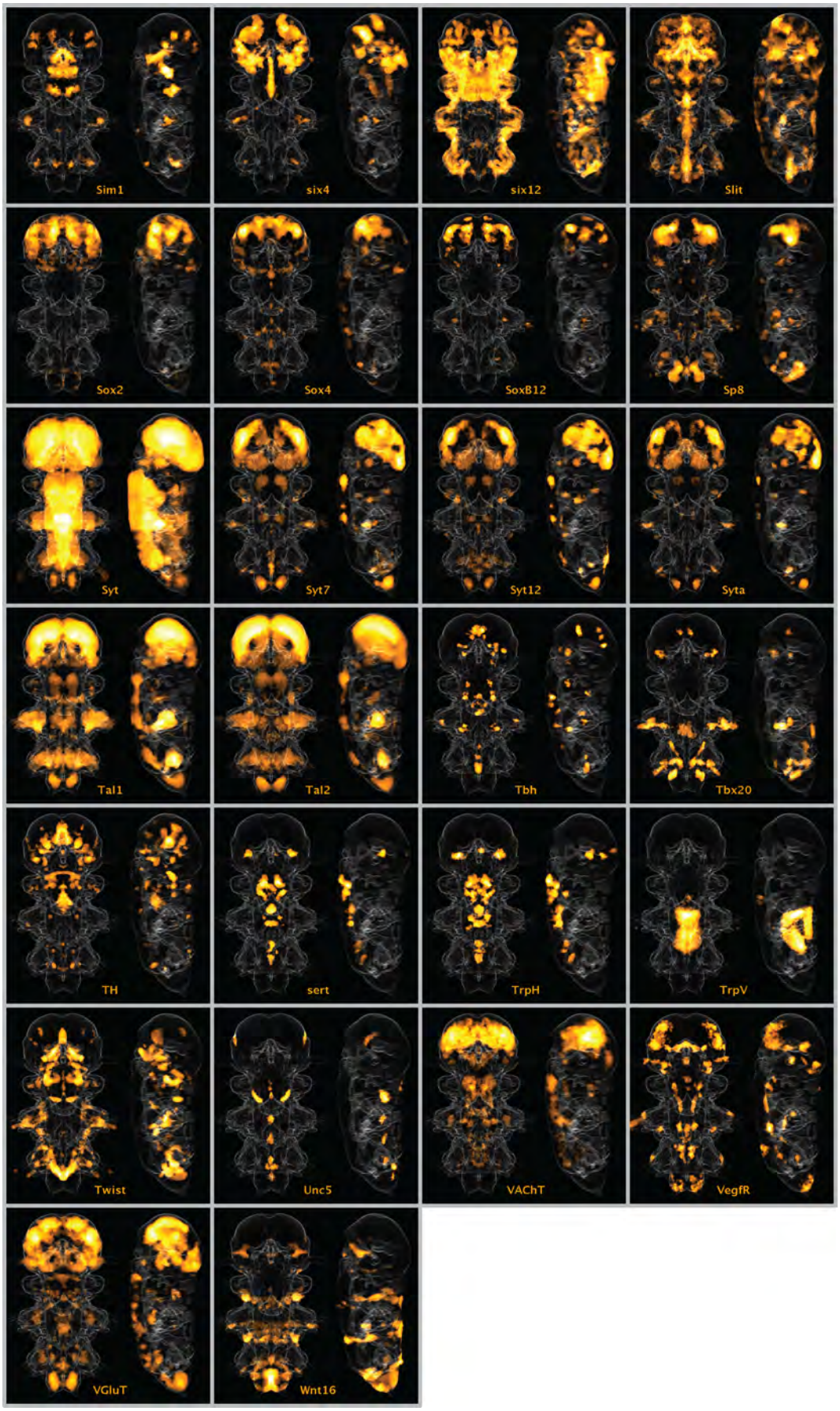

FIGURE 3.25: Genetic markers included in 6dpf ProSPr (V).

The outcome of this pipeline are groups of spatially proximal voxels with similar expression profile, most likely representing signal from real cells. I will now detail the different steps of the workflow.

Group voxels in sub-cellular three-dimensional grid

This step is necessary in order to reduce the dimensionality of the data, given that the atlas is composed of 15.374.324 pixels. To do this, a spatial grid is applied to the atlas, generating virtual cubes (super-voxels) that will contain the average information (gene expression) of the voxels that they cover. We can use the information in the super-voxels to measure the expression pattern correlation between the different markers in the atlas (Figure 3.26). A heatmap of these correlations forms clusters of genes that are associated with specific regions in the atlas. For example, the markers extracted from gut transcriptomes form three strong clusters representing different structures in *Platynereis* gut. It is not surprising to find many genes associated with nervous system, as this is the reason they have been selected to be included in the atlas. Interestingly, neural genes divide into two main categories. Those that are strongly restricted to neural tissue (marked as "Nervous system" and "Posterior peripheral brain" in Figure 3.26) and that correlate negatively with gut markers, and those that do not show a strong correlation with any other marker and are therefore not part of big clusters.

As a requisite to maintain cellular information, the volume of the super-voxels cannot be bigger than that of a real cell. Ideally, a cell should be represented by many super-voxels, at the same time that the dataset generated is small enough to efficiently handle its analysis. Given that the cell diameter in the raw data is roughly between 4μm and 6μm, and that the imaging resolution is 0.55μm (isotropic), each cell will be likely represented in the atlas by 250 to 1000 voxels. Using virtual cubes of 3 voxels per side (see Figure 3.27c), the generated super-voxels contain 27 voxels, which means each cell is covered by 10 to 40 data points, giving enough resolution to reconstruct the cells in subsequent steps of the workflow (see below). Before proceeding to group the super-voxels into cells, those with low correlation with their neighbouring super-voxels were removed (Figure 3.27d). These discarded super-voxels were likely representing cell boundaries and/or locations with data of poor quality.

Recursive partition of super-voxel clustering

In order to identify super-voxels representing signal from the same cell, I applied unsupervised hierarchical clustering to the data to group the super-voxels just by their expression profile, and followed an iterative procedure of matrix subdivision and reclustering to retrieve expression-coherent groups of super-voxels that

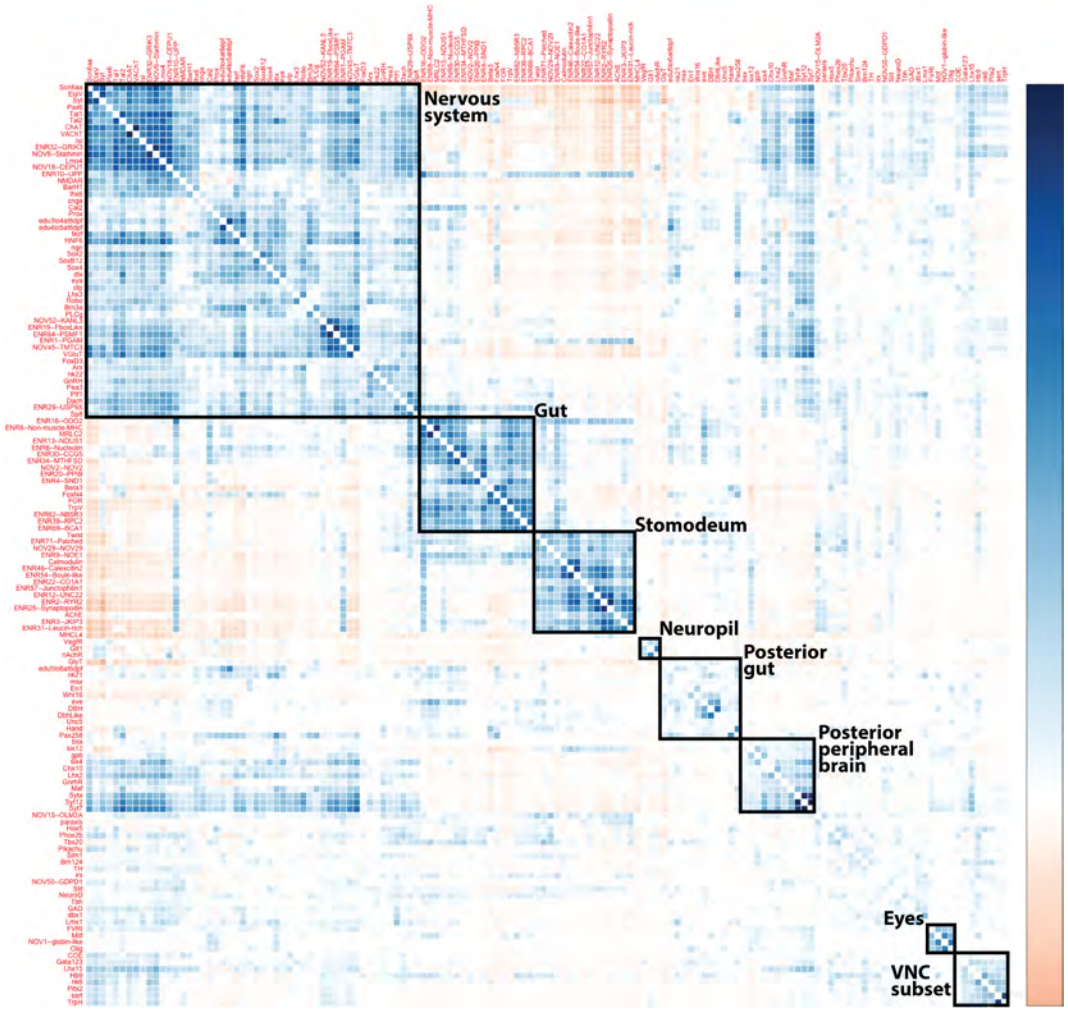

FIGURE 3.26: Cross-correlation of genetic markers in the atlas.

Animal Region	SV size	Total SV number	Cleaned SV	Cluster number	Cleaned clusters	Volume coverage
VNC	3	49878	46336	1553	**884**	**75.53%**
PNS	3	32593	29638	936	**381**	**53.77%**
Brain	5	22943	22274	3493	**2995**	**92.51%**
Pigid	3	12432	12334	424	**319**	**90.16%**
CS	3	10095	9480	323	**208**	**80.51%**

TABLE 3.2: Results of cellular reconstruction in different animal regions.

matched the size of a cell. I validated those clusters based on their spatial coherence (if representing the same cell they should be close in space[2]), and their inner-group correlation (similarity of expression profile). The workflow of the algorithm is outlined in Figure 3.27a. Only high quality super-voxel clusters were selected for further analysis (Figure 3.27 e and f). I refer to these clusters as cells.

Segmentation of tissues from the reference

Despite the data reduction process using super-voxels, the size of the atlas was still too big to efficiently apply the recursive partition and clustering algorithm to the entire animal. Luckily, the quality of the reference (see Section 3.1.6) is good enough to identify different tissues by hand. I therefore segmented different regions of the animal using TrackEM (Cardona et al., 2012), and reconstructed the cells for each of them separately. These regions include the main nervous system components of the body plan: the ventral nerve cord (VNC), the brain, the peripheral nervous system (PNS), the cryptic segment (CS), and the pygidium. Figure 3.27b shows the results of this segmentation process. Table 3.2 summarizes the results of the recursive partition and clustering algorithm for each of these regions.

The recursive partition algorithm performs very well on the data, as it retrieves as reconstructed cells bilateral groups of super-voxels that show a high spatial correlation (visual inspection of the data). Furthermore, after applying the high-quality filter, I obtained a total number of clusters that was very similar to the expected number of cells for each region, based on independent measures from confocal and electro-microscopy data (data not shown). This is also illustrated by the total volume covered by the super-voxels falling into clusters, although these numbers are under-representative as it is expected for the algorithm to remove super-voxels located across and in between cell boundaries.

[2]Bilaterality is taken into account by virtually folding the animal along the anterio-posterior axis.

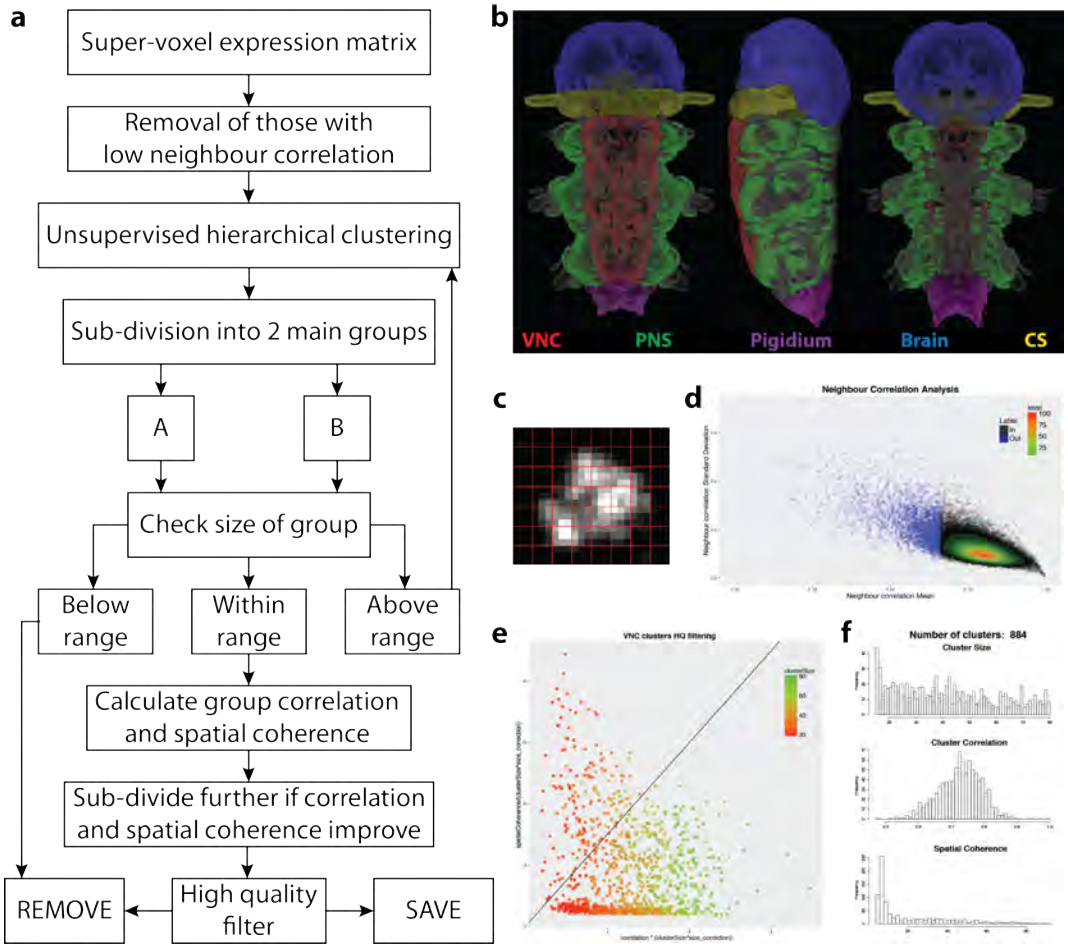

FIGURE 3.27: Reconstruction of a cellular model of the 6dpf *Platynereis* atlas.
(a) Workflow of the recursive partition of hierarchical clustering to group super-voxels into cells. **(b)** Manual segmentation of the reference to profile different tissues. VNC, ventral nerve cord; PNS, peripheral nervous system; CS, cryptic segment. **(c)** Cellular grid to build super-voxels with volume of 27 voxels. Illustration using the signal for the *GAD* cell used to calculate the resolution of ProSPr. **(d)** Removal of super-voxels with low neighbour correlation. **(e)** Selection of high quality super-voxel clusters based on inner-group correlation, spatial coherence (measured as medium distance between the super-voxels), and size. **(c)** Histograms of high quality clusters properties.

3.2.3 Relationship between the cells in *Platynereis* nervous system

In order to start exploring the similarities in expression profile of the 4782 bilateral cells reconstructed from the segmented regions, I applied two different unsupervised clustering methods to the dataset. First, I used hierarchical clustering to visualize these relationships as a heatmap, which offers a direct read of the expression profile for all the cells (Figure 3.28). Clustering methods are directly dependent on the distance measure that is used to calculate the similarity of multidimensional points. I tested many different distance methods, and the one grouping the data with more structure was the *jaccard* method from the vegan package in R (https://github.com/vegandevs/vegan).

Hierarchical clustering shows five big groups of cells. One (not labelled in Figure 3.28), contains a mixture of cells coming from different parts of the body plan. Nevertheless, clear clusters of cells, strongly defined by the coexpression of a handful of specific genes, can be appreciated across this group. Both the brain and the VNC, besides having cells in this group, are contained mainly in two groups each. One way of mining the data is to label interesting groups of cells in the reference to see where they are located (Figure 3.29). For instance, the two brain groups correspond to the inner brain and part of the cortex, and the two VNC groups indicate that the cells in the inner part of the segments are more related to one another than to the periphery of the segments. Group number 5 is shown to illustrate the usefulness of these approaches to discover distinct structures within the atlas. I have developed several tools to guide this process of clustering, selection, mapping and reclustering to efficiently mine the dataset using hierarchical clustering.

Given the nature of the data, I used a dimensionality reduction algorithm (t - distributed stochastic neighbour embedding or tSNE) to visualize the relationship between cells across *Platynereis* nervous system as a scatter plot (Figure 3.30), where clusters of nearby cells indicate similar expression profiles. Again, we can see that both the brain and the VNC show a higher general coherence than the other three tissues, although with this clustering algorithm all the tissues are clustered in some defined groups, and many of these clusters are formed by cells located in different parts of the body. We are currently analysing the expression profile and spatial location of the cells forming these clusters to investigate their developmental and evolutionary relationships. As for this thesis, I will focus on the analysis and interpretation of the VNC, as it is in this structure where the locomotor CPG related neurons are most likely located.

3.2.4 Analysis of cells in the VNC

With the aim of characterizing the molecular substructure of the VNC, I used tSNE on the 884 cells that were extracted from this segmented region, using the signal

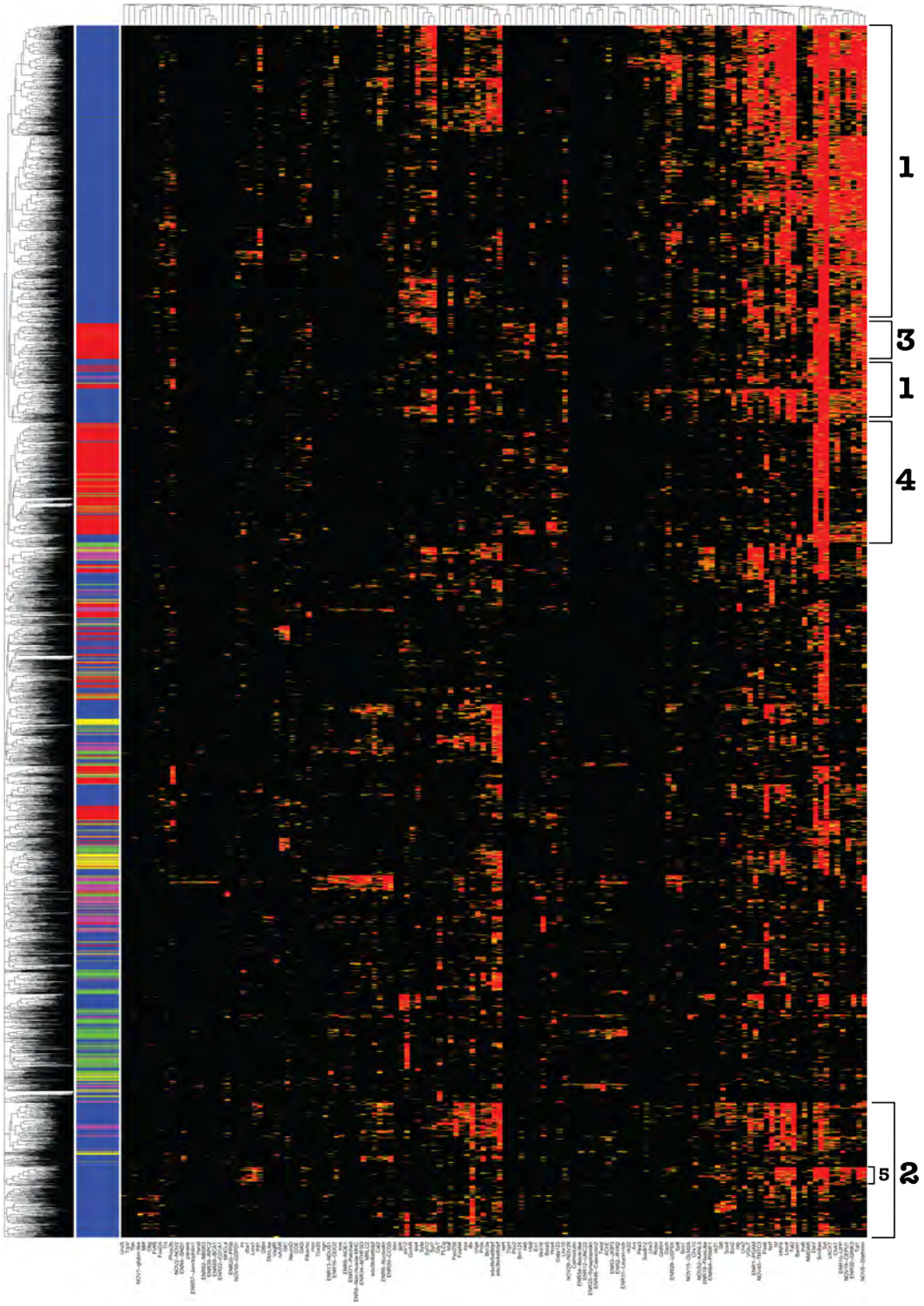

FIGURE 3.28: Hierarchical clustering and heatmap of reconstructed cells.
Heatmap representing the *jaccard* hierarchical clustering of 4782 cells and 137 markers. From yellow (50% coverage), to red (100% coverage), heatmap colourmap indicates the gene coverage value for every cell. Colours in the left indicate the region of each cell (see Figure 3.27b). Cell clusters selected on the right are depicted in Figure 3.29.

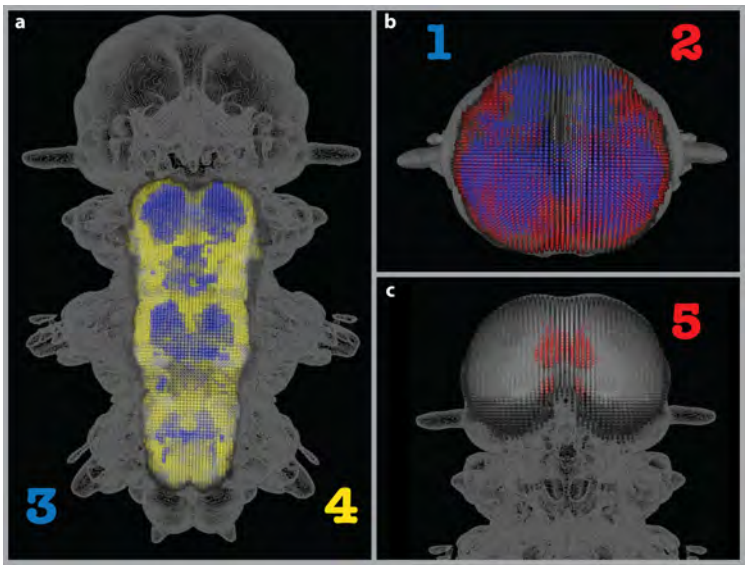

FIGURE 3.29: Spatial location of heatmap cell clusters.
Spatial mapping for some clusters depicted in Figure 3.28. **(a)** Ventral view of the animal, anterior side up. **(b)** Apical view of the head, ventral side down. **(c)** Ventral view of the top half of the animal, anterior side up.

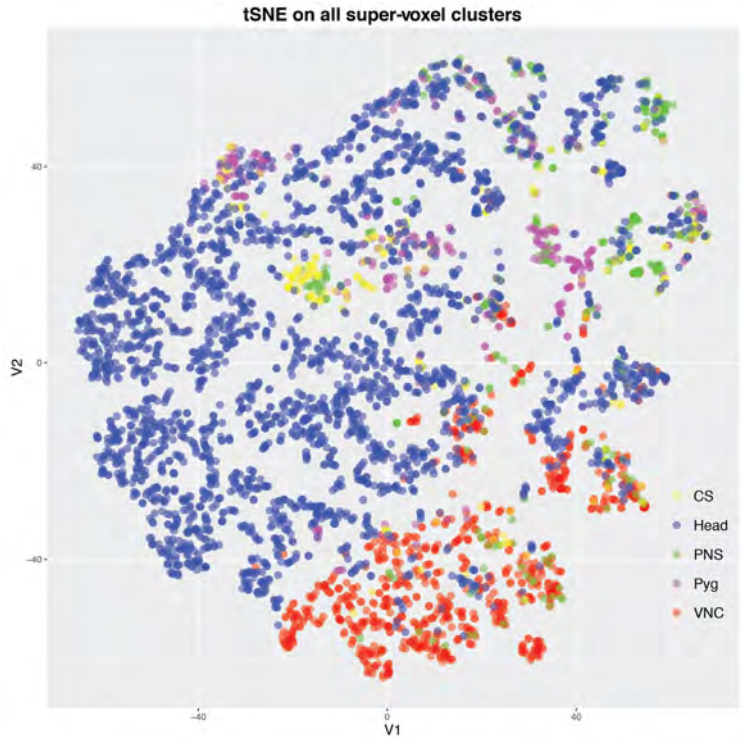

FIGURE 3.30: tSNE clustering of reconstructed cells.

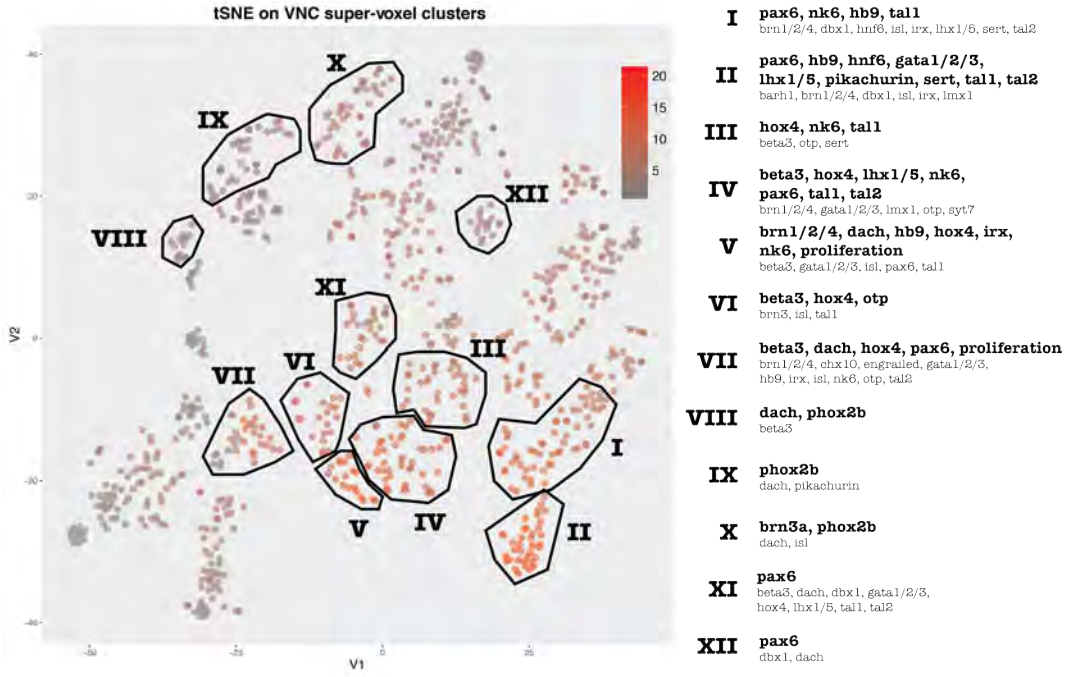

FIGURE 3.31: VNC tSNE clustering.
tSNE analysis of the VNC reconstructed cells and manual annotation of clusters. Cells are color-coded according to how many genes they show expression for. Genes strongly expressed in a cluster are indicated in bold, and those with expression in a subset of cells in regular font.

from 75 informative genetic markers in the atlas[3]. tSNE distributes the cells generating clusters that differ in size and point spread, and are not correlated with total expression in the cells (Figure 3.31). I inspected the expression of genes over the tSNE plot (see Figure 3.32) to determine which genes were strongly associated with gene clusters, and manually defined twelve groups of cells based on their clustering by tSNE and their gene content (Figure 3.31).

Ventro-dorsal arrangement of the VNC

In order to visualise these molecularly-defined cellular clusters, I mapped the spatial location of their constituent cells into the atlas (Figure 3.33). Interestingly, all of these clusters show a strong spatial coherence in their ventro-dorsal, medio-lateral, and segmental location of the VNC. For instance, clusters VIII, IX and X are deeply embedded in the ganglia, positioned in the most dorsal part of the VNC. There is also a strong separation between these three clusters and the rest in the tSNE plot, and they specifically express the transcription factor Phox2. Cluster VII is located on the most ventral surface of the VNC and, as expected, is strongly defined by proliferation markers (see also Section 2.1.4).

[3]Genes not expressed in the VNC were not included, as well as genes from the gut transcriptomics experiment.

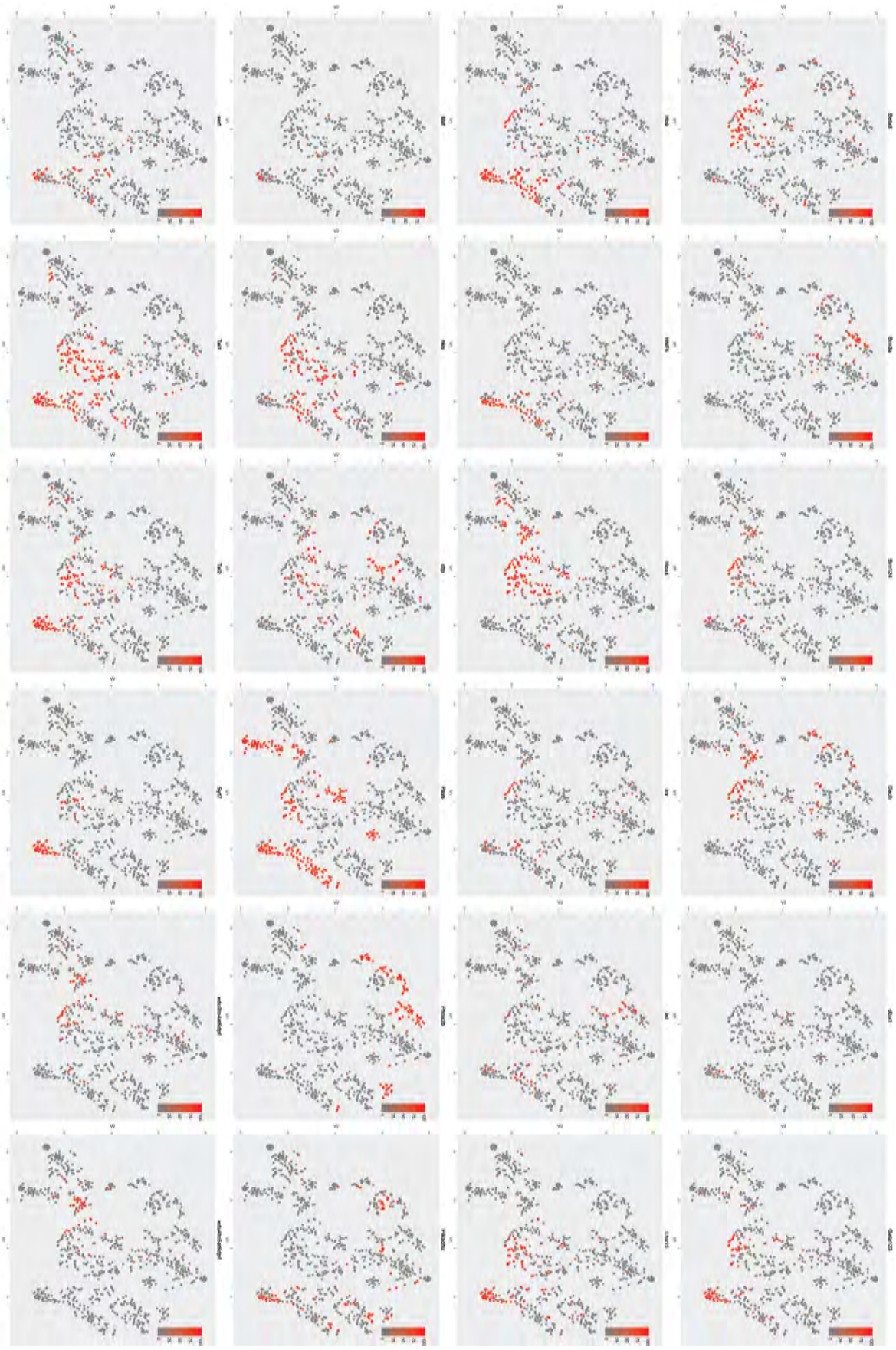

FIGURE 3.32: VNC tSNE cluster-defining molecular markers.
Gene expression mapped into the tSNE VNC clustering for some of the cluster-defining molecular markers.

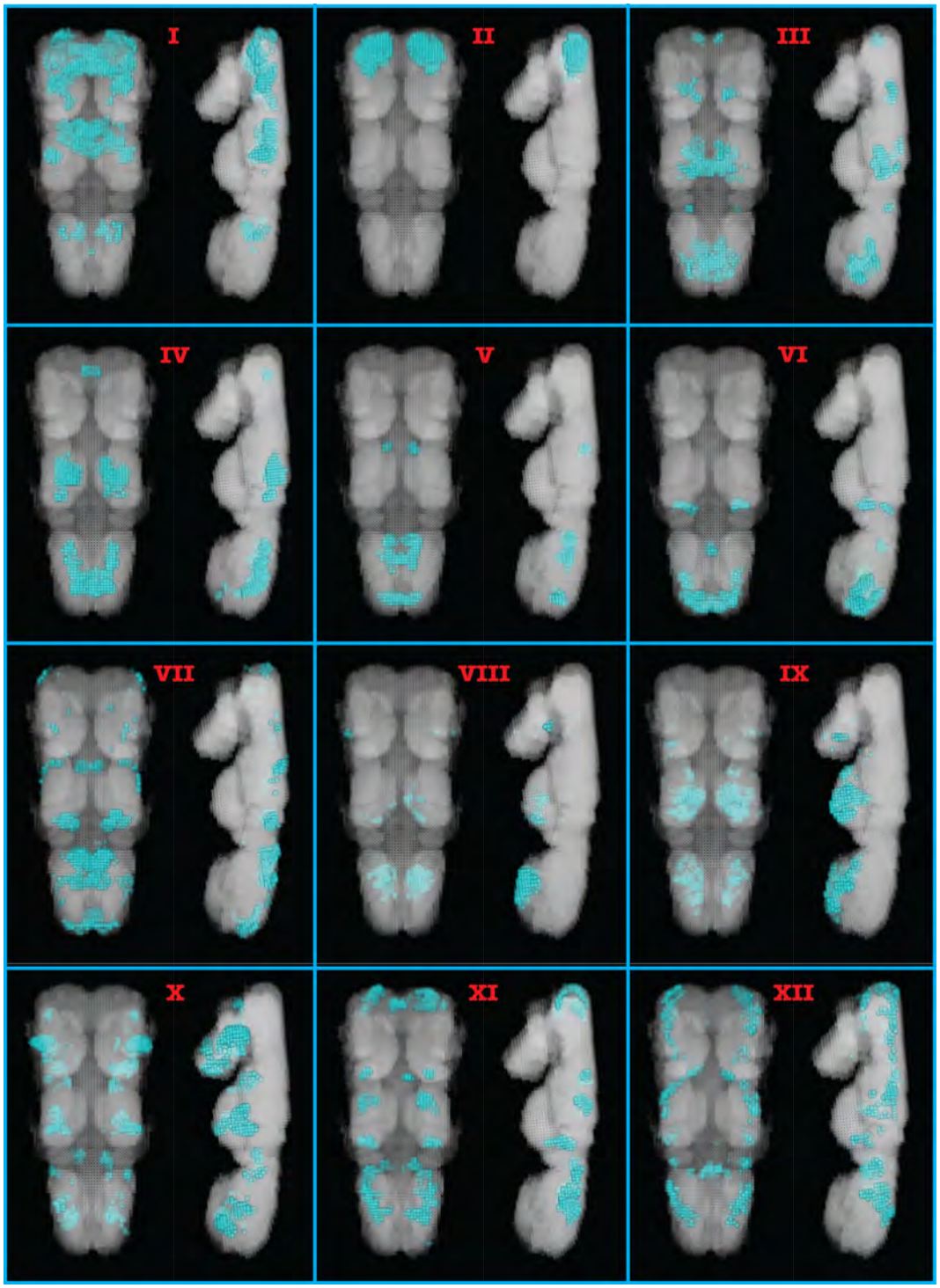

FIGURE 3.33: Spatial location of tSNE VNC clusters.

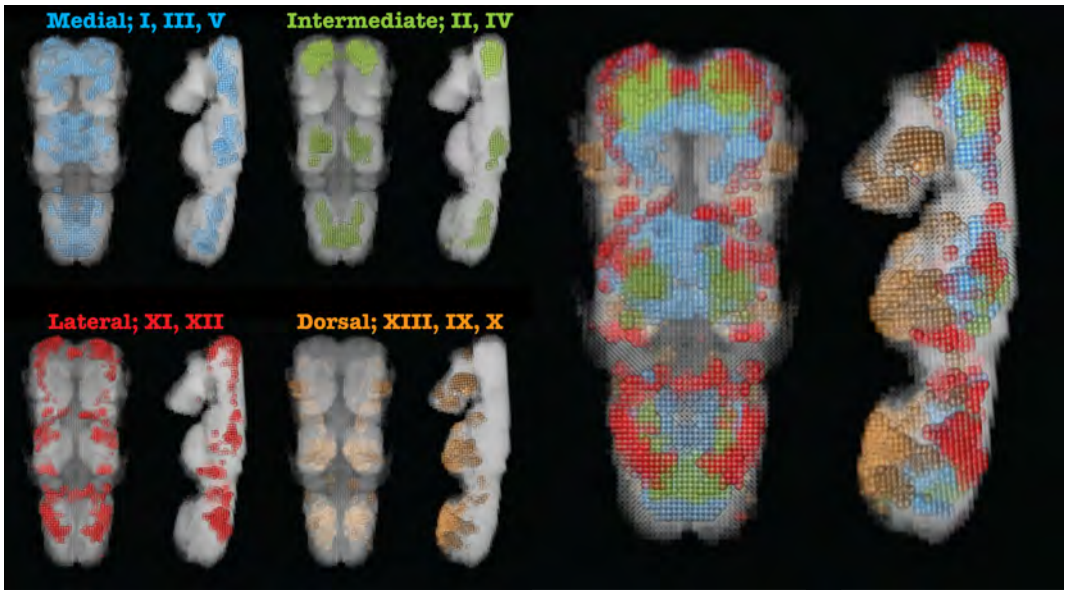

FIGURE 3.34: Medio-lateral arrangement of clusters.
Segmental grouping of the clusters in Figure 3.33.

The first segment contains a unique population of cell types

The rest of the clusters occupy a more intermediate position along the ventro-dorsal axis, and show a distinct medio-lateral disposition (analysed in the next section), and a segmental bias. Most strikingly, cluster II shows a unique population of cells, which only exist in the first segment, and expresses up to 20 out of the 75 markers (see Figure 3.31). Complementary to this, clusters IV, V and VI show cells similarly located (respectively), in the second and third segment, but not in the first. This indicates that the first segment strongly differs in composition (at least with respect to some cell types), with the second and third segments, and possibly a specialization of the first segment into other tasks not only related to a direct control of the parapodia during crawling. This is consistent with the little role of the first segment for locomotion (see Results section 3.3), and its later incorporation into the head after metamorphosis (Fischer, Henrich, and Arendt, 2010).

3.2.5 The identified clusters are arranged in medio-lateral columns

One of the most regular properties of the clusters was their specific medio-lateral arrangement within the segment or segments. Indeed, if we group them by their distance to the midline, a clear columnar pattern arises for all the segments (Figure 3.34).

For instance, clusters I and III, although distinct between them (cluster I is located in the anterior part of the segments and cluster III in the posterior part), occupy a medial position. Cluster V most likely represents the developing part of this group

of clusters, as it is located in the medial region of the third segment and expresses proliferative markers. Slightly on top (on the ventral surface) and laterally to clusters I and III we find clusters II and IV. These clusters form an intermediate ring, and despite showing differences that relate to the first segment with the second and the third (mentioned in the previous section), they represent cells that are characterized by similar markers (e.g. Lhx1/5 and Tal proteins). Interestingly, a repeated segmental pattern is found in between the pairs I-II and III-IV (first and second-third segments respectively), both in their spatial location as well as in the relative position of these clusters in the tSNE scatter plot. Clusters XI and XII are located in the most lateral domains of the VNC and span the three segments. In addition, some cells seem to be located in between the segments. This group is strongly characterized by the expression of Dbx1, and is the cell population that I have begun to characterize functionally (see Section 3.4). As previously mentioned, and following this medio-lateral route, we next find the most dorsally located clusters (XIII, IX and X).

In the discussion (see Section 4.2) I elaborate more on what this medio-lateral arrangement means in terms of evolutionary conservation of cell types, comparing the genetic markers enriched in these regions with those that specify the vertebrate spinal cord.

3.2.6 General neurotransmitters are not specific to particular clusters or columns

I analysed the location of neurotransmitter-defined cells (gabaergic, cholinergic and glutamatergic)[4] within the molecular clusters found by tSNE in the VNC. Interestingly, they are not particularly enriched or specific to any of the identified clusters, and are rather dispersed throughout the plot, and except for GAD, which shows an intermediate and lateral restriction, VAChT and VGluT (which are expressed in many cells in the VNC) can be found across all the medio-lateral domains (Figure 3.35). Despite the domains of neurotransmitters are largely non-overlapping, a few cells show co-expression for some of them (also including serotonergic markers), indicating that some neurons might use a combination of neurotransmitters.

3.2.7 ProSPr applied to other developmental stages

One of the major limitations of defining the molecular profile of cell types by their differentiation stage is that their developmental specification is not captured. This

[4] I used specific enzymes in the production of neurotransmitters to define their transmitter identity.

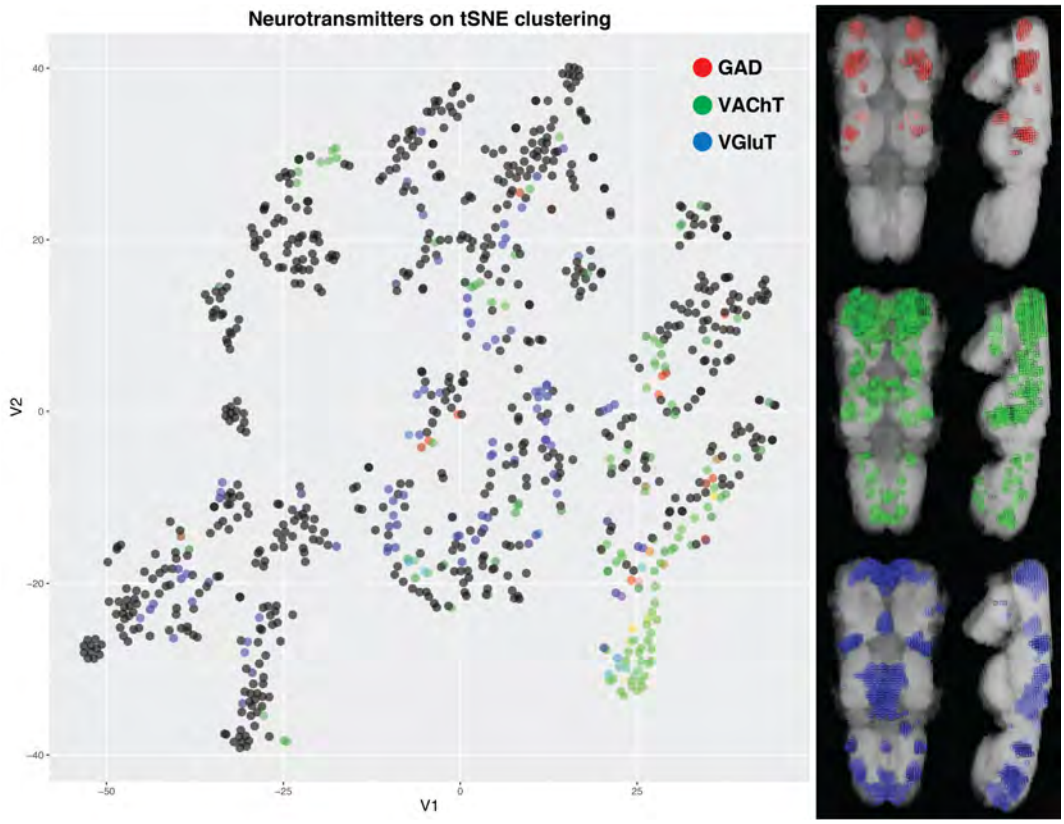

FIGURE 3.35: tSNE and spatial location of major neurotransmitters.
The scatter plot shows the expression of the cells for three major neurotransmitter markers using an RGB colour code system depending on their expression (indicated in the legend). On the right, the location of the cells within the VNC expressing these markers.

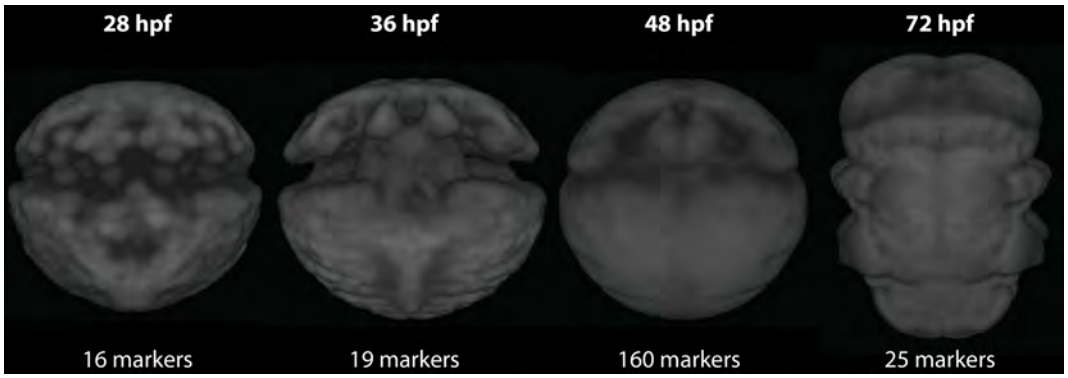

FIGURE 3.36: ProSPr references through development.
References not at the same scale.

is the case with the results presented above, where we have generated a comprehensive but static picture of a stage of *Platynereis* nervous system. Adding the dimension of time to an expression atlas is not a trivial thing to do, but ProSPr provides a perfect tool to start filling the temporal gaps by building different atlases for crucial developmental stages. On our way to do this, Dr. Kaia Achim, Dr. Paola Bertucci and I generated four additional atlases for 28, 36, 48 and 72 hours post fertilization using ProSPr pipeline. As expected, given that these stages contain less cells, are highly stereotypic, and the WMISH is less problematic, the results (SPMs and MEDs) are easier to achieve than at 6dpf.

Figure 3.36 shows the four references and the amount of genetic markers in the atlas at the moment of writing this paragraph, as these numbers are currently growing every day. The resource at 48hpf now contains enough data to perform the analysis described in previous sections, and we are using this atlas to spatially map single-cell transcriptome data to a full organism (work in progress and manuscript in preparation). The reference of 48hpf is courtesy of Sanja Jasek from the Jékely lab, as we teamed up to generate common references that could be used to group data across different labs into a unique resource for the field.

3.3 Monitoring larval locomotion

At 6dpf, *Platynereis dumerilii* larvae regularly perform two main modes of locomotion: swimming using ciliary bands and crawling using the parapodia. In the first case, similarly as they do in earlier stages, the animals use a synchronous beating of the cilia of the multi-ciliated cells located in concentric bands distributed along the anterio-posterior body axis. This mode of locomotion allows the animals to travel long distances very fast and efficiently, and to move up and down the water column. The crawling mode of locomotion develops between 4 and 6 dpf, as it is associated with the three pairs of parapodia, located in the trunk of the animal (see Section 2.1.3). At these stages, *Platynereis* larvae exhibit an exploratory behaviour of the substrate, when they pause from swimming and settle to the floor (pelago-benthic transition) crawling over small areas, probably sensing them in order to feed, as this ability develops at the same time. This exploratory behaviour is characterized by an alternation of an anterio-posterior movement (stroke) of the six parapodia, so the larva can crawl on the substrate over short distances. These rhythmic oscillations of parapodia, that result in a highly coordinated locomotion, indicate the presence of neural circuits assembled into Central Pattern Generators (CPGs). Valuable information about the functionality of these CPGs can be extracted by just observing the animal movement.

In order to characterize *Platynereis dumerilii* crawling behaviour, I have designed a behavioural setup that forces the animals to perform these movements continuously while being recorded, and I have developed a series of image analysis and tracking routines to process and analyse the videos of these recordings. These algorithms are coded in Matlab® and assembled into a Graphical User Interface (GUI) that I have called **PduLoco**, that stands for ***P****latynereis* ***du****merilii* **Loco**motion monitoring tool. The behavioural set-up is very simple and I will briefly describe it first. Second, I will explain the functionality of PduLoco step by step. Lastly, I will present what we have learned about *Platynereis* crawling behaviour using this tool.

3.3.1 Behavioural set-up and video acquisition

Recording free swimming *Platynereis* larvae is possible and indeed has been done to monitor the vertical migration in the water column of hundreds of individuals while keeping track of single larvae (Guhmann et al., 2015). Even though this is an ideal approach for quantifying behaviour (as the natural actions of the animals are monitored for many larvae simultaneously), it cannot be used in the case of crawling locomotion for two main reasons. First, high resolution image acquisition is needed so that the different appendages (parapodia) of the individual larvae can be easily recognized. Second, animals can switch from crawling to swimming very fast, so any set-up without vertical confinement will lose the focus on the animals

and the tracking of individuals would be unsuccessful. Therefore the requirements of the set-up must be as follows:

- Vertical confinement, but free exploration in 2D (substrate).
- High spatial resolution imaging, focusing on one individual and achieving a high contrast of the body plan.
- High temporal resolution imaging as the movements of the parapodia can occur very fast.

Such a behavioural chamber can be achieved by positioning a microscope cover-slip on top of a water drop that contains a single larvae. The vertical space created by hydrostatic pressure is such that the animals can crawl in any direction, but cannot swim vertically. The use of a transparent cover-slip allows the larvae to be recorded with a high-speed (50-60Hz) camera in a normal stereoscope using bright field imaging. In my case, grey scale images at 640x460 were taken, as these were the constraints of the cameras[5], although the software should be able to work with any video composed of .bmp formatted images. Given that PduLoco can identify and track the larvae by zooming in on it (see below), the field of view can be large enough so that the animal can move around a certain distance.

3.3.2 PduLoco Graphical User Interface

PduLoco was initially inspired by an animal body posture tracking software described in Gomez-Marin et al., 2012, and although it clearly differs from the original source in many ways, some functionalities have been kept. PduLoco can track the movements of the most posterior parapodia[6] with relation to the body of the animal, in order to quantify the phases and timing of movements. In order to do this, the pigments associated to the parapodia[7] and the eyes of the animal are being tracked using image analysis algorithms based on intensity and size of these structures. Figure 3.37 shows a snapshot of PduLoco GUI, outlining the different panels of the program, which will be explained in the following paragraphs. The GUI is divided in two parts: the tracking step (Figure 3.37 I and II), and the analysis step (Figure 3.37 III and IV). The first one performs the actual tracking, and is done in such a way that the user can guide the tracking if needed, and spot and correct mistakes very easily. The second part analyses the tracking, where the user can adjust a few parameters interactively to properly identify the parapodial strokes.

[5]Cameras used for the recordings: DMK 42BUC03 and DFK 23UM021 from The Imaging Source®.

[6]Initial analysis of the crawling videos identified the parapodia of the second and third segment as the ones moving coordinately. The parapodia of the first segment are smaller and don't seem to contribute much to the crawling behaviour.

[7]Fortunately, *Platynereis* larvae shows six pigments in the trunk, each in close proximity to the base of the chaete of each parapodia. As the parapodia is doing the stroke movement, the pigment moves accordingly, making it a very reliable source to monitor the locomotion of the animal.

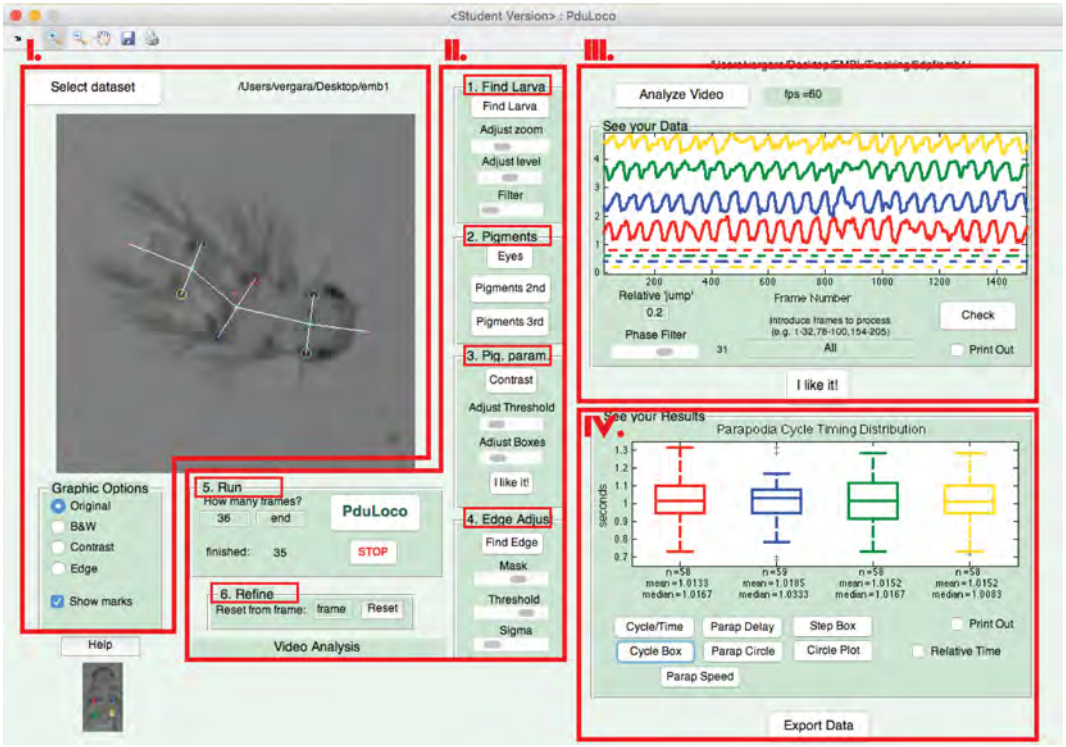

FIGURE 3.37: PduLoco Graphical User Interface.
Panels **I** and **II** are used for tracking purposes. Panels **III** and **IV** are used for the analysis of the tracking results. See main text for a detailed explanation of the software functionalities.

PduLoco I: Tracking performance display

This panel of the GUI displays the video frame by frame as the movements of the parapodia are being tracked. During the tracking, PduLoco computes several parameters of each frame, all related to information of the body of *Platynereis*, such as the binary image of the segmented larvae, or the contour of the body plan. The user can select, in an interactive and dynamic manner, the type of information that is being displayed, as a way to supervise the performance of the method. The different modes of information are related to the parameters explained in the following paragraph. The live tracking results (identification of the eyes and pigments) are displayed, as well as the skeleton of the animal body plan.

PduLoco II: Tracking parameters and command panel

This part is composed of several panels that are meant to guide the user, step by step, through the adjustments of different parameters for the correct tracking of pigments and eyes. The output panel (Figure 3.37 I) is automatically updated during all the steps to monitor the results of the changes.

1. Find the larvae. As the recorded area is larger than the animal, in order to restrict the analysis to the immediate surroundings of the larvae, an image of the background arena is subtracted from the initial frame. The black and white binarization threshold and a size filter can be adjusted by the user until the larvae is correctly identified. To track the whole animal during the full video, the centroid of the body is then calculated from the previous frame to center the view.

2. Identify the pigments. In this panel, the user clicks on the posterior eyes and the four most posterior parapodia-associated pigments, to indicate to the program the approximate location of the objects that need to be tracked. This panel (as well as the following two) can be updated and refined along the video in case of tracking errors. The correct order of the user clicks, as well as the colors associated with each parapodia both on the graphic output and the resulting graphs, can be seen in the help section image (lower left corner of the GUI).

3. Adjust pigment parameters. The tracking of the pigments and eyes works by computing, for each frame, the local intensity contrast on the areas indicated by the user, or in the case of continuous tracking, the position of the object in the previous frame. The user can adjust both the contrast range and the local area until good segmentation of the pigments is achieved, as well as change it for different parts of the movie[8].

4. Find the edge. In order to find the head and tail to create a skeleton of the larvae, the contour of the body plan is calculated in a segmentation procedure. Three parameters can be adjusted by the user until the result is satisfactory. Together with

[8] Adjustments of the contrast parameters might be needed as the illumination might change depending on the recorded area or the relative position of the animal to the light source, for example.

the mid points of the eyes and pigments, the head and tail points are used to draw a skeletonized version of the larvae. Currently I have not implemented any measures or further implementations from this, so this step is not critical for the results presented here, but it can be useful for future modelling purposes (see Discussion section 4.3).
5. Run. Command panel used to control how many frames to process, to start the tracking, and stop it in case an error is detected in the output. As the program is running, it outputs two pieces of information. The first one, that will be used in the analysis part of PduLoco, is a text file containing, for each time point (frame of the movie) the coordinates for each tracked object, and for the head and tail. The second one is a graphical output for an easy screening of error (see next paragraph).
6. Refine. As the processing of the images and the tracking can be relatively slow (half a second per frame) and there are usually many frames in each video[9], processing a one minute movie can take half an hour in a regular computer. As staring at the screen looking at such a slow movie for that long is quite boring, the user can let the program run without supervision as long as no errors are detected. As part of the output, PduLoco stores every processed frame as an image containing several pieces of useful information, like the centered larvae, the tracking objects, the segmentation of the pigments and eyes, the frame number and the contour of the body (see Figure 3.38). The user can then play a faster version of the processed movie (real time for example), and identify any errors very efficiently. This panel can then be used to reset the tracking from that point onwards.

PduLoco III: Identification of crawling strokes

After the tracking has been performed, the user can explore the raw results using this panel. In order to have a straight-forward representation of the movements, the relative distances of the four pigments to the head[10] are plotted over time. As can be seen in Figure 3.37 III, the data is normalized for each pigment so the four tracks can be represented as non-overlapping curves to illustrate the time relationship between the parapodia strokes. The user can then quickly explore these curves in order to select, using the window located in the bottom of this panel, those time ranges characterized by a continuous non-disrupted crawling[11] in order to restrict the analysis to these cases.

The goal of this step is to identify, on the different curves, the time points corresponding to the start and the end of each forward stroke. To do that, the local maxima and minima are computed from the curves, using a window size specified by the user interactively using a slider in the GUI and a live-updated graph (small lines below the curves). These local maxima and minima correspond, respectively,

[9] Videos contain between 50 and 60 frames per second.
[10] An invariant point is chosen, in this case the middle point between the two eyes.
[11] Sometimes animals stop, change direction, or even crawl backwards.

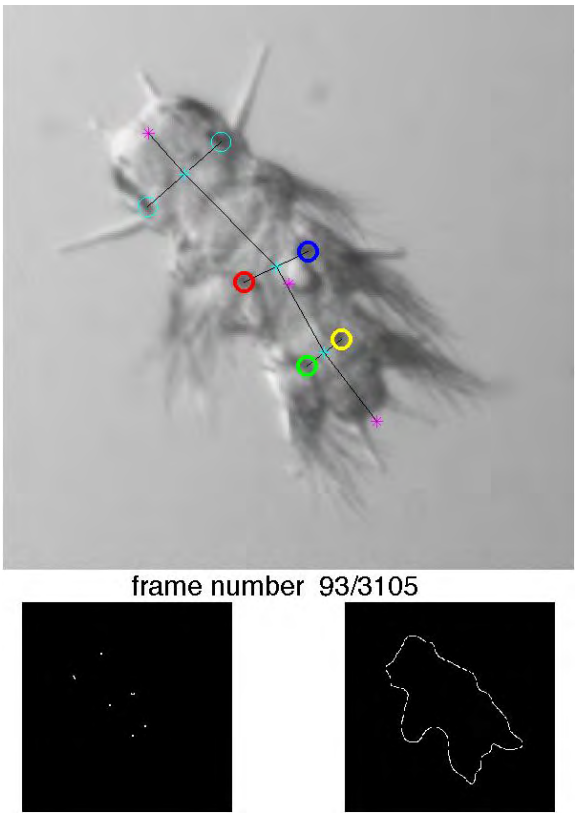

FIGURE 3.38: Graphical output of PduLoco tracking.
PduLoco outputs a video (one image per analyzed frame) with the raw result of the tracking, as well as the result of two important channels (the contrast of the pigments, and the outline of the larvae).

to the most distant and closest point of the pigment to the head, or in other words, the start and end of the forward movement of the stroke. As the parapodia move forward and backwards, I refer to these phases of the stroke as **protraction** and **retraction**, respectively. All the data that I am presenting in the next sections uses only the information related to the start point of the protraction for each of the four most posterior parapodia.

PduLoco IV: Quantification of results

The last panel of PduLoco is very useful to screen through the results in an efficient way. The data from the previous panel (starting points of every stroke within the selected range), is analysed as temporal relationships between stroke events. These relationships are calculated both for data within each parapodia, as well as across different parapodia. The user can always change the parameters of the previous panel to see how that changes the final results, in a dynamic way. Several types of plots representing different aspects of the locomotor behaviour can be represented in the graphical panel just by clicking the buttons below, and are easily exported if desired (see Figure 3.39 for examples of these plots). I will explain these plots here, and also refer to some of them in the following sections. The same colour-code related to the identity of each parapodia, similarly as in PduLoco GUI, is used to plot all the data.

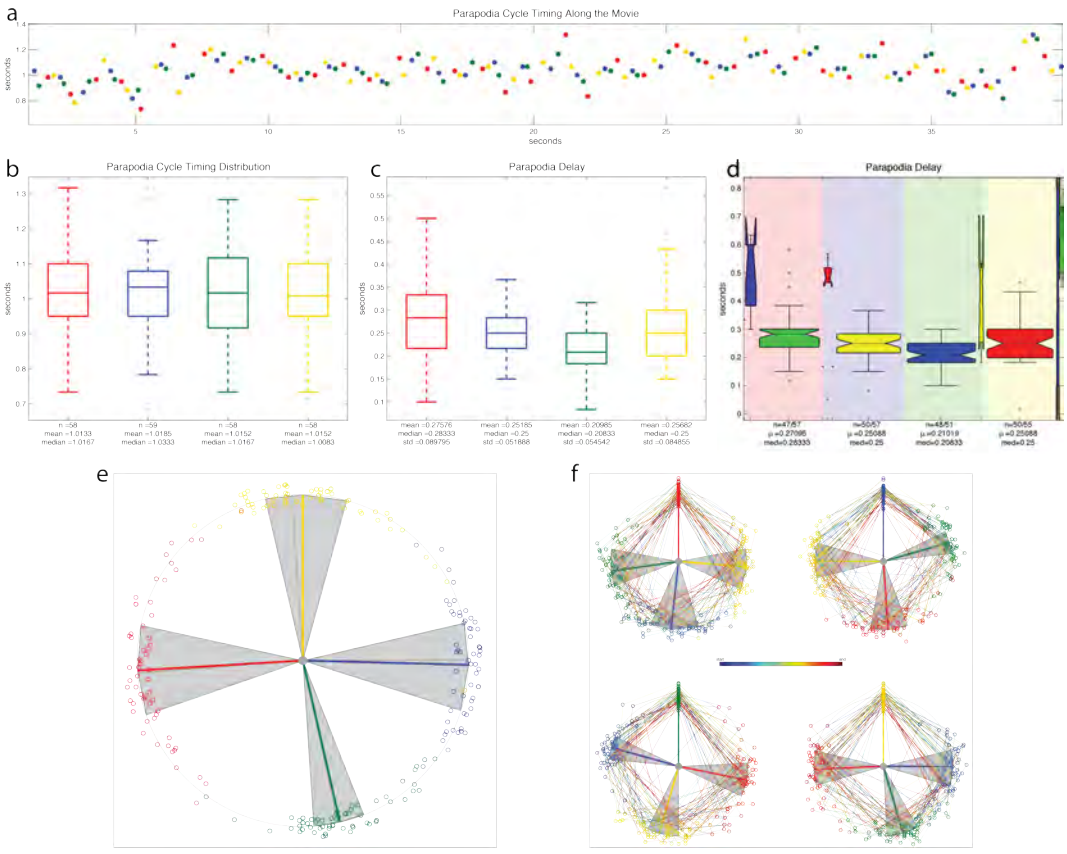

FIGURE 3.39: Plots generated by PduLoco as results of the tracking analysis. **(a)** Representation of stroke occurrences over time. **(b)** Time distributions of cycles duration. **(c)** Time distributions of normal cycles steps duration. **(d)** Quantification of the step order. **(e)** Distributions of relative time relationships between cycle steps. **(f)** Distributions and time track of relative time relationships between cycle steps. Plots explained in the main text.

Stroke occurrences over time. Figure 3.39a. This plot represents all the occurrences for the start of each protraction movement. Points are plotted along the x axis depending on which time point during the movie a particular protraction (corresponding to a specific parapodia) occurs. The position in the y axis of each point relates to the time difference with the occurrence of the previous protraction from the same parapodia (I refer to this difference from now on as a **cycle**). This way, the curve represented by all the points indicates the speed of crawling (eg. the curve will descend if the animal crawls faster).

Intra-parapodia delay distributions. Figure 3.39b. This plot shows, for each parapodia independently, the distribution of cycle lengths. The total number of cycles, as well as the mean and median values, are indicated below each boxplot.

Inter-parapodia delay distributions. Figure 3.39c. This plot is similar to the previous one, but instead of representing cycles it represents naturally occurring steps. I have defined a **step** as the period between two consecutive protractions, regardless of the parapodia that performs them. Naturally occurring steps refer to the normal alternation of parapodia (see Section 3.3.5). For instance, the red boxplot represents the distribution of time differences between the protractions of the "red parapodia" and the "green parapodia", as red immediately follows green. The mean, median and standard deviation values are indicated below each boxplot.

Quantification of parapodia alternations. Figure 3.39d. This plot is similar to the previous one, but represents a more quantitative and non-biased measure for the alternation of parapodia. That is, all the possible steps can be represented, and not only the naturally occurring ones. For each parapodia (represented by the graph background color), the steps are classified according to which of the parapodia moves immediately before (any of the other three and also itself). The distributions of the possible different steps are then plotted as boxplots. The width of the boxplots is proportional to their distribution sample size (in relation to the total number of occurrences). For instance, in the case of the "red parapodia" in the example plot (see boxplots over the red background), there are two possible steps. The naturally occurring step (as seen before green precedes red) is the most frequent. There are some instances where the "blue parapodia" is preceding the red, with a time delay twice as big as in the green case. Given the natural sequence of movements (see Section 3.3.5), this indicates that in these instances either the "green parapodia" skipped the stroke without affecting the overall step sequence delay, or that the program failed to identify the green strokes, or a combination of the two[12].

Strokes distributions over crawling cycles. Figure 3.39e. The crawling behaviour of *Platynereis* larvae is based on a repetition of a cycle, which is composed of a determined sequence of stroke movements or steps (see Section 3.3.5). In order to plot

[12] All the plots in Figure 3.39 have been obtained from the same example (same one as shown in Figure 3.37). To get plot **d** to show several boxplots for each step, I intentionally gave bad parameters in step **III**, so whereas all the other plots have been done with the same data, this particular one has a slightly different set, despite being based on the same tracked movie.

both the stroke sequence, as well as the variability of each step within the cycle, a circular coordinate system proves to be useful. To do this, we can normalize the cycles by their duration given that the speed of crawling does not affect the step sequence (see Section 3.3.6). The distributions are represented in gray and the stroke occurrences, as well as the distribution mean, in the color of each parapodia. One parapodia (the yellow one) is arbitrarily selected to set its mean to the origin and plot the other events relative to this one[13] in a clockwise fashion. The distance of each point to the center of the circle is relative to the speed in that particular cycle. Using this plot, we can get an idea about the phases (steps) of a cycle and their robustness. Note that in order to plot the data in this format, we have to normalize it by the mean cycle length (sum of the means of each step).

Strokes distributions over cycles along the movie. Figure 3.39f. These plots use a similar principle to the previous one, but the data is not normalized by the mean cycle length. Instead, the step phases are represented relative to a cycle exclusively defined using two consecutive strokes of the same parapodia (used also to set the origin). This way, we get a "parapodia-subjective" view of the locomotion, in terms of the different steps relative to the cycles defined by one parapodia only. This is useful because it allows us to also plot the sequence of movements (sequence of strokes) during the movie. Lines connecting the points represent this sequence, colour-coded according to the time point of the full video (scaled by the video total length).

3.3.3 PduLoco is a useful and efficient tracking tool

The development of PduLoco has been carried out side by side with the acquisition of initial behavioural videos, and both approaches have been adjusted to accommodate each other to achieve the best tracking results. This has resulted in the establishment of a robust software which reliably quantifies the crawling behaviour of *Platynereis* larvae in an efficient way. The tracking of a behavioural recording containing thousands of time points requires only a quick setup by the user to identify the tracking objects and adjust some parameters. In cases where the illumination during the recording does not result in a good contrast of the pigments[14], the tracking might need to be reset and pigments identified again from the time point where the tracking failed. This is nevertheless done in a time-efficient way using PduLoco, and I have only needed to perform three to five adjustments per movie.

The image analysis modules that PduLoco implement are simple and successful in order to reliably perform the tracking, as well as to calculate a skeleton for the body plan to hopefully, in future, model this behaviour. The implementation of an

[13] This selection does not have an effect on the quantification.

[14] The body of the animal can produce some shadows depending on the direction of the light, and the lipid droplets in the proximity of the pigments can interfere with their correct identification.

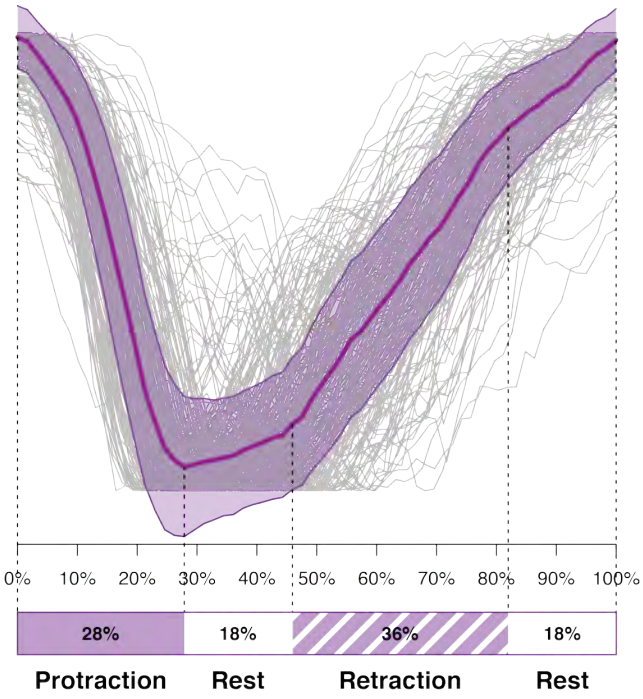

FIGURE 3.40: Analysis of *Platynereis* crawling stroke stages.
Canonical stroke movement cycle (dark purple line), averaged over 233 full strokes from the four most posterior parapodia of a 6dpf *Platynereis* larvae (grey lines). Shadowed area indicates the standard deviation. Cycles are defined as the period between two consecutive initiations of protraction movement. Curves are as in Figure 3.37 III, normalized by time and amplitude. Below the graph, illustration of the stroke stages and their relative cycle length.

integrated platform to directly report the result analysis is very helpful to both understand and identify the singularities of each video and to correct errors or refine the parameters if needed. The different plots made automatically by PduLoco can accurately describe several aspects of the crawling behaviour in *Platynereis*.

The initial intention for developing PduLoco followed two main purposes. First, to describe the natural phenotype of this animal, and the following sections will describe some aspects we have learned about *Platynereis* crawling by recording different wild type individuals. And second, to screen mutant larvae and identify specific crawling defects that could indicate a functional role for particular cell types (see Results section 3.4).

3.3.4 The stroke movement

The continuous monitoring of the relative position of parapodia with respect to the animal body plan allows for a detailed analysis of their movement during a full cycle (protraction, retraction, and time in between). In order to do this, I combined the measurements of 233 cycles (data from the four parapodia), and created an average stroke movement, normalizing by speed and amplitude (Figure 3.40). From this analysis, I have defined four stroke stages based on the speed of movement of the parapodia. As I have defined a cycle as the period between two consecutive protraction movements of the same parapodia, the first occurrence is therefore the protraction stage. This is also a natural starting point, as before any movements the muscles are relaxed, although during continuous crawling any reference point is valid as the stages are cyclic.

The protraction stage occurs when the parapodia moves towards the head of the animal, and positions itself perpendicular to the anterio-posterior axis of the body (from an initial parallel position). This stage occupies 28% of the total cycle length, and the shape of the speed curve is sigmoidal, meaning that there is an acceleration and a deceleration phase in this movement. It is followed by a resting period corresponding to 18% of the cycle, and followed by the retraction stage, which appears to have a linear function, meaning that the speed of retraction is constant. It is also considerably longer (38% of cycle length) than the protraction, meaning that the parapodia is recovering its initial position more slowly. Interestingly, there is also a resting period consuming 18% of the cycle length between the retraction and the protraction (as it is the other way around).

Note that these curves are representing the total distance between the pigments and the middle of the head, and this is the result of both local movements of the parapodia, and general bending of the body affecting the relative orientation of segments. This probably has a bigger effect on the resting period before the protraction and that is why the distance seems to increase. I speculate on the muscles involved in these movements, as well as possible explanations for these stages in the Discussion section 4.3.

3.3.5 Alternation of parapodia

Resolving the phase shift of stroke movement stages between the different parapodia can be very instructive about the neural circuit architecture controlling the inter-parapodia coordination. As seen in Figure 3.39a, a simple plot of stroke occurrences over time already shows that the movement of the parapodia in the second segment immediately follows the movement of the posterior parapodia of the same side. Also, the posterior parapodia follow the movement of the anterior parapodia of the contralateral side. The quantification of the stage shift is better illustrated in Figure 3.39d, and the complete cycle step alternation can be best appreciated following the Figure 3.39e in a clockwise fashion. I have made a schematic (Figure 3.41a) explaining the progression of steps within a cycle, and given that the different measures performed with PduLoco (see Figures 3.39b and 3.40) indicate that the four parapodia perform comparable cycles, I have overlapped the stroke stages of different parapodia (Figure 3.41b) to illustrate their temporal phase shift.

This can be interpreted as having one caudo-rostral wave of locomotion per side, in which every four consecutive parapodia are in a different stage of the cycle. Between the two sides, this wave is completely antagonized, so within each segment the two parapodia are in opposite stages of the stroke movement. The different points I have selected to represent the steps of the canonical crawling cycle are based on the animal morphology, which probably indicates complete muscle contraction and extension. It is yet unknown whether the phases of neuronal activity

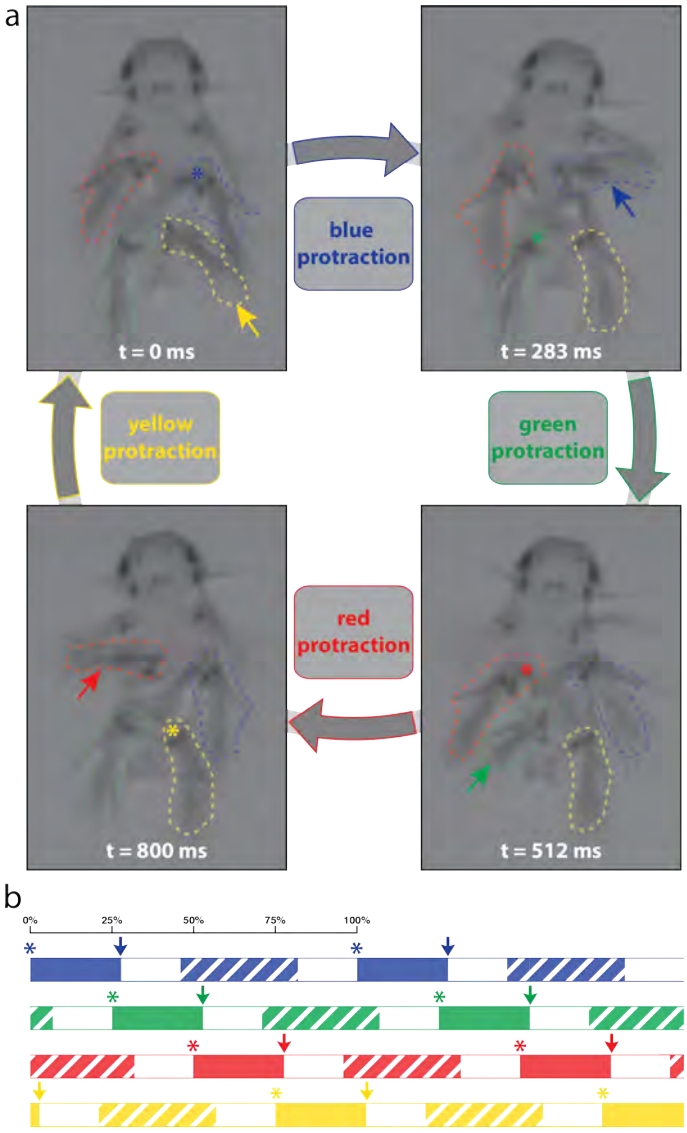

FIGURE 3.41: Steps of the crawling cycle. **(a)** Illustration of the step sequence that composes a crawling cycle. Four frames of the movie that is being used as an example in this section are shown, corresponding to these points where, for each parapodia, the forward stroke (protraction movement), has just been completed. **(b)** Superposition of the stroke stages for each parapodia during two full cycles. Bars explained in Figure 3.40 and main text. The filled coloured arrows indicate the completion of a forward stroke. Asterisks indicate the parapodia that is contracting (moving forward) in the upcoming cycle. t, time; ms, miliseconds.

controlling these muscle movements overlap temporally with these steps, although it would be surprising if they did not. In the Discussion section 4.3 I interpret these results with regard to the structure of the neural circuits controlling these movements and coordinating the temporal contraction of the different muscles involved in this complex behaviour.

3.3.6 Crawling speed

Another aspect I looked into using PduLoco is the speed of crawling. As can be seen in Figure 3.39a, animals can modulate their speed continuously during a recording. The smooth curves produced by these graphs indicate that this modulation affects the entire system and not only individual parapodia. Analysis of the speed distributions during entire movies (see Figure 3.42a) shows that despite this continuous modulation, individuals show a preference for certain speed ranges, at least during short and medium-length (10-120 seconds) crawling periods. Also, although the total range of speeds seems to be quite broad, there is a general preference for speeds ranging in between 0.8 and 1.1 cycles per second.

Given this variety of speeds across and within individuals, I asked whether the speed of crawling influenced the coordination in between the parapodia. To check for this, I analysed the correlation between the speed of the cycles and the protraction delay for the four parapodia, and found a similar sequence for all the speeds, indicating that the the relative cycle steps scale with speed variation. This allowed me to generate the data presented in previous sections by scaling the steps according to movement speed and analysing the stroke phases relative to the total cycle length. Note that these results indicate that the phase shift between the different parapodia scales with speed, but it remains to be determined whether this is the case for all the stages within the stroke (see Discussion section 4.3).

Lastly, and despite the fact that there is no indication that the speed changes from 5 to 6 dpf (see Figure 3.42a), during this developmental time the coordination between the parapodia becomes more precise and the robustness of the cycles within each parapodia increases (Figure 3.42c)[15], probably corresponding to the reinforcement of already established synapses due to functional input and/or developmentally encoded processes.

[15]The shift of the "green parapodia" is due to the inclusion in the analysis of the animal shown as an example in this section, which shows an abnormality in the delay of this parapodia from the blue one (not seen in any other individual).

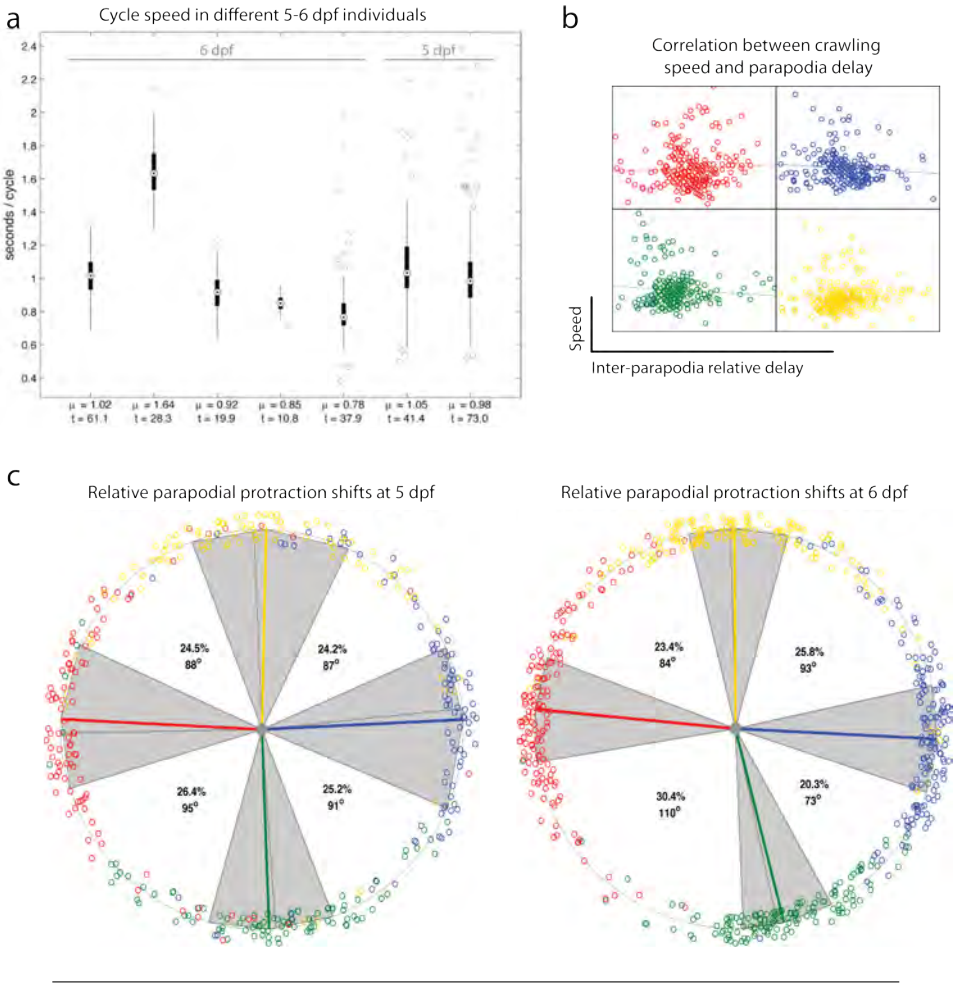

FIGURE 3.42: Analysis of the speed properties of crawling behaviour.
(a) Boxplots representing the distribution of crawling speed, as seconds per cycle, for individual animals at 5 and 6 dpf. Mean of the distributions and total time tracked per individual are shown below the boxplots. **(b)** Analysis for correlation between the crawling speed and the inter-parapodia delay time for the four parapodia using the data of the five 6 dpf animals shown in **a**. **(c)** Circular plots showing the distributions over a cycle of the start of protraction movements for the four parapodia, for 5 and 6 dpf animals. See Section 3.3.2 for details.

3.4 Functional relevance of *dbx1* cells in *Platynereis*

A powerful way to test for the conservation of locomotor-related cell types in *Bilateria* is to analyse the functional relevance of specific neuronal populations across different animals. Until very recently however, following this approach was a tedious task as genetic editing was very costly for non-model organisms, which, like in the case of *Platynereis*, offer a greater insight into evolution than established laboratory models. Fortunately, my early PhD period coincided with the development of the genetic editing technique Crispr-Cas9, which substituted the expensive Zinc-Finger and TALEN approaches with a cheap and fast RNA-based method to mutate very specific sequences in the genome.

Using Crispr-Cas9, I aimed to target the main transcription factors specifying the principal vertebrate locomotor CPG cell-types (see Section 2.2.2). This section describes the first (to my knowledge) use of this technique in *Platynereis*, and focuses on the description of the *dbx1* mutant phenotype. Results on the efficiency of Crispr-Cas9 for other important transcription factors are explained in the Appendix section A.3.

3.4.1 Crispr-Cas9 mediated targeting of the transcription factor *dbx1*

dbx1 is a transcription factor that in vertebrates specifies a mainly commissural population of cells involved in the control of the left-right coordination during locomotion (see Section 2.2.2). We decided to first target *dbx1* in order to test the efficiency of Crispr-Cas9 for two main reasons. First, the origin of commissures represents a very interesting evolutionary question as this neuronal population can be considered a novelty of bilateral animals (see Section 1.5). Second, during *Platynereis* VNC development the commissures are easily visible with a general staining against Ac-Tub, and any phenotype affecting them could be easily scored.

In order to aim for high efficiency *dbx1* mutants, I selected the target site of the gRNA following several criteria:

- Upstream of the homeobox region as close as possible to it (see Figure 3.43a for a schematic of *dbx1* locus and relative location of the gRNA used in this study). This ensures the mutation of the main function of the gene as a TF, and ideally prevents transcription of active gene transcripts from alternative start sites.
- Nucleotide sequence as specific as possible to avoid off-target effects in other places of the genome.
- Target site located in a region free of single nucleotide polymorphisms (SNPs), to make the gRNA efficiency reproducible across experiments.

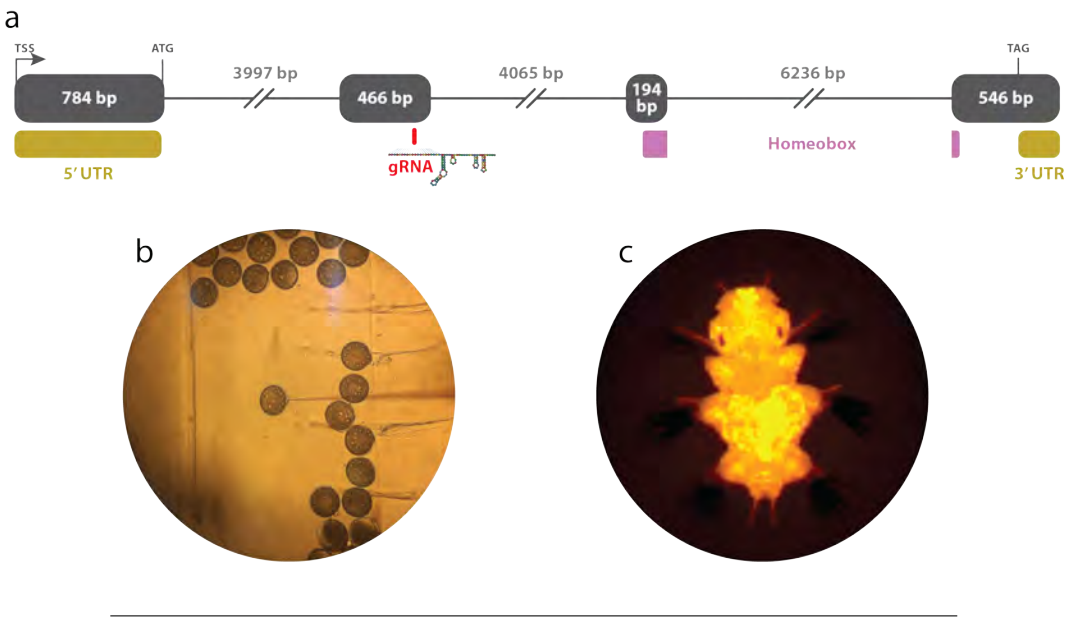

FIGURE 3.43: Design and injection of Crispr-Cas9 in *Platynereis*.
(a) Schematic of *dbx1* locus in *Platynereis*. The gene model was reconstructed from data from genome and transcriptome reads. Homeobox identified using NCBI Conserved Domains database. **(b)** Image of the injection procedure. **(c)** Representative example of dextran-fluorescent individual at 5dpf.

- Presence of a unique restriction enzyme site in the near sequence surroundings, to allow for a fast genotyping of worms (see scheme in Figure 3.44b).

See Materials and Methods section 6.8 for specifics on how to select the target site, design the oligonucleotides, perform the cloning, generate the gRNA and the Cas9 mRNA, and perform the microinjection procedure in *Platynereis* fertilized oocytes.

The gRNA (Dbx gRNA) was microinjected (Figure 3.43b) into the oocytes with mRNA coding for the Cas9 protein. Dextran fluorophore was added to the injection cocktail in order to screen for successfully injected larvae (Figure 3.43c). Only normal-looking animals showing strong dextran fluorescence were selected for analysis. Non or low fluorescent animals, and larvae strongly affected by the injection procedure were discarded.

3.4.2 Crispr-Cas9 is an efficient tool for genomic editing in *Platynereis*

In order to quantify the efficiency of the Crispr-Cas9 system in *Platynereis*, I extracted the genomic DNA of several injected larvae and performed a nested PCR of 500bp around the binding site of Dbx gRNA. I used several approaches to quantify the mutation rate, all depicted in Figure 3.44.

First, I ran the PCR product through a DNA BioAnalyzer to sample, at high resolution, the different sizes of the PCR products. As can be appreciated in Figure 3.44a, the use of Crispr-Cas9 to mutate the genome produces a wide range of possible

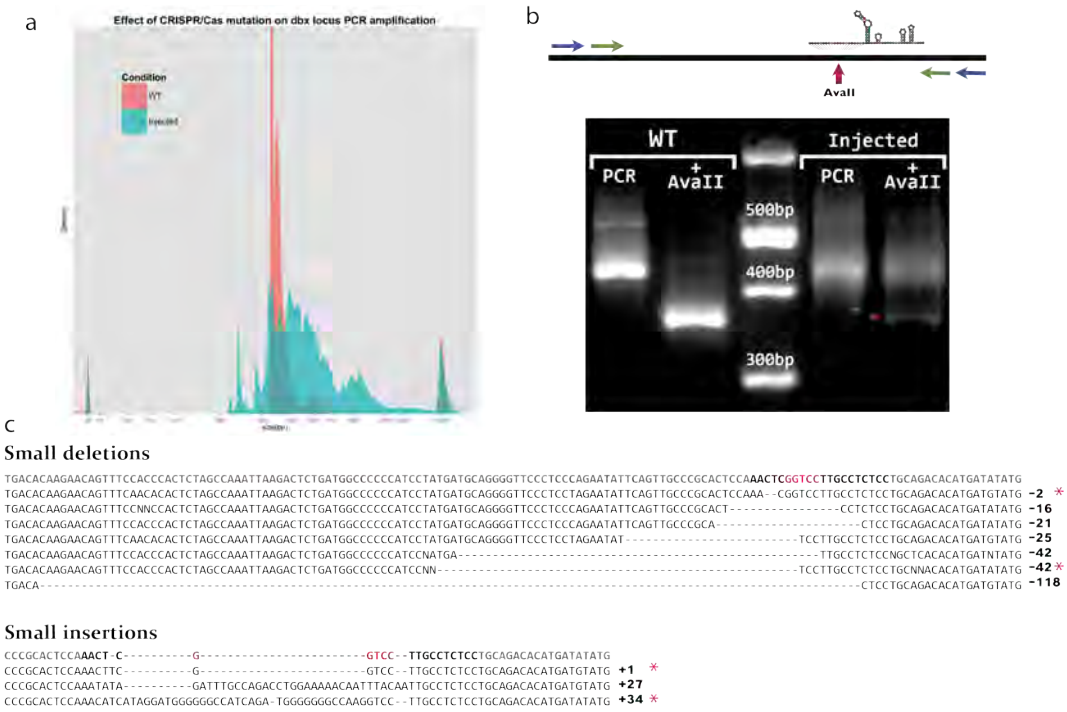

FIGURE 3.44: Effectiveness of Crispr-Cas9 in *Platynereis*.
(a) DNA BioAnalyzer size profile of the PCR product (green primers in **b**) of part of the *dbx1* locus containing the Dbx gRNA target site. **(b)** gRNA was designed to target a region containing a unique restriction enzyme site (AvaII) for a fast genotyping of the worms. In the scheme above, this is indicated by the red arrow. Blue and green primers indicate the position of primers used for the genomic and nested PCR, respectively. Gel shows, for WT and injected animals, the PCR product of the green primers and its restriction result using AvaII. Asterisk indicate possible WT sequences (see **c**). **(c)** Examples of small indels obtained by sequencing (WT in top row, gRNA binding sequence in bold). In red, the restriction site for AvaII. Asterisks indicate those sequences that despite being mutated still contain the AvaII site.

modifications, from deletions of 100bp or less to insertions up to 1kb. The shape of the size distribution suggests that the efficiency of mutation is very high, given that Crispr-Cas9 is supposed to be biased towards small indels.

The second method to screen the efficiency of the gRNA makes use of the unique restriction enzyme site explained in the previous section. Given that the Crispr-Cas9 system will preferentially cause nucleotide mutations in the gRNA-DNA binding region, the unique presence of a restriction site in such region can be exploited to estimate mutation efficiency by doing a simple PCR product digestion. Those PCR products resistant to the digestion would represent amplifications from Crispr-Cas9 mutated genomic regions, and a simple agarose gel can be used to quantify the results. Figure 3.44b shows the side by side comparison of the nested PCR and the AvaII (unique site for the Dbx gRNA) digestion products for genomic DNA, extracted from WT and injected larvae. Similarly as with the results from the BioAnalyzer (Figure 3.44a), the PCR product shows a widespread size distribution. Consistent with this, and unlike the WT case, the great majority of the PCR-amplified

"injected" sequences are resistant to the digestion with AvaII, although some digested product can be seen in the gel (asterisk in Figure 3.44b).

The third method used to quantify the results, as well as to identify in more detail the type of mutations produced by this approach, was to sequence individual PCR products (see Figure 3.44c). I detected a great variety of insertions and deletions in terms of size, all affecting, as expected, the gRNA binding site and its surroundings. In the case of the small deletions, they seem to spread towards the 3' end of the gene, although this can be caused by a PCR product bias due to the relative position of primers to the gRNA binding site. Interestingly, and unlike the case of the BioAnalyzer density plot, more deletions than insertions were detected. This can be caused, however, by a bias in the ligation procedure towards small fragments. This analysis also indicated that the restriction digestion is an underestimate of the mutation rate as in some mutated regions the enzyme binding site is still intact (sequences marked with an asterisk in Figure 3.44c).

Alltogether, these results indicate that Crispr-Cas9 is a very efficient genetic editing tool for *Platynereis*, and that Dbx gRNA can produce a mutation rate of more than 95% in the F0 population.

3.4.3 No off-targets detected with Dbx gRNA

One of the main concerns of the Crispr-Cas9 related approaches is the relatively high proportion of off-target effects given the inherent promiscuity of this bacterial immune system derived technology. To evaluate the possible off-target effects of Dbx gRNA, I identified the three places in the genome with stronger sequence similarity to the *dbx1* target site, and designed primers to perform a nested PCR and AvaII digestion as described in previous section. Figure 3.45 shows the sequence similarities of these sites and the results of the experiment. No digestion resistant sequence was found in any of these three genomic places. Off-target effects were also assessed by sequencing, and no mutant sequence was detected in any case (data not shown). Its important to note that these results can be biased by the selection of the gRNA[16], and a proper genome-wide analysis should be carried out to evaluate completely the off-target effects of this procedure.

Nevertheless, the complete absence of affected sequences in the three cases where they most likely would have happened (based on general sequence similarity), in particular given the high efficiency in the target site, made us very confident in that Dbx gRNA is very specific.

[16] As the gRNA is selected based on specificity across the genome, any part of the genome that is not represented in the database cannot be tested for.

dbx1 locus	AACTCGGTCCTTGCCTCTCC
Site_no.1	GTCATGGTCCTTGCCTCTCC
Site_no.2	GAGTGGGTCCTTGCCTCTCC
Site_no.3	TACTCGGTCCTTGCCTATTT

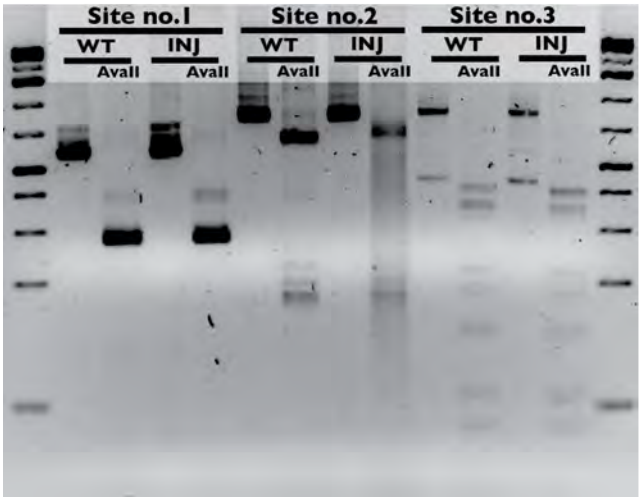

FIGURE 3.45: Off-target effects of *dbx1* gRNA in *Platynereis*.
On the top, the three places in the *Platynereis* genome that are most similar to the *dbx1* Crispr-Cas9 target sequence. Gel image showing the result of the nested PCR for each of the possible off-targets, and the restriction digestion with AvaII. WT, wild type; INJ, injected.

3.4.4 Commissural axons are reduced in *dbx1* mutant larvae

One of the reasons we chose to target *dbx1* locus was because this TF controls the specification of commissural neurons in vertebrates, and a phenotype affecting contralateral neuronal connections would be easy to score in *Platynereis* trochophores (see Section 3.4.1). To this end, we[17] fixed several *Platynereis* larvae at the stage of 50hpf, previously injected with the Dbx gRNA, Cas9 mRNA and dextran. We also fixed animals injected only with Cas9 mRNA and dextran from the same batches of larvae, as well as WT controls, and performed an AcTub immunostaining to label the nerve tracks.

We used confocal imaging at high resolution to image the VNC scaffold of the animals in the three groups (Figure 3.46a). In order to quantify the number of commissures, I designed an unbiased approach based on a straight-forward image analysis routine of the scans. The principle is shown in Figure 3.46 a and b and works as follows:

- Selection of a predefined[18] volume between the connectives covering the entire anterio-posterior span of the commissures.
- The volume was virtually scanned from side to side, calculating, for every pixel in the anterio-posterior direction, the projected sum of intensity along the ventro-dorsal axis.

[17] Some of these results have been generated with Ryan Prestil, an undergraduate student who worked under my supervision during the summer of 2015.
[18] Same frontal area was used for all individuals.

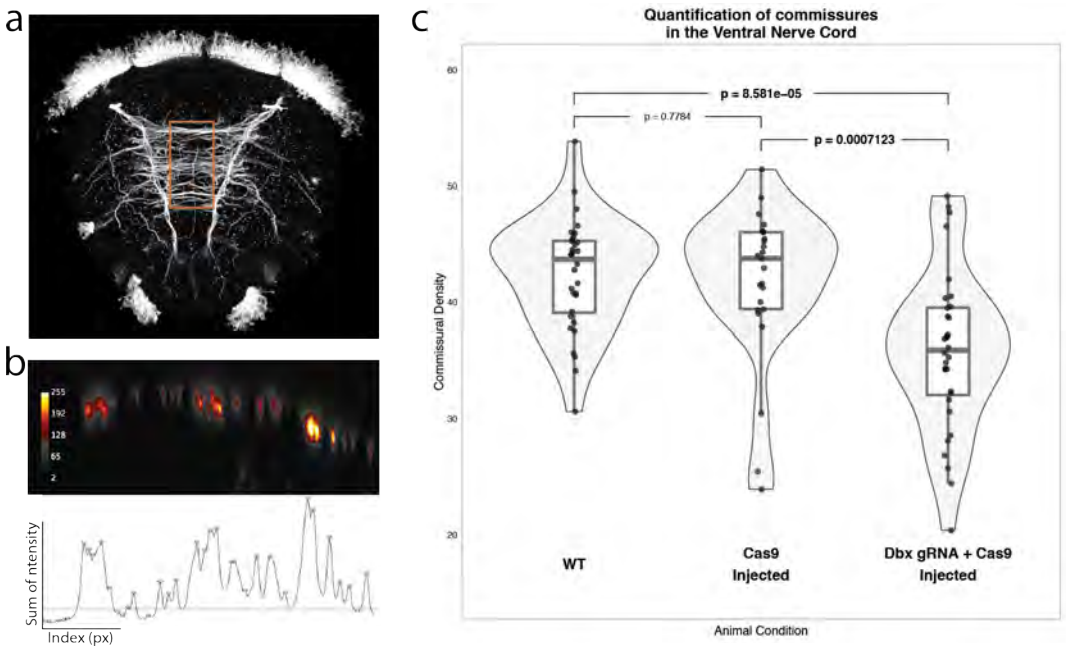

FIGURE 3.46: Quantification of commissures in the trochophore larvae.
(a) Maximum projection of the acetylated tubulin confocal scan of a WT trochophore larvae (50hpf), showing the VNC nerve tracks. **(b)** Scheme showing the principle of the image analysis routine to automatically determine the commissural density in the VNC. The nerve tracks of the VNC are virtually scanned in between the two lateral connectives to determine the intensity peaks (orange box in **a**). The analysis is done for every pixel from left to right, and a median value is obtained for every individual. **(c)** Plot showing the result of the image analysis for WT, Cas9 (only) injected, and *dbx1* mutant animals. Data plotted as violin plots, boxplots, and points indicating individual measures. p values from Wilcoxon test. [Data from Hernando Martínez and Ryan Prestil].

- The previous step generates an intensity profile from which the local maxima of the plot are quantified as commissural axons. This relative measure ensures that no bias is made to strongly or weakly stained animals[19].
- The two previous steps were performed several times along the left-right axis using a sliding window to reduce the noise, and a median value, representative of the total commissural density of the VNC, was calculated for each animal[20].

Figure 3.46c shows the result of the quantification. We pulled data from experiments performed at different injection sessions on different batches of larvae and different days as no difference was detected between them (data not shown), showing also the robustness of the detected phenotype and the quantification method. We detected a 20% reduction in the total number of commissures when *dbx1* was mutated using Crispr-Cas9. This phenotype, however, is not immediately obvious

[19]The AcTub staining protocol was performed side by side to all the groups of animals and no difference was detected between them in terms of intensity.

[20]No significant differences were detected within any group of animals between restricted sections of the frontal area (e.g.: area closed to the connectives vs. central area).

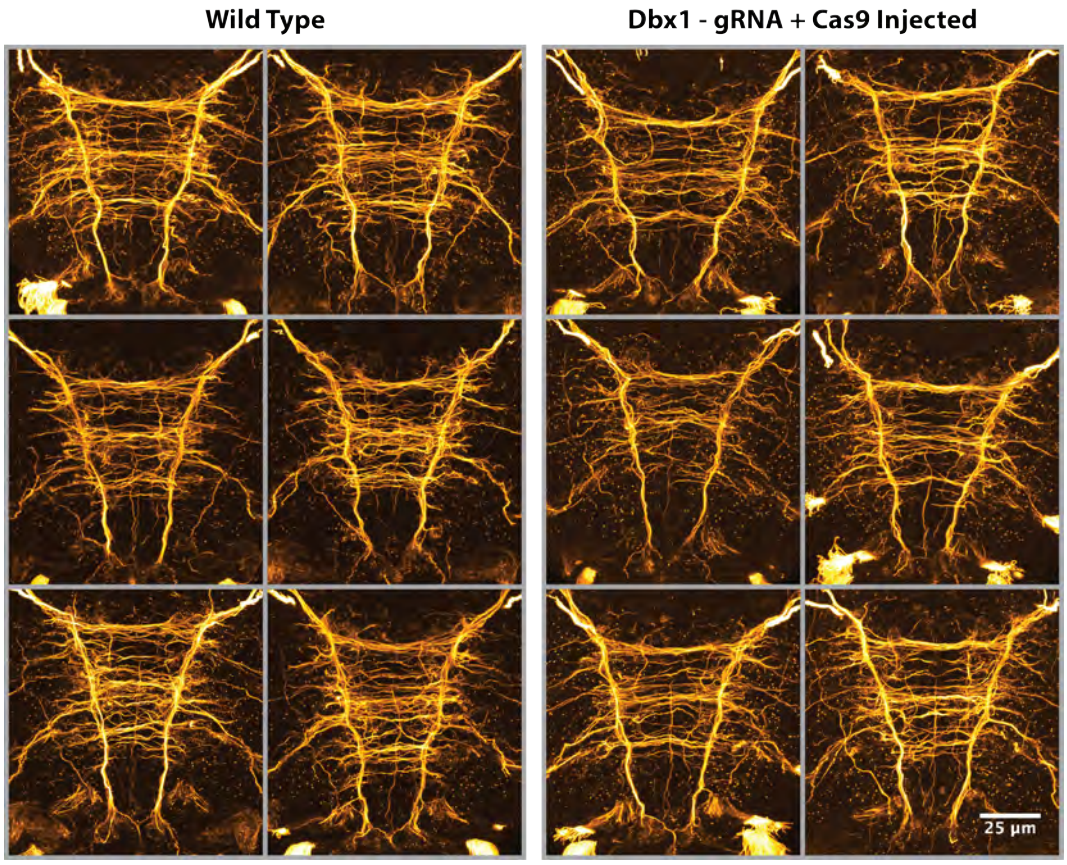

FIGURE 3.47: VNC axonal scaffold of WT and mutant 50hpf trochophore larvae. Panels showing the AcTub staining confocal scans (maximum projections) for the most representative (selected as those individuals with a value closer to the median calculations in Figure 3.46c) larvae of WT and *dbx1* mutant cases.

just by looking at the AcTub stainings (Figure 3.47 shows maximum projections of some representative data for the WT and mutant groups). This strengthens the relevance of using an unbiased image analysis approach to properly quantify this phenotype.

From this experiment, we can conclude that *Dbx1* mutants show a highly significant reduction in the commissural number that is not caused by the injection procedure itself or the exogenous action of the Cas9 protein.

3.4.5 *dbx1* mutant larvae do not show periods of crawling

In order to examine the behavioural phenotype of *dbx1* mutant larvae, I made use of PduLoco (see Section 3.3). Measuring the crawling parameters of the mutants this way was however, not straight forward for one main reason: *dbx1* mutants very rarely perform crawling movements. Out of 12 injected (Dbx1 gRNA and Cas9 mRNA) individuals for which more than 20 minutes of video were recorded in total

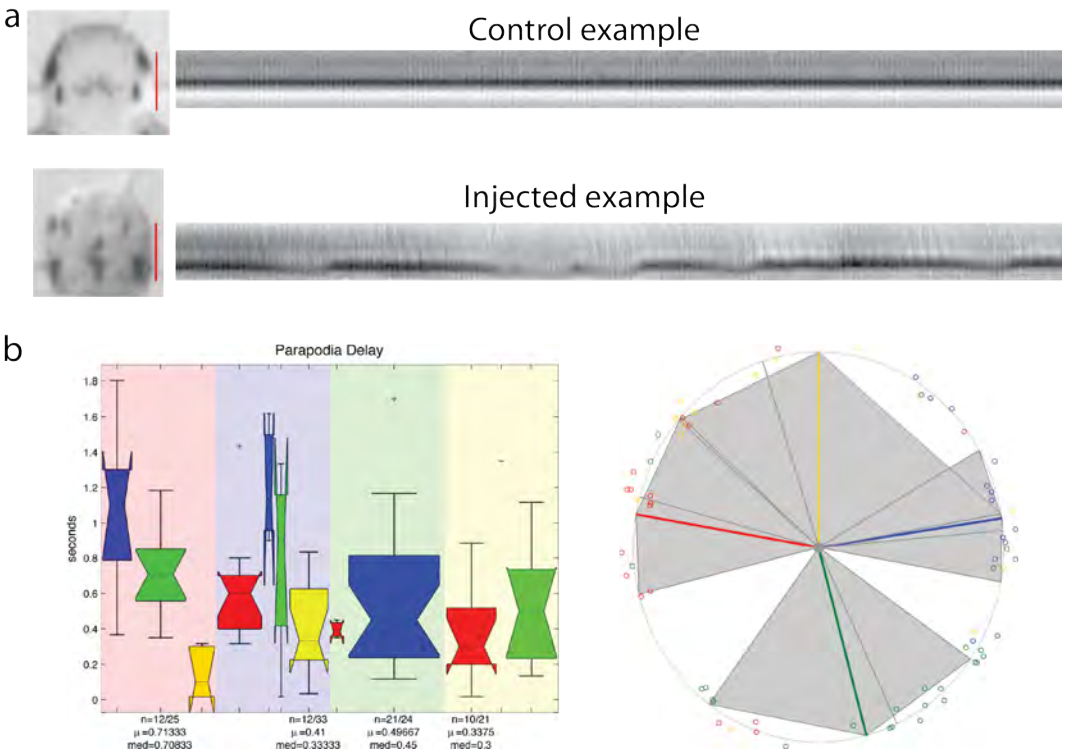

FIGURE 3.48: Behavioural observations on *dbx1* mutants.
(a) Frame and kymograph analysis of recordings of the ciliary movement in the head for control and mutant larvae. In red, the region used to create the kymograph. Duration of the movies 8.35 seconds. **(b)** PduLoco results of the only mutant individual that crawled during the recordings. For an explanation of the graphs see Section 3.3.

(as explained in Section 3.3.1), only one animal exhibited a continuous crawling period worth being analysed with PduLoco. Instead of crawling, animals were either swimming by cilia or immobile. Another difference between injected and control larvae was that, whereas wild-type larvae close and open the cilia depending on when they swim or rest, the cilia of injected animals were in constant beating mode (see Figure 3.48a). Out of the 12 observed individuals, the only animal that showed an attempt of crawling, did so in a highly uncoordinated way (see Figure 3.48b), exhibiting spasms in single parapodia, and what appeared to be an independence of the pair of parapodia on one side of the body with respect to the pair on the other side.

This data is not enough to conclude any significant defects caused by the absence of *dbx1*, and to properly quantify the defects in the locomotory behaviour of the *dbx1* mutants, two problems needed to be solved. First, a way to force the animals to attempt crawling, and second, given that the animals being recorded are from F0, it needs to be assessed whether the individual being analysed is indeed a *dbx1* mutant. The two following sections address these issues.

3.4.6 5-HT and NMDA activate a locomotory response in *Platynereis*

The external administration of excitatory molecules like serotonin and glutamate elicit fictive locomotor behaviours in nervous system preparations[21] across different animals in the bilaterian tree. I aimed to test whether the simple application of 5-HT (serotonin) and NMDA (glutamate) to the sea water could induce locomotor activity in *Platynereis* larvae. To do this, we continuously monitored the behaviour of 6dpf larvae before and after the application of the drugs, and quantified second by second, whether each individual was resting, crawling, or swimming. In order to properly diffuse the drug we pipetted the water up and down, and did the same with sea water for the controls to eliminate any current-related effects. Different concentrations of the drugs were tested, and the behaviour of ten or more individuals was recorded during one minute for each condition.

Figure 3.49 shows the results of the experiment. Animals normally spend most of the time in the resting state, lying on the substrate, exploring the near surroundings using their head appendages, but without performing continuous crawling or swimming. Proportionally with the increase in drug concentration, animals tend to be more active. Significant differences between the crawling-time before and after drug administration were detected using 500μM of 5-HT and NMDA. Higher drug concentrations were more effective in increasing the overall activity but animals were preferentially swimming, although this could be because this cohort of larvae was already pretty active in crawling before drug administration. Regardless, we settled to use the lower possible concentration of each drug (in this case 500μM) for future behavioural recordings.

3.4.7 Single-larvae genotyping

As it is yet not known when during *Platynereis* development the activity of the Cas9 takes place, it is unpredictable whether an injected larvae will be a complete mutant, or whether some cells might still contain one or two intact alleles of the targeted gene. The mutation level can be also dependent on which blastomeres receive more or less amount of the injected material, as the lineage of *Platynereis* is stereotypic (Ackermann, Dorresteijn, and Fischer, 2005). In order to assess, for each injected individual, its mutation level, we extracted the genomic DNA from single larvae right after acquiring the behavioural videos, and performed a digestion-based genotyping as explained before (see Section 3.4.2).

Figure 3.50a shows that the protocol for single-larvae genotyping performs very well, and that complete mutant individuals can be identified using this approach. Coupled with the *a priori* recording of behavioural videos, it is possible to analyse

[21] Spinal/Ventral-nerve cords are dissected, immersed in a saline bath, and activity is recorded from the nerves.

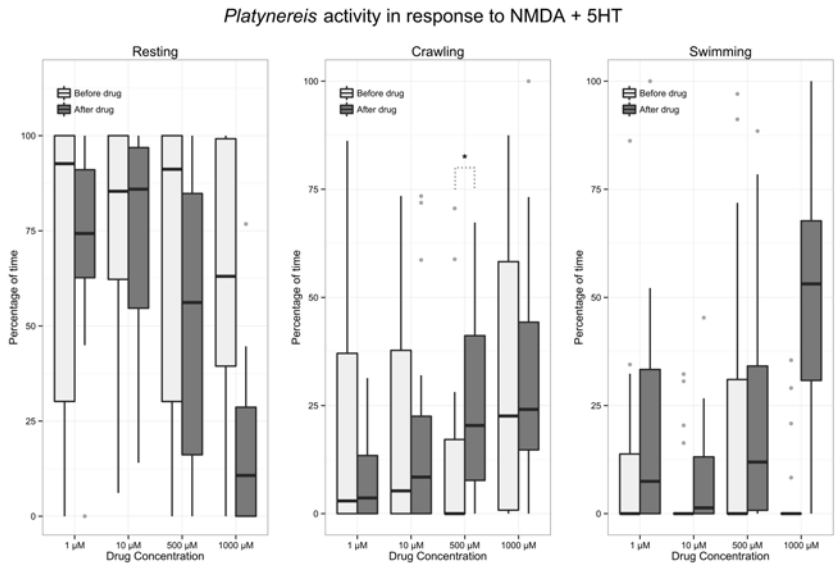

FIGURE 3.49: Effects of 5-HT and NMDA in *Platynereis* larvae locomotory response. Boxplots showing the differences between the locomotory behaviour of a cohort of 6dpf larvae before and after the administration of a cocktail of 5-HT and NMDA at different concentrations. **(*)** p-value of 0.03376 using Wilcoxon rank sum test with continuity correction. [Data by Ryan Prestil and Hernando Martinez].

those videos with PduLoco to quantify the locomotor-related behaviour of the mutants. Unfortunately, due to problems with the illumination during the recording of the videos, only the quality of one of the four mutant videos (see Figure 3.50a) was good enough for the tracking to work. The main results of the PduLoco analysis for this individual are shown in Figure 3.50b. This animal[22] showed a more normal behaviour that the one in Figure 3.48b, as the proper sequence of parapodia was maintained. However, there is a clear difference between the timing of movement within the parapodia in one side, and the parapodia in the other. The period that the contralateral parapodia (green and yellow) take to move after the previous protraction (blue and red respectively), is much shorter than the period between the parapodia in the second segment with respect to the third. This experiment is nevertheless planned to be repeated in the near future to collect more data, and these results, although promising with regards to a different ipsilateral and contralateral controlling *dbx1*-specified cell population, are still preliminary.

3.4.8 Signalling cues for midline crossing are conserved in centralized nervous systems

Complementary to the results explained above, I also followed a more general phylogenetic approach to analyse the commissural cells in *Bilateria*. I analysed the conservation of the binding sites of two important receptor-ligand pairs that have been

[22]Note that this animal is 8dpf, given that the tracking is easier because the pigments are more clear at older stages, as the lipid droplets, that sometimes cause shadows, disappear.

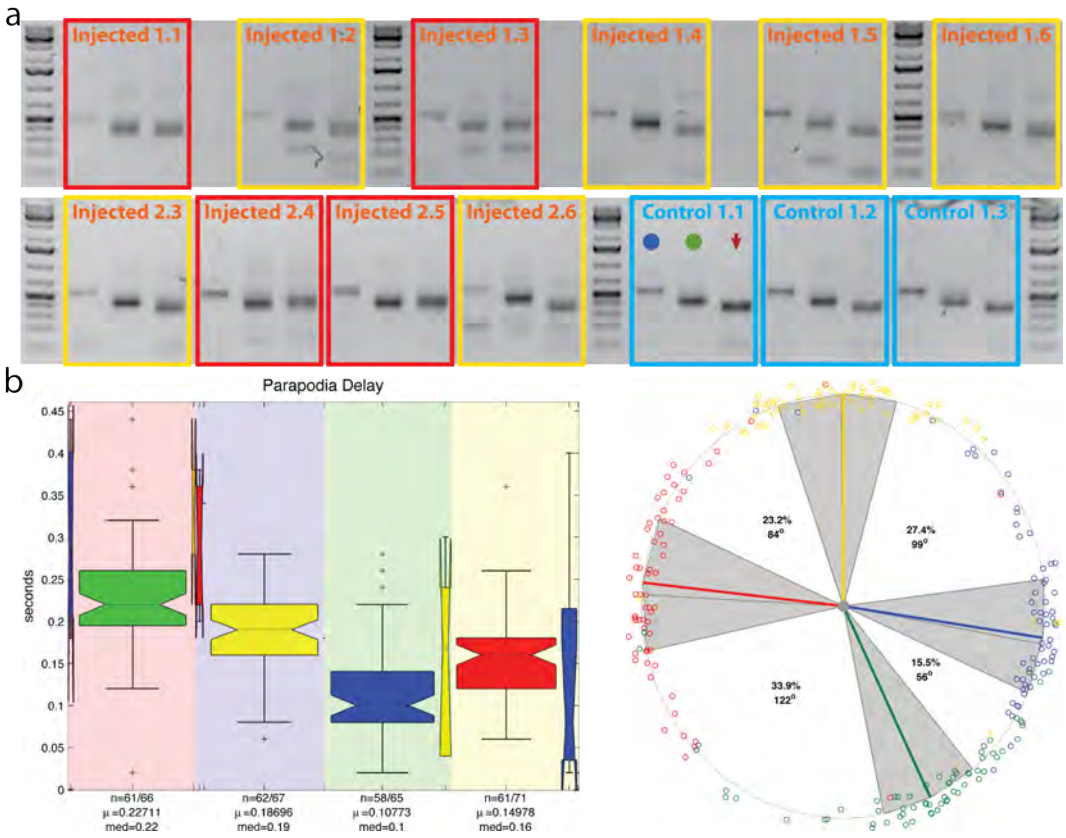

FIGURE 3.50: Genotyping and PduLoco analysis of single *Platynereis* mutant larvae. **(a)** Agarose gels showing some examples of the results of the genotyping of single animals. The strategy is similar to the one explained in Figure 3.44b. The squares contain the data for a single individual, either injected with *dbx1* gRNA and Cas9 mRNA, or non-injected WT larvae. Each square shows the genomic PCR band (blue circle in Control 1.1 and blue primers in Figure 3.44b scheme), the nested PCR (green circle and green primers), and the AvaII digested product (arrow). Red squares show high-efficient mutants. [Data by Ryan Prestil]. **(b)** PduLoco analysis results for animal "Injected 2.5" in **a**.

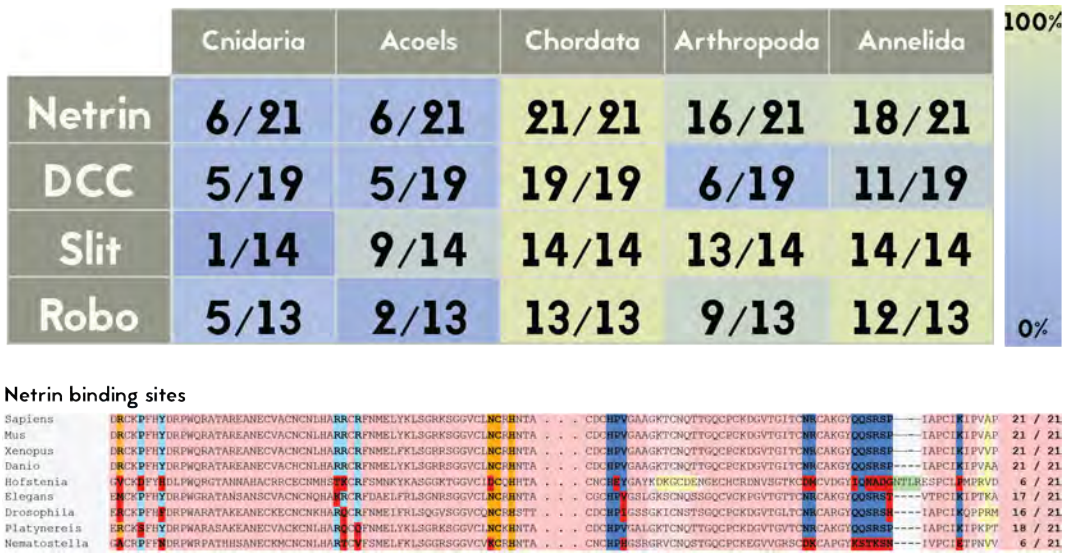

FIGURE 3.51: Comparison of binding site residues conservation in molecules regulating axonal midline crossing.
Quantification of the conservation of residues in the binding sites for the ligand-receptor pairs Netrin-DCC/Neogenin and Slit-Robo in five different metazoan phyla. Below, examples of the quantification procedure for Netrin and DCC/Neogenin. Protein sequence alignment performed using T-Coffee. [Data on the binding site sequences after Finci et al., 2014 and Morlot et al., 2007].

shown to control the commissural axon guidance in both vertebrates and *Drosophila* (see Appendix Section B). These receptor-ligand pairs are Neogenin/DCC-Netrin, and Robo-Slit, and their binding residues have been determined in vertebrates (Finci et al., 2014; Morlot et al., 2007). I found the orthologues of these four proteins in different animal species corresponding to five different phyla (Cnidaria, Acoels, Chordates, Arthropods and Annelids), and aligned them to find the conservation of the binding sites.

The quantification of the residues conserved with respect to Chordates (Figure 3.51) shows a high conservation in all of them (specially Robo-Slit) for animals with a centralized nervous system (all but Cnidaria and Acoels). Also, the similarities of annelids (*Platynereis*) with vertebrates is higher than arthropods and vertebrates. This indicates that the signalling cues that guide the commissural crossing are highly conserved in animals having a centralized nervous system.

Chapter 4

Discussion

In this chapter I discuss the relevance of the technical implementations I presented in the Results chapter, as well as potential improvements and future directions. Similarly, I reason about the implications of the biological findings unravelled using these techniques in what relates to *Platynereis* locomotion, molecular composition of the ventral nerve cord (VNC), and characteristics of *dbx1+* neurons.

4.1 A new method for generating expression atlases

In Section 3.1 I have demonstrated that gene expression atlases with cellular resolution can be achieved for animals with a complex and not highly stereotypic body plan. The method that I have developed in order to do so is called **ProSPr** (**Pro**filing by **S**ignal **Pr**obability mapping), and uses medium-throughput sampling and image analysis routines that are built upon previous implementations of image registration techniques (see PrImR in Section 2.3). The result is a highly automated and easy to customize method, available in free license software, to generate canonical gene expression maps, that are subdivided into expression regions for which quantitative information can be extracted.

Analysis of resolution

Previous implementations (Tomer et al., 2010; Asadulina et al., 2012) have relied on the position of the larval eyes to assess the resolution of PrImR, which are likely to develop in highly stereotypic invariable locations. To determine the precision of different methods in determining the canonical spatial location of expression signal, I used a unique cell located in the middle of neural tissue. This represents a "worst case scenario", and therefore it was expected (and confirmed), that the general performance of the method would be better for other cases. First, we must take into account that this analysis has relied on the gene *GAD*, which is a very difficult *in situ* signal to obtain, taking up to 8 days to develop at 6dpf larvae with the standard protocol. This means that the variability between the different samples

increases considerably. Second, this cell is embedded in a VNC ganglion, which consists of many packed cells without distinct DAPI landmarks to guide the registration, thus likely representing one of the tissues with more variability that will register less accurately. Lastly, the references used in the resolution analysis experiment were built from far less samples than the final one, probably causing a higher technical error during the registration of the samples used to calculate the resolution than the ones included in the final data set. The analysis used here therefore represents a high quality standard for determining the resolution of image registration methods.

High-quality of the atlas

Given the systematic approach that I undertook and the inclusion of dozens of different markers, the quality of the atlas[1] has been outstanding and ProSPr has performed beyond expectation in many different aspects. The high quality reference (see Results section 3.1.6) has allowed manual segmentation of tissues, thanks to the definition of structures defined by spatial coherence of cell nuclei. The molecular differences between these tissues is reflected on the identification and grouping of their constituent cells by unsupervised clustering methods. The two-step algorithm used to reconstruct the cells (see Results section 3.2.2) from the atlas imposes a high-quality filter on the data, as it first groups super-voxels just by expression, and then validates them spatially. Nevertheless, and for the majority of the tissues (see Table 3.2), a high percentage of the total volume was successfully grouped as being part of a cell. Is therefore reasonable to think that this process will ensure the recovery of clusters that most likely represent real biological cells (see also next section). It will be interesting to analyse the robustness of this "atlas cellularization process" once more markers are added to the atlas, although it is likely that both the quality (correlation and spatial coherence) of the recovered cells and the total volume reconstructed (if missing areas are covered by new genes) will improve.

One of the major reassurances for us about the quality of the data is the strong bilateral similarities seen along all the process. For instance, the recovered expression profiles for all the genes was highly bilateral, and so were the properties of the recovered minimal expression domains (MEDs). Moreover, visual inspection of reconstructed cells confirmed that these objects corresponded to groups of super-voxels representing two cells, one in each side of the body. Not all cells in *Platynereis* are bilaterally symmetrical. For these cases that we detected asymmetric expression, SPMs, MEDs and recovered cells faithfully reflected the biological case (see the example of MED number 6 in Figure 3.13). Similarly, we measured a high correlation between genes labelling the same cells (e.g. *sert* and *trpH*) with all the methods used during the analysis (e.g. Figure 3.26).

[1] Only the 6dpf data is discussed here. The 48hpf atlas is also outperforming previous versions.

Future implementations in ProSPr

The application of MEDs as a way to dissect, analyze, and threshold SPMs has several advantages as already mentioned in Section 3.1.5 (cellular resolution and process automatization), plus a big potential for the development of future applications of ProSPr. First, each MED is associated with a specific three-dimensional location, volume, and shape, as well as with the processing parameters that have been applied, including the individual measurements in every sample included in the analysis. It is possible then to use the MEDs to quantify and analyze variability in gene expression. In the cases where the expression of a gene is restricted to an isolated cell (see several examples in Figure 3.1.5), the resulting MED will correspond to one cell only, and therefore will be located in its canonical 3D position and will have its canonical volume. With enough similar cases for different cells, MEDs can be used to construct a cellular model, and if every cell is different from its neighbours in at least one marker, this can potentially lead to a coverage of the full animal body plan. This method of generating a cellular model will represent a better solution than the one used in this dissertation[2] because it contains the spatial position and volume of cells and it does not need general parameters like cell size to group super-voxels. Nevertheless, one potential validation of the cells reconstructed by the recursive partition clustering is to use the subset of small MEDs retrieved by the markers in the atlas to correlate the results, but I have not explored this possibility yet.

Potential improvements for ProSPr

Despite the improvement in the automatization of the process and rigorous unbiased analysis to represent gene expression in the atlas, the performance of the algorithms is not entirely customized to every possible scenario and some arbitrary choices still need to be made. For instance, the generation of a mask to binarize individual registered images (see Figure 3.19) is still somehow arbitrary to user selection. It might be possible to improve the method by feeding the MEDs data back into the binarization part and run this iteratively until an optimal solution is found, so all the genes will be processed unbiasedly. This is nevertheless not trivial and also does not guarantee a complete unbiased process, as it still depends on the arbitrary method of determining expression by the WMISH protocol, as it is up to the experimenter to determine the length and strength of the staining. It could be potentially avoided by having transgenic lines or antibodies to label endogenous proteins, but at this stage this approach is unfeasible for non-animal models if one is interested in hundreds of markers. The implementation of a completely unbiased method to model expression and the generation of a cellular model using the MEDs will allow to measure the error in expression assignment of objects (super-voxels

[2]The reason it has not been used is because of the lack of markers labelling specific cells.

or cells), and assign a binary value for expression (either the gene is expressed or not expressed) for all the cells in the atlas.

4.2 Molecular architecture of *Platynereis* VNC at 6dpf

Using an unsupervised recursive partition process with constrains based on *a priori* knowledge of the size of the cells, I grouped the sub-cellular super-voxels of *Platynereis* 6dpf ProSPr atlas into spatially coherent objects that are likely to represent biological units (cells). I performed this reconstruction process for different neural tissues in the animal, and defined the expression profile, using 137 genetic markers, for 4782 bilateral cells in the body plan of *Platyneris* nechtochaete (Results section 3.2). This dataset represents an unprecedented opportunity to compare the molecular fingerprint, in a system-wide manner, of cells within an animal. Similarly, this analysis can be extended to include other tissues, and applied to other developmental stages (see Results section 3.36). Last, the full methodology (see previous section) can be potentially applied to other organisms, which will allow unbiased comparisons to help reconstruct the history of cell types.

Motor-visceral duality of the neuroepithelium

Due to its relevance with regards to animal locomotion (see Section 2.1.4), I have focused my analysis on comparing and classifying the cells in the ventral nerve cord of *Platynereis* larvae. Using unsupervised clustering methods, the cells in the VNC grouped into discrete populations with a consistent ventro-dorsal and medio-lateral arrangement across the three different segments (Results section 3.2.4). The more pronounced subdivision is seen along the ventro-dorsal axis, separating a *phox2*+ region from a *pax6*+/*nk6*+/*hb9*+ one. This dichotomy is very similar to that seen between the alar and basal plate in the neural tube in vertebrates (Gray, 2008; Gray, 2013). It is therefore likely that these two domains in *Platynereis* are differentially involved in visceral and sensory versus motor functions. Investigation of the early neuroectoderm patterning by these markers (data not shown) shows a laterally located *phox2*+ region and a ventrally located *pax6*+/*nk6*+ region, which indicates that during development, part of the lateral plate relocates in the dorsal portion of the VNC, and part relocates lateral to the ventral longitudinal muscles in the peripheral nervous system (peripheral ganglia show strong expression of *phox2* at 6dpf). Consistent with this scenario, sensory markers like *brn3a* and *barh1* are also expressed in the dorsal ventral nerve cord and peripheral ganglia. This "folding" is, relatively to the body plan, inverse (lateral inside) to that of the vertebrate neural plate (medial inside). Given the dorso-ventral inversion that likely happened in the vertebrate lineage (Arendt and Nübler-Jung, 1999), the dorsal and

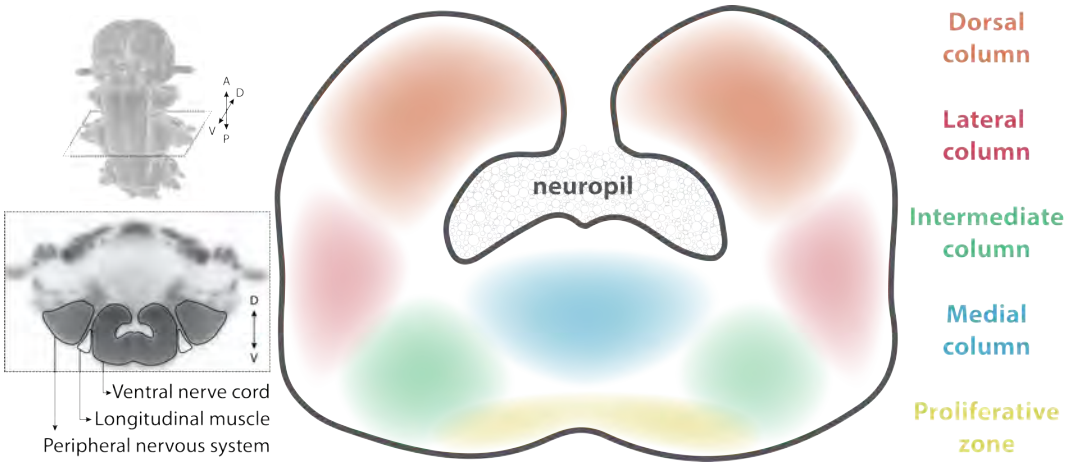

FIGURE 4.1: Summary of the structure of *Platynereis* VNC at 6dpf.
On the left, transversal section of the reference at the level of the second VNC ganglia, indicating the ventral structures. A, anterior; P, posterior; D, dorsal; V, ventral. On the right, depiction of the major genetically-defined territories in the VNC.

ventral regions of the VNC in *Platynereis* would correspond, by homology, to the dorsal and ventral (alar vs. basal) parts of the vertebrate neural tube, respectively.

Medio-lateral subdivision of the post-mitotic VNC

The subdivision of the ventral territory of the VNC is more relevant to discuss here due to the accumulation of markers associated with locomotor cells in vertebrates. The cellular clusters found in this analysis show a clear medio-lateral arrangement forming three major territories: medial, intermediate and lateral (Figure 4.1). Within the most medial column, strongly expressing *nk6*, we find a segmentally-repeated anterior *pax6*+ and posterior *pax6*- subdivision. This territory does not seem to have a molecular counterpart in the vertebrate nervous system, where *pax6* is not expressed in medial domains. Nevertheless, this post-developmental midline territory is very similar to that seen in *Drosophila* following the expression of *pax6* genes (Kammermeier et al., 2001). The most lateral domain expresses markers characteristics of the most dorsal part of the ventral neural tube (*dbx1*, *pax6* and *lhx1/5*), and is discussed in detail below (section 4.4).

The intermediate territory is a segmentally consistent group of cells, strongly expressing *tal1*, *tal2*, *lhx1/5* and *gata1/2/3*. In vertebrates, Tal2 is specific to the lateral floor plate (Pinheiro, Gering, and Patient, 2004), and Tal1, Gata2 and Lhx1/5 are distributed differently in the V3, MN and V2 domains, being important for the sub-specification of locomotor cell types (Francius et al., 2015). Consistent with the evolutionary relationship between the cell types composing the intermediate territory in *Platynereis* and the most ventral part of the neural tube, is the co-expression

of markers like *brn1/2/4*, *hnf6*, *lmx1*, *isl*, *nk6* and *pax6* (Gray, 2008) (see also Figure 2.3).

Evolutionary relationship of the intermediate territory and the CSF-cNs

Very interestingly, Tal1, Lhx1/5 and Gata2 markers define the cerebrospinal fluid-contacting neurons (CSF-cNs) in mice (Petracca et al., 2016), homologous to the Kolmer-Agduhr cells in zebrafish, for which recently a role in locomotion has been described (Wyart et al., 2009). CSF-cNs develop from the V3, MN and V2 territories in a late round of neurogenesis, remain in contact with the neural tube canal by an extension of the primary cilia, and provide a positive drive to the CPGs to initiate locomotion (Wyart et al., 2009). This is similar to the role of the V2a cells and the reticulospinal neurons (see Section 2.2.2), but in contrast to V2a, CSF-cNs do not express *chx10* (Petracca et al., 2016). Interestingly, *chx10*+ cells are located adjacent of the *gata1/2/3* domain in *Platynereis* at 6dpf, but the developmental pattern of these genes has not been investigated yet.

CSF-cNs are also different from V2b in that they do not express FoxN4, which together with Lmo4, Ascl1, Tal1, Gata2 and NLI promote the formation of the V2b interneuron type at the expense of the Lhx3-directed differentiation of V2a (Barrio et al., 2007; Joshi et al., 2009). Lhx3 has also been shown, together with Lhx4, to suppress Kolmer-Agduhr development and to regulate the developmental interplay between interneurons and motorneurons of the V2 and MN domains (Seredick et al., 2014). Neither *lmo4* or *lhx3* are expressed in the intermediate domain of *Platynereis* VNC at 6dpf, indicating a higher affinity of cells in this territory with CSF-cNs than with vertebrate-specific subpopulations. Additionally, Kolmer-Agduhr cells are transiently serotonergic until the formation of another stable serotonin-expressing cell population (Montgomery et al., 2016), and serotonergic cells are prominently found within the intermediate domain in the first segment.

Evolutionary origin of locomotor cell types

The unbiased classification of *Platynereis* VNC subdivides this structure into mediolateral regions that are very similar in molecular identity to main ventro-dorsal subdivisions of the vertebrate neural tube. However, whereas clearly correspondent locomotor cell types have not been identified within these regions, the data suggests a general "combination of identities" of today's vertebrate subpopulations, such as described above for the V3, MN and V2 domains. This indicates a distinct diversification of related cell types in both lineages, which is consistent with the increase in cell type diversity detected in the vertebrate lineage (Goulding, 2009; Grillner and Jessell, 2009) and the less apparent genetic repertoire complexity as

we move down the tree (mice, zebrafish, *Xenopus*, lamprey). Guided by the data, we[3] noticed the molecular similarities of the intermediate domain cells with the CSF-cNs in vertebrates. Given these similarities, and the presence of several ciliated cells in the ventral surface of *Platynereis* VNC, where this region is located, it is tentative to speculate that the CSF-cNs[4] cells derive from an ancient cell type that was sensing the environment, connecting to the motorneurons (or CPG) to elicit a proper locomotory response.

Potentials and limitations of the analysis

The analysis presented here has been restricted to a major overview of the VNC at 6dpf. However, it is important to note that ProSPr atlases (see Results section 3.36) provide the opportunity to navigate gene expression in the complete animal and guide, unbiasedly, more detailed questions regarding the evolution and development of cell types in other parts of the body and at different stages. Similarly, the developmental and functional characterization of specific cell types within the general territories can reveal more refined evolutionary relationships across animals (see Discussion section 4.4).

As a word of caution, there are regions in the VNC that are not covered by specific markers, and therefore the molecular fingerprint and evolutionary relationships of their constituent cells remains obscure. The addition of more markers to the atlas should help characterize those territories. Furthermore, the use of other clustering and data exploration tools could potentially retrieve other information from the data in terms of cellular relationships. The incorporation of new markers can potentially change the composition and interrelationship between the clusters. Nevertheless, adding the genes identified in the transcriptomics analysis (see Section 3.2.1) to the 75 markers used here to classify the VNC cells, does not change the identity of the clusters (data not shown), which is indicative of the robustness of the analysis.

Finally, these comparisons will benefit tremendously from the incorporation and integration of other parallel system-level approaches that I describe in this dissertation (see Discussion section 4.5). For instance, the reconstruction of the entire animal connectome (see Appendix section A.1) will allow the morphological classification of the cells within the intermediate column and their comparison with vertebrate CSF-cNs, as well as the similarities and differences between *Platynereis* circuits and the CPGs described in other animals.

[3]The interpretation of the data was done together with my supervisor Detlev Arendt.

[4]The folding of the neural plate would have repositioned the cilia of these cells to the inner canal.

4.3 PduLoco and *Platynereis* crawling behaviour

Muscular units controlling crawling

The monitoring and analysis of the crawling behaviour of *Platynereis* larvae using PduLoco has resulted in a detailed description of the temporal relationship between the sequence of movements of the trunk appendages (parapodia). These results can help understanding the muscular and neural basis of this behaviour. Observing the anatomy of the larval muscular system (Figure 4.2 a and b), it seems very likely that the main muscles involved in the stroke movement are the oblique muscles. The oblique muscles are attached to the axocord (Lauri et al., 2014), a ventromedial muscle located dorsally to the VNC. The contraction of the anterior oblique muscle would produce the protraction movement (forward stroke), and the antagonistic contraction of the posterior oblique muscle would produce the retraction movement (backward stroke). This is supported by the GCaMP signal seen in the location of the oblique muscles during crawling (see Appendix section A.2).

The difference in the speed and duration of these two movements (see Figure 3.40) is likely due to the fact that retraction is the phase that pushes the animal forward, and therefore offers more resistance, whereas protraction needs to be fast to position the parapodia for the next retraction movement. It is therefore likely that the parapodial muscles lift the parapodia during protraction to avoid any friction with the substrate. The role of these and other muscles potentially involved in locomotion (ventral longitudinal muscles and axocord) has not been included in this analysis, but the skeleton calculated by PduLoco (see Results section 3.3.2) can be used as a starting point to model in more detail the locomotion of *Platynereis*.

Similarities in locomotion kinematics in protostomes

Interestingly, walking movements in arthropods (Orlovsky, Deliagina, and Grillner, 1999; Büschges et al., 2008) show similar kinematics than the ones described here for *Platynereis*. The arthropod's leg movement (e.g. crayfish and stick insect) shows two main phases: the swing phase and the stance phase. During the swing phase, levator muscles lift the leg and protractor muscles position it in its most forward state, in a movement that follows a sigmoidal speed curve (see Results section 3.3.4). This is followed by a slower and linear backward movement of the leg controlled by retractor muscles and corresponding to the stance phase. In arthropods, the duration of the stance phase is strongly affected by the speed of movement, whereas the duration of the swing phase varies considerably less. This fact has not been assessed here but, despite the range of *Platynereis* speeds (at least in the larval stages) does not seem as broad as in arthropods, it is possible to use PduLoco to analyse stroke curves (see Figure 3.40) for different speeds and check if there are

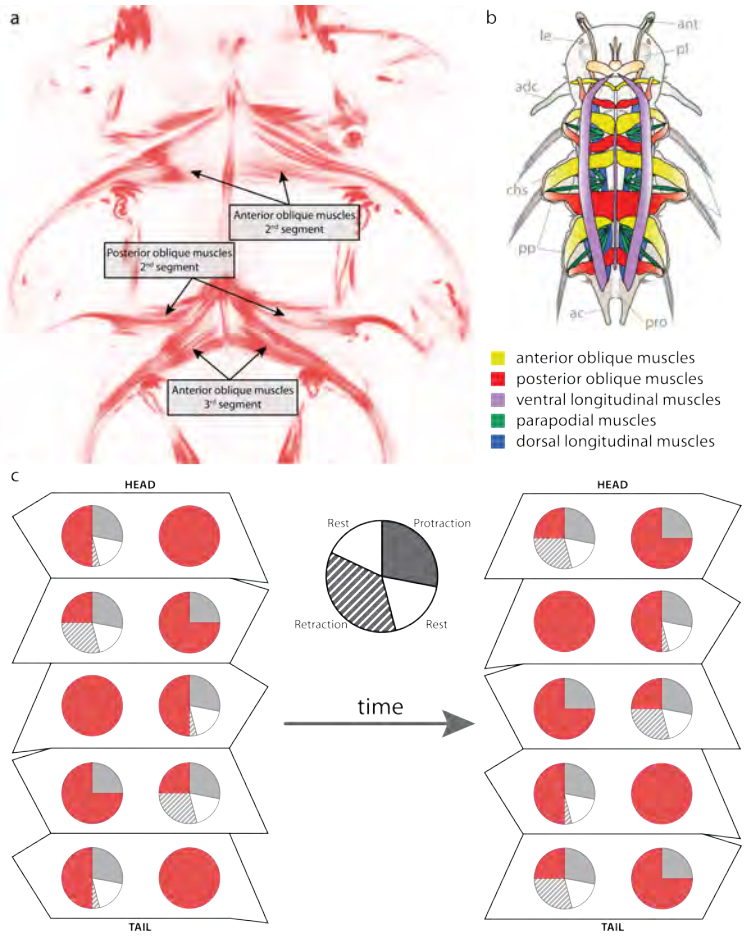

FIGURE 4.2: Muscular and neural basis of locomotion. **(a)** Confocal scan of phalloidin-stained 5dpf larvae labelling the muscles. **(b)** Diagram of muscles in the 5dpf larvae (ventral view). [Image from Fischer, 2010]. **(c)** Schematic of phase shift between segments. Explanation in the main text.

also similarities in this regard. Finally, it will be extremely interesting, once the interneurons of the *Platynereis* CPG are identified, to compare them with the structure of the CPGs controlling the stance and swing movements in arthropods (Orlovsky, Deliagina, and Grillner, 1999).

Direction of locomotor waves

The posterior to anterior sequence of movement shown by *Platynereis* differs considerably with that seen in other animals with undulatory coordination like swimming vertebrates (zebrafish, *Xenopus* and lamprey) (Grillner and Jessell, 2009) and even other annelids like the leech (Kristan, Calabrese, and Friesen, 2005) (anterior to posterior direction), and is similar to the movement of swimmerets in crustaceans[5] (Mulloney and Hall, 2007; Smarandache, Hall, and Mulloney, 2009; Smarandache-Wellmann, Weller, and Mulloney, 2014). This is nevertheless not very relevant to conclude any differences with regards to the evolutionary relationship between the neurons composing these CPGs, as many of this animals can also reverse the sequence using the same neurons and circuits. It is also interesting to note that

[5]Note that both the leech and the crustaceans swimmerets show a left-right synchrony during locomotion.

Platynereis adult worms change dramatically the direction of the locomotor wave when they transition from asexual atokes (poterior to anterior) to matured epitokes (anterior to posterior)[6]. This transition occurs once, seems instantaneous, and is never reversed.

Phase relationship between appendages during crawling

From the analysis of the phase relationships between the four most posterior parapodia in the 6dpf larvae, two observations can be made. First, the phase of the anterior segment follows the phase of the posterior one, first in one side of the body, and then in the other. Second, this alternation produces a representation of one complete cycle by the four parapodia in the two segments. Given these results, an interesting question arises: In an animal with more than two crawling segments (larvae at older stages), are the phases between parapodia modified to accommodate for segment number? This would result in every parapodia being in a different stroke stage and therefore representing one full cycle along the full body. Alternatively, are the new segments added to the established phase shift? The alternation between parapodia would be then maintained and more complete cycles would take place along the trunk length. The first scenario appears to be unlikely, given that the structure of the entire circuit would have to rearrange for every addition of a new segment. The other possibility supporting the first scenario is that there is only one command center controlling all the oblique muscles and that the completion of a cycle is fed back into this command center so a new cycle can start. Nevertheless, this is probably not true as short pieces of trunk dissected from the body can crawl (data not shown).

I will explain why the second scenario seems more likely. The coordination between the four most posterior parapodia of the 6dpf larvae can be explained as two posterior to anterior waves of locomotion, one per side. Between the two sides at any segment, these waves are in anti-phase, and within each side, a complete cycle would span four segments. Therefore, every fifth segment the pattern of both parapodia would be in synchrony, and every parapodia will be at the same stroke stage with the contralateral parapodia within a two-segment distance (Figure 4.2c). I tested and confirmed this hypothesis by recording locomotion in older animals, whose trunk is composed by more segments, and monitoring the synchrony of the parapodia located two segments apart (data not shown). Despite this has been confirmed using recordings at high temporal resolution and playing the videos in slow motion, it would be interesting to use PduLoco and make a properly quantified analysis across larval development.

[6]The amplitude of the bending of the body during the lateral undulations differs considerably between the two modes of locomotion.

Conceptual organization of *Platynereis* locomotor CPGs

The results I have obtained indicate that the crawling movement of the larvae is controlled by multiple CPGs (or Unit Burst Generators), probably two per segment[7], coupled anterio-posteriorly and contralaterally to coordinate the movements between the parapodia, and ensure a proper transition of the locomotory waves. Besides the observations discussed here and the results in section 3.3, this possibility is supported by several additional observations:

- Repeated (segmental) structure of the trunk nervous system into discrete ganglia.
- Stronger molecular similarity between the second and third segments than with the first segment, which is not involved in locomotion.
- Parts of the trunk, and even individual segments, can perform periodic alternation of parapodia if physically separated from the rest of the body.
- The structure of the nervous system in other segmentally-organized animals (protostomes and deuterostomes) follows the "chain of CPGs" model (McCrea and Rybak, 2008; Guertin, 2009).

It can be then considered that each parapodia is controlled by repetitions of the same neuronal circuits, and that the modulation of the speed can occur via a brain-directed tonic drive affecting all CPGs, and/or by neuromodulation via neuropeptides. It will be interesting to see if these observations are supported by the discovery of identical circuits once we have access to the connectome of the animal (see Appendix section A.1). Similarly, we will be able to compare the intra and interconnectivity of these circuits with the structure of other well characterized locomotor CPGs, like the crayfish swimmerets (Mulloney and Hall, 2007; Smarandache, Hall, and Mulloney, 2009; Smarandache-Wellmann, Weller, and Mulloney, 2014), the leech (Cacciatore, Rozenshteyn, and Kristan, 2000; Cang and Friesen, 2002; Friesen and Kristan, 2007) and the lamprey (Kozlov et al., 2009; Sarvestani et al., 2012). A remarkable level of conservation of CPG structure and cell type composition has been shown in vertebrates with very different modes of locomotion (lamprey, *Xenopus*, salamander, zebrafish and mice) whose common ancestor (Bagnall and McLean, 2014; Bem et al., 2003; Goulding, 2012) lived 450Mya[8]. Characterizing *Platynereis* locomotor CPGs will allow us to evaluate to which level this conservation extends to the protostomia branch in the evolutionary tree, and will help to comprehend the locomotor capabilities of the common ancestor (550Mya) of this group of animals.

[7] It could be even possible that protractor and retractor muscles are controlled by coupled oscillators, resulting in two Unit Burst Generators per hemisegment.

[8] Million years ago.

4.4 Lateral commissural neurons controlling locomotion

The origin(s) of commissural cells is an intriguing evolutionary event as these cells constitute an innovation of bilaterians (Arendt, Tosches, and Marlow, 2016). In this dissertation I have characterized in *Platynereis* a population of cells expressing the transcription factor *dbx1* (see Results section 3.4), that both in vertebrates (Lanuza et al., 2004) and *Drosophila* (Lacin et al., 2009) specifies commissural neurons involved in locomotion. In order to do that, I established the Crispr/Cas9 system in *Platynereis*[9], devised an unbiased image analysis strategy to quantify the morphological phenotype, and used Pduloco to examine behavioural defects.

dbx1 may specify commissural cells in *Platynereis*

Results section 3.4.2 shows that Crispr/Cas9 is a highly efficient tool to edit the genome of *Platynereis* and that it can be coupled with very efficient ways of genotyping (see also Results section 3.4.7). As mentioned, despite no off-targets effects have been found using the system (Results section 3.4.3), a genomic screen would be needed in order to properly quantify these effects. The high efficiency of the Dbx gRNA in editing *dbx1* locus (more than 95% of mutant sequences) allowed me to find significant phenotypic defects in the F0 population. For instance, I showed that *dbx1* mutants show a 20% reduction in the total number of commissures in the VNC at 50hpf, and that this is not caused by the injection protocol (Results section 3.4.4). I examined the expression of *dbx1* over development (Figure 4.3b), and found just a subset of cells expressing this TF in every segment. These cells are located laterally and are part of the lateral column found using ProSPr atlas (Results section 3.2.5). Consisting with previous results, the first segment shows an expanded population of *dbx1*+ cells (see Results section 3.2.4). The expression at early stages of *dbx1* in the midline cells is discussed below. These results indicate that there are many other commissural cell types, specified by other developmental programmes. More importantly, and despite we are missing direct evidence of the laterally located *dbx1*+ cells having an axon crossing to the other side, the strong effect on the commissural number indicates that these cells indeed project to the contralateral side of the body. We are currently working on using the *dbx1* promoter and the Crispr/Cas9 system (see Appendix section A) to genetically label the projections of this population of cells. An alternative to this is to integrate the connectomics data and the gene expression atlas (see Discussion section 4.5).

[9] To my knowledge this was the first time this system was shown to work in this species.

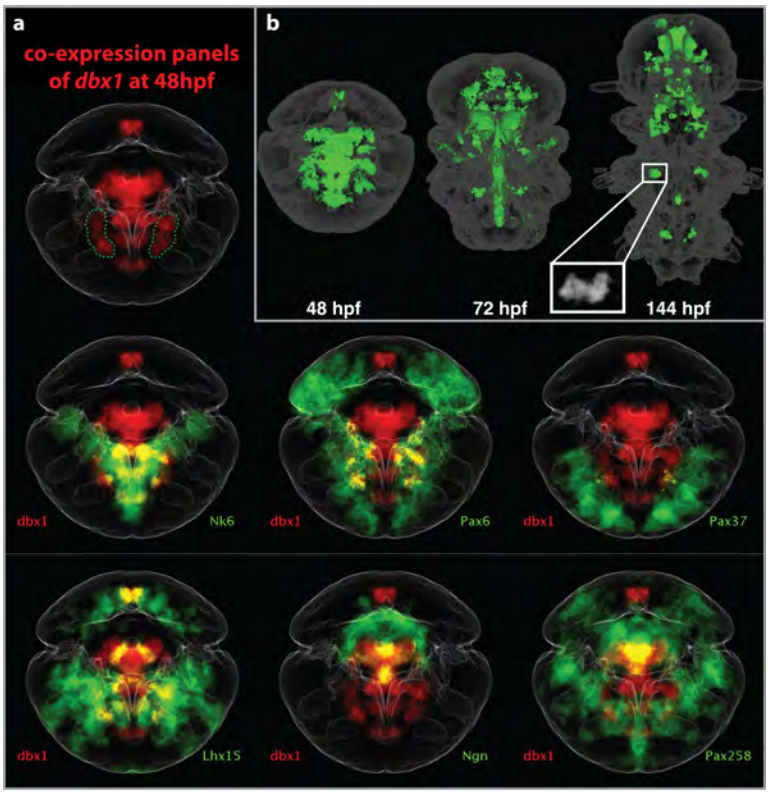

FIGURE 4.3: Expression of *dbx1* over development.
(a) Coexpression panels of *dbx1* with some relevant markers at 48hpf. Lateral *dbx1* cells outlined in green in the top panel.
(b) 3D rendering of *dbx1* expression at 48, 72 and 144 hpf. *In situ* signal from one individual showing two *dbx1* cells per side in the second segment.

Locomotion defects of *dbx1* mutants

Using PduLoco I found a locomotion defect in *dbx1* mutants specific to the left-right coordination during crawling, consistent with the elimination of specific commissures involved in CPG coupling. As already indicated in Result section 3.4.7, this is a preliminary result that requires the analysis of more mutants, but we can still draw some valid general conclusions of the data presented in this dissertation. First, *dbx1* mutants show a clear locomotor phenotype as they do not show episodes of rhythmic crawling. As decapitated animals can still crawl, it is unlikely that this defect is caused by *dbx1*+ cells in the head (see Figure 4.3b), although it is still possible that these cells are required to initiate crawling. Second, the use of excitatory drugs elicits locomotion in *Platynereis* larvae, making possible the analysis of mutant phenotypes with PduLoco. Specifically for *dbx1* mutants, preliminary data suggest that locomotion defects are caused by a miscoordination of the contralateral circuits.

Midline expression of *dbx1*

The expression of *dbx1* by the midline cells at early larval stages (Figure 4.3b) constitutes a problem for the interpretation of these results. This is because the midline cells are instrumental for the process of commissural guidance, expressing attractants and repellents that direct the growth and direction of axons (see Appendix B).

Therefore, the phenotypic effects described above could be consequence of the removal of the midline cells. However, there is some indirect evidence that suggests this is not the case:

- Only a subset of the commissures are affected. This is unlikely to be caused by a mutant-mosaic effect given the high efficiency of Dbx1 gRNA and the consistency of the result across batches.
- No axons show misguidance effects, which is a recurrent phenotype seen when there is a defect in the midline cells.
- There is no apparent phenotype in the midline, and this structure seems normal over development.
- The structure of the VNC (except for the 20% reduction in commissures) and the attachment of the muscles (highly dependent on midline signalling) are normal at different larval stages.

Therefore, this very precise phenotypic effect seems to be caused by the lack of lateral *dbx1* commissural cells and not by the misregulation of the midline cells. Nevertheless, these observations do not represent a definite proof of lateral *dbx1* cells being commissural and controlling locomotion in *Platynereis*. For instance, *dbx1* could control the midline expression of a chemokine specific for just a subpopulation of commissural cells, which would generate the phenotypic effects measured here. Several possibilities to solve this problem include the use of an optogenetically activable Crispr/Cas9 system to direct the activity just to the lateral domains, the generation of Cre-recombinant lines to selectively knock out both cell groups independently, or the finding of a more specific marker for the lateral *dbx1* cells using the expression atlas in combination with single-cell sequencing (see Discussion section 4.5).

Molecular profile of *dbx1* cells across development

Making use of the gene expression atlases, we can determine the molecular profile of the *dbx1* cells. At 6dpf, unfortunately, we don't find many co-expressed markers for the laterally located segmental cells. The only co-expressed gene, apart from markers indicating that these cells are neurons (*ElaV* and *Syt*) is the transcription factor *pax6*. It is interesting to note that *dbx1*+ cells in vertebrates develop from the *pax6*+ territory. However, vertebrate cells also express *Lhx1/5*+ postmitotically (Lu, Niu, and Alaynick, 2015), whereas *Platynereis* cells do not. We are currently limited in this comparison as there are not many more general markers known for this population in vertebrates (see Figure 2.3). Interestingly, none of the *dbx1*+ cells in *Platynereis* co-expressed *evx* at 6dpf (see Section 2.2.2). The cells in the the first segment, with no counterpart in the second and third, are also gabaergic and co-express *hb9*.

Nevertheless, we can make use of the 48hpf atlas to look at another time point during development, as these lateral cells can also be easily identified (Figure 4.3a). Indeed, we find that at 48hpf lateral cells do co-express *dbx1* and *Lhx1/5*, and are located in the *pax6* territory, adjacent to the *pax3/7* one. It is likely that one of the two cells in each side (Figure 4.3a shows two cells per side per segment) also expresses *nk6* at 48hpf. Unlike vertebrates though, no co-expression with *ngn*, which is not expressed in the trunk, or *pax2/5/8* is detected. Nevertheless, the lateral cells are already differentiated at 48hpf, and similarly to the *Lhx1/5* case, these cells might co-express these markers at earlier stages. The use of the 28 and 36 hpf atlases (see Results section 3.36) will help resolving this issue.

Comparison of *dbx1*+ cells in bilaterians

In order to unbiasedly compare neuronal cell types across animals to accurately distinguish between homology and homoplasy, we have to gather information at different levels (expression, connectivity, function...) for various branches of the evolutionary tree. *dbx1*+ cells represents an ideal case for bilaterian comparison, as much information is already known. For instance, both in vertebrates (*Xenopus*, zebrafish and mice) (Lanuza et al., 2004) and in *Drosophila* larvae (Lacin et al., 2009; Lacin et al., 2014), *dbx1*+ cells are mainly commissural, gabaergic, located in the medio-lateral neural plate, and involved in the control of locomotion. I found a similar case in *Platynereis*[10], although the neurotransmitter identity of these cells still remains to be identified.

With regard to *dbx1* subpopulations, there are some discrepancies across the tree. For instance, in *Drosophila* some of the *dbx1*+ cells co-express *engrailed* (Lacin et al., 2009). In vertebrates however, *engrailed1* specifies the V1 population and these TFs co-repress each other to define specific developmental programmes (Thelie et al., 2015; Gribble, Nikolaus, and Dorsky, 2007). Similar as in vertebrates, I found no co-expression of these two markers in *Platynereis*. A similar scenario occurs with *eve*, where in vertebrates *dbx1*+ controls the expression of *eve* in a subpopulation, and in *Drosophila* they downregulate each other. In this case, and more similar to the *Drosophila* case, no co-expression is found between *eve* and *dbx1* in *Platynereis* at 6dpf[11]. These discrepancies for neuronal subpopulations are consistent with the scenario outlined above (see Discussion section 4.2), where there seems to be an expansion of a basic common bauplan in the vertebrate line, and very likely in other lineages too.

The data presented in this dissertation highly supports a conserved role of *dbx1*+ commissural neurons in controlling locomotion, found in bilateral animals and

[10]Note that the domain called "lateral" in this dissertation, refers to the VNC at 6dpf, and is indeed located in a medio-lateral position during development (see Figure 4.3a).

[11]This still needs to be tested at earlier stages

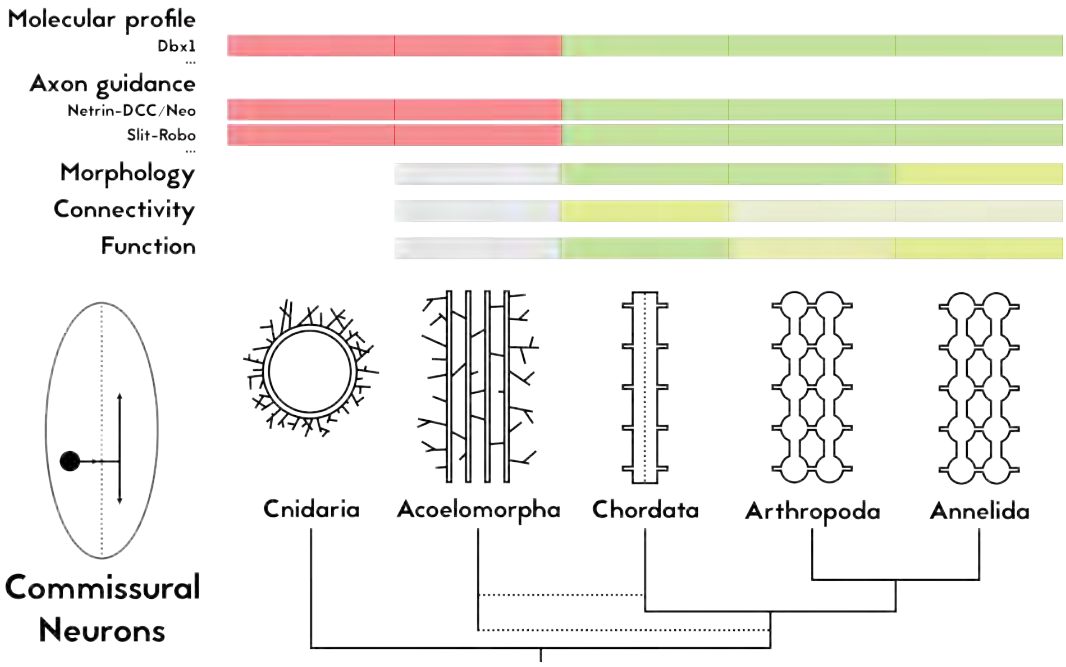

FIGURE 4.4: Commissural cells in centralised nervous systems.
Schematic of different blastoporal nervous systems, and properties of a subpopulation of commissural cells specified by *dbx1*. A color scheme, red and green, is used to indicate similarity. Red indicates low conservation, whereas green indicates high conservation. The fading of the color illustrates the amount of information available for that particular property.

probably present in the common ancestor (Figure 4.4). The evolution of this new type of cell (commissural neuron) represents a true bilaterian innovation, and this work strengthens its association with Dbx1, as this transcription factor is not found in animals without a centralised nervous system. During 550 million years, this cell type may have diversified and specialised further into different lineages to accommodate the specific locomotor adaptations seen in extant animals.

4.5 Towards a system-level understanding of *Platynereis dumerilii*

As I mentioned at the beginning of this dissertation (see Section 1.3), *Platynereis* represents an ideal model for which we can realistically aim towards a system-level understanding of its larval nervous system. The approaches that I have carried out during my PhD work have always been directed towards this goal. I have described some of them in the Results chapter, and of special relevance is the development of ProSPr (see Results section 3.1). I also suggest the reader to refer to Appendix section A, where I have summarised other results that I have obtained during these past four years, and despite these projects are not yet in an advanced

state, they represent parallel approaches to unravel different aspects of *Platynereis* nervous system, necessary for its system-level understanding.

The full connectome of *Platynereis* at 6dpf

I aimed to unravel the complete connectome of the 6dpf larvae. This means reconstructing the morphology of all the neurons, and identifying the synapses between them. Together with the EM facility in EMBL Heidelberg, we explored different fixation methods, and achieved a high-quality standard (see Appendix section A.1) that will soon result in a high resolution electro-microscopy sampling of *Platynereis* 6dpf larvae in three dimensions (the dataset is being currently acquired by the EM facility in the FMI in Basel) using SBFSEM (Wanner et al., 2016). The reconstruction and tracing of neural processes still represents the bottleneck for these approaches, as the current best standard is to do manual tracing. We estimate the average neurite length to be around 300μm, and considering the total number of neurons at 6dpf, it could take up to 10 human years to trace the full connectome. Nevertheless, the need for automatic methods, caused by the worldwide interest in neuroscience, is motivating the development of image analysis algorithms that can currently speed up the tracing process (Maco et al., 2014; Kreshuk et al., 2014; Helmstaedter et al., 2013; Helmstaedter, Briggman, and Denk, 2011), and will most likely outperform humans in the near future. It is still too soon to predict which method we will employ to trace the dataset, as this requires an analysis of the quality (contrast between the membrane and the cytoplasm, and imaging properties of the synapses) of the neuropil in different parts of the animal.

Parallel to the tracing process, I aim to integrate the connectome data with the ProSPr expression atlas. The signal from the nuclei in the connectome can be easily classified using Ilastic (ilastic.org), as they have very unique features that allow automatic segmentation (data not shown). It will be possible then to produce a "rough" three dimensional alignment to the ProSPr reference. It is important to note that this registration will most likely perform much worse than with *in situ* data, as the fixation protocol for the EM is very different and produces other deformations. Nevertheless, manual annotations of those tissues that can be identified in the reference (e.g. examples in Figure 3.27b, glands, neuropil...) and that can definitely be annotated in the EM data, can be used as landmarks to guide the registration. In order to correlate the cells within the neural tissue (brain and VNC show a high packing of cells that will not register using the above-mentioned general landmarks), we can make use of antibody labelling (e.g. 5HT, neuropeptides...), that at the same time provide the information about cell identity (used to identify the cell in the ProSPr reference) and the cell projections (used to identify the cell in the connectome). The cross correlation of these points through the neural tissue can be done using multiple antibodies that label cell projections, and corrections on

areas that did not match can be made until no further improvement is seen when adding new markers. At that point, we could be confident that all cells (at least the ones within areas covered by landmarks) are correlated across the two datasets, specially if we detect a bilateral confirmation.

An extra layer of information can be incorporated into this integrative atlas, as it is possible to use ProSPr to spatially map the transcriptome reads from single cells (Achim et al., 2015). Together with Dr. Achim, we are mapping 350 single-cell transcriptomes into the newly developed full body expression atlas for 48hpf (manuscript in preparation and data not shown). This way, expression atlases created with ProSPr can be used as a "bridge" between the connectome and the transcriptome to obtain, for potentially all cells in an organism, their complete set of transcripts, their morphology, and their circuitry connectivity. Furthermore, the cells within these atlases can be linked through development as ProSPr performs very well for earlier stages (see Results section 3.2.7), we have collected transcriptomes for single cells at different stages (work by Dr. Achim and Dr. Bertucci), and other labs have already succeeded in reconstructing full animal connectomes for the 72hpf larvae (Randel et al., 2014; Randel et al., 2015).

Functional analysis on cell types and neural circuits

The integration of the above mentioned descriptive technologies into virtual atlases would be extremely useful to generate unbiased hypothesis about the functional role of certain cell types and neural circuits within the animal. But in order to test the validity of these hypothesis, we will need to make use of techniques to functionally assess the relevance of specific cell types in specific circuits. During the work described in this dissertation, I set up some of these techniques to assess the functional role of locomotor circuits components. These techniques include the exploration of neuronal activity fluorophores coupled with high-speed recordings (Appendix section A.2), and the proof of concept for the use of optogenetic techniques (Appendix section A.4). Of particular relevance is the establishment in *Platynereis* of Crispr/Cas9 technique to knock down (Results section 3.4 and Appendix section A.3) the transcription factors that in vertebrates specify the main locomotion-related cell populations, as well as the development of methods to quantify the phenotype (Results section 3.3). The genetic markers selected in this study represents a biased selection (driven by the knowledge of the vertebrate spinal cord), but having the integrative atlases mentioned above we would be able to generate unbiased targets based on the structure of the circuits. Motorneurons would be easily identified by their connection to the muscles, and classified based on their specific muscular targets according to what we have learned using PduLoco (see Discussion section 4.3). Interesting interneurons (based on their position

within the network, their number of synapses, etc.) connecting to these motorneurons, mono and poly-synaptically, could be genetically targeted as we would have several markers (and potentially even their full transcriptome) using the expression atlas.

The work presented here brings us closer to the nowadays realistic goal of generating, for model and non-model organisms, highly informative atlases that integrate system-level information. Guided by these "virtual embryos", we would be in a position to direct functional studies to understand, at the level of cell types, the neuronal circuits that drive the behaviour of animals. This would represent an unbiased methodology to compare the body plans of different species, which will hopefully help unravelling the history of cell types, as well as the evolutionary paths that resulted in today's wonderful diversity of nervous systems and animal behaviours.

Chapter 5

Conclusions

Whereas the evolutionary relationships between animals within phyla are widely accepted in the scientific community, the correspondences across phyla are still a major topic of debate, specially whether centralized nervous systems are homologous or convergent (Arendt et al., 2008; Holland et al., 2013; Strausfeld and Hirth, 2013; Holland, 2016). This question is crucial for our understanding of biology, as from very similar molecular territories (Arendt, Tosches, and Marlow, 2016; Denes et al., 2007) animals develop very diverse neural structures capable of multiple different behaviours. Traditional evolutionary comparisons lack the resolution to solve this issue, as they either look into the structure of tissues (e.g. palaeontology), or the structure of the genome (e.g. phylogenomics). Instead, in order to fully explain the composition of nervous systems, we need to describe the components that they are made of, the cell types, and how they connect and relate to each other (Arendt, 2008). Fortunately, we have reached an era where technological developments in various fronts allow the access to cell type information, and for the first time, it is possible to build integrative approaches to achieve a comprehensive view of the body plan of metazoans.

The work presented in this dissertation provides a powerful technical framework to exploit the cell type concept to molecularly classify animal body plans in a system-level manner. Of special relevance is the implementation of an automated pipeline called ProSPr, used to build gene expression atlases with cellular resolution for complex body plans. ProSPr represents a very useful resource to study the development of the annelid *Platynereis dumerilii*, and its potential applications to other animals can significantly change the throughput of the research in the EvoDevo community. This methodology has allowed me to gain very interesting evolutionary insights into the relationship between the protostomes and deuterostomes nervous system. For instance, and unbiased analysis classifies the post-mitotic ventral nerve cord of *Platynereis* into medio-lateral territories that much resemble, in molecular identity, groups of adjacent ventro-dorsal subdivisions of the spinal cord. Interestingly, one of these territories shows a relative position (with respect to the nervous system) and molecular specification much like the cerebrospinal fluid-contacting

neurons (CSF-cNs or Kolmer-Agduhr cells) in vertebrates. These similarities support an evolutionary scenario where the CSF-cNs derive from an ancestral neuronal population contacting the environment and controlling locomotion, that with the progressive specialization of vertebrate life-styles gave rise to the most ventral cell types of the locomotor CPG.

The design of an image analysis software (PduLoco) to monitor *Platynereis* crawling behaviour has proven useful to describe in detail the kinematics in this animal, that indicate a structure of the locomotor circuits similar to other segmental animals. Additionally, and together with the implementation of Crispr/Cas9 technology, PduLoco has shown very practical in characterizing a commissural population of cells, specified by Dbx1, controlling locomotion in *Platynereis*. The analysis of the molecular fingerprint of these cells during development indicates its conservation in vertebrates and *Drosophila*, and its likely role in contralateral coordination during locomotion in the bilaterian ancestor.

Besides the generation of gene expression atlases and tools to quantify behaviour, I have explored the implementation of other parallel approaches to complement the system-level analysis of *Platynereis*. These include setting the stage for generating a full connectome of the larvae, and techniques to monitor and manipulate neuronal activity. I have demonstrated here the usefulness of complementary system-level approaches to describe cell types and guide the research in the evolution of locomotor circuits. However, their potential extends far beyond, and I expect that the integration of these approaches into multi-dimensional atlases can lead towards a comprehensive understanding of nervous systems and metazoan evolution.

Chapter 6

Materials and methods

6.1 Reagents

PTW: 1x PBS pH 7.4 + 0.1% Tween-20, sterile filtered
Hybmix: 50% formamide, 5xSSC, 50 mg/ml heparin, 0.1%Tween20, 5 mg/ml torula yeast RNA (Sigma®)
20XSSC: 3M NaCl + 0.3M Dinatrium citrate
SSCT: 1XSSC + 0.1%Tween
MAB: 0.1M Maleic Acid + 0.15M NaCl
MABT: MAB + 0,1% Tween-20
Staining buffer 7.5: 100 mM TrisCl pH 7.5, 100mM NaCl, 0,1%Tween20
Staining buffer 7.5 (+$MgCl_2$): Staining buffer 7.5 + 50mM $MgCl_2$
Staining buffer 9.5: 100 mM TrisCl pH 7.5, 100mM NaCl, 0,1%Tween20, 50mM $MgCl_2$
Staining solution: Staining buffer 9.5 + 4.5μl/ml NBT + 3.5μl/ml BCIP
Annealing buffer: 10mM Tris, 50mM NaCl, 1mM EDTA, pH: 7.5-8.0

6.2 Biological samples

6.2.1 *Platynereis dumerilii* culture

All animals used in this study have been derived from a laboratory breeding culture set in the EMBL Heidelberg, established for more than twenty years. Animals are grown during their maturation phase in plastic boxes in natural sea water (NSW). Animals are fed with *Tetraselmis* up to three months, and then with a mixture of *Tetraselmis* and *Dunaliella tertiolecta*. Sexual maturation is synchronized by using a light bulb to emulate the moon cycle (rotations of 3 weeks off and 1 week on). Temperature during the entire life cycle is maintained to 18°C. Crossings are achieved by introducing one female and one male in the same plastic cup in NSW. Fresh batches are raised during the first week of development (or less depending on the experiment) in the same cup, with a change of water 24hpf to remove the oocyte

jelly protection layer. Fresh batches are kept in an incubator (Type KB53, Binder, Tuttlingen, Germany) to maintain a constant temperature and ensure proper developmental synchronization between batches. Light cycles in the incubator are 16 hours light and 8 hours dark.

6.2.2 Larvae fixation protocol

Animals were fixed following the standard paraformaldehide (PFA)-based fixation protocol:

- Collect animals in the mesh.
- Wash with filtered natural sea water (FNSW).
- Treat with 100μg/ml of proteinase K diluted in PTW for 2 minutes[1].
- Wash with FNSW.
- Relax muscles with 300mM of $MgCl_2$ in FNSW during one minute.
- Fix in 4% PFA in PTW for 2 hours.
- Rinse 3-4 times in PTW.
- Transfer to 100% MeOH and store at -20°C.

Note that the fixation is different in the case of SBFSEM preparation (see Section 6.5 below).

6.2.3 DNA extraction

Extraction of DNA from *Platynereis* larvae was performed using the QIAamp DNA Micro Kit from Qiagen® following the manufacturer instructions for DNA extraction from tissues. The number of larvae used to extract DNA from varied from 1 to 50 individuals.

6.2.4 RNA extraction and cDNA libraries

RNA was extracted from a pull of larvae at different stages using the RNeasy Mini Kit from Qiagen®, following manufacturer instructions. cDNA libraries were prepared using the Superscript III Reverse Transcription Kit from Invitrogen®, following manufacturer instructions.

[1]This time is for 6dpf larvae. Duration varies between stages

6.2.5 Microinjection protocol

Platynereis fertilized oocytes were injected following the established injection protocol, using a microinjector, glass capillaries and an agar-based support (see Figure 3.43b for illustration). A thorough description of the injection protocol can be found in Dr. Tosches PhD thesis.

6.3 Gene expression detection

6.3.1 Generation of probes

PCR and cloning

Primers were designed using ApE software, setting an aligning temperature of 59 to 63°C. PCRs were performed using the HSTaq DNA polymerase from Qiagen®, according to manufacturer instructions. PCR products were purified using the QIAquick Nucleotide Removal kit from Qiagen®, and cloned into the Topo vector using the Topo TA Cloning Kit from Thermo Fisher Scientific®.

Probe synthesis

Plasmids were linearized and purified using the QIAquick Nucleotide Removal kit from Qiagen®. The transcription reaction was assembled as follows:

- 1μg of linearized plasmid (up to 6μl)
- 0-5μl of H_2O
- 2μl of 10x Transcription Buffer Sigma-Aldrich®
- 2μl of DTT 0.1M
- 2μl of Dig-labelled nucleotides Sigma-Aldrich®
- 1μl of RNA inhibitor Sigma-Aldrich®
- 2μl of RNA polymerase (T7/SP6) Sigma-Aldrich®

The reactions were set at 37°C for 4 hours. 2μl of TurboDNAse was added to the mixture, and the reaction was left for 1 hour at 37°C. The Dig-labelled RNA was cleaned using the RNeasy Mini Kit from Qiagen®, and it was eluted in 30μl of water from Sigma®. After measuring the RNA concentration, probes were diluted to 50 ng/μl in Hibridization Buffer.

6.3.2 Whole mount *in situ* hybridization protocol

The start of this protocol assumes animals have been fixed according to Section 6.2.2.

Rehydration

- Rehydrate the animals: 5 minutes in 75% MeOH/25% PTW, 5 minutes in 50% MeOH/50% PTW, 5 minutes in 25% MeOH/75% PTW.
- Rinse twice with PTW.

Permeabilization

- Permeabilize with 100μg/ml of proteinase K diluted in PTW for 3 minutes[2].
- Stop the reaction with two washes of 2mg/ml of Glycine in PTW for 2 minutes.
- Rinse twice with PTW.

Acetylation

- Wash the larvae 5 minutes in 1% trietaniolamine (TEA) 99%PTW (prepare fresh).
- Wash the larvae 5 minutes in 1% trietaniolamine (TEA) 99%PTW with 3μl/ml of acetic anhydride. Shake the solution (after adding the acetic anhydride and before adding it to the animals) during 15 seconds.
- Wash the larvae 5 minutes in 1% trietaniolamine (TEA) 99%PTW with 6μl/ml of acetic anhydride.
- Rinse three times with PTW for 5 minutes each.

Fixation

- Immerse larvae 20 minutes in 4% PFA in PTW.
- Rinse five times with PTW.

Hybridization

- Incubate animals in Hybmix solution for 1 to 2 hours at 65°C.
- Dilute the probes to a final concentration of 1-5ng/μl in Hybmix solution.
- Denature the probes at 90°C for 10 minutes and transfer immediately to ice.
- Remove Hybmix solution from the animals and add the probe.
- Hybridize during 24-72 hours at 63°C shacking gently.

[2] see footnote 1

Washes

- Wash the animals for 15 minutes in Hybmix solution at 65°C.
- Wash the animals twice for 30 minutes in 50% formamide - 50% 2xSSCT at 65°C.
- Wash the animals for 15 minutes in 2xSSCT at 65°C.
- Wash the animals twice for 30 minutes in 0.2xSSCT at 65°C.
- Bring the animals to RT during the last wash and transfer them to PTW.

Immunohybridization

- Wash the animals for 5 minutes in MABT buffer.
- Block in 2.5% sheep serum in 1xMABT, for 1 hour at RT, shaking at 450rpm.
- Dilute anti-DIG Fab fragments 1:4000 in 2.5% sheep serum in 1xMABT. Incubate the animals for one hour at RT and one night at 4°C, shaking at 450rpm.
- Wash the animals 6 times for 5 minutes in PTW shaking at 450rpm. For a better signal to noise ratio, the last PTW wash can be done overnight at 4°C.

NBT/BCIP staining

- Equilibrate the larvae 5 minutes in staining buffer, pH 7.5, without $MgCl_2$, while shaking.
- Equilibrate the larvae 5 minutes in staining buffer, pH 7.5, with $MgCl_2$, while shaking.
- Equilibrate the larvae 5 minutes in staining buffer, pH 9.5 while shaking.
- Transfer the animals to the staining solution and develop the *in situ* in the dark. The timing of this step is dependent on many factors like probe concentration and mRNA abundance (between others). It is recommended to check the animals periodically and monitor the colorimetric reaction to achieve a good signal to noise ratio.
- Stop the reaction by transferring the animals to PTW.

6.3.3 Confocal imaging of WMISH

- Rinse the embryos with PTW a couple of times.
- Incubate the animals in 1μg/ml of DAPI in PTW for one night at 4°C.
- Wash the embryos twice in PTW for 10 minutes.
- Transfer the animals to 33% TDE (2,2'-thiodiethanol) - 66% PTW for 15 minutes shacking at 450rpm.

- Transfer the animals to 66% TDE - 33% PTW for 15 minutes shacking at 450rpm.
- Transfer the animals to 97% TDE - 3% PTW.
- Shake at 450rpm for 1 hour.
- Prepare the animals for imaging in a microscopic slide. Use two layers of tape to create a space between the slide and the coverslip.
- Image using the following settings: 40x objective, 0.75 zoom, 696x696 pixels, z step size of 0.55μm [3].
- Image the DAPI channel using the 405nm laser and the NBT/BCIP signal using the 633 laser in reflection mode. (Optional: image the bright field and/or the autofluorescence of the animal with the 488 or the 633 laser).

6.4 Other stainings

6.4.1 Inmunostaining

The start of this protocol assumes animals have been fixed according to Section 6.2.2.

- Rehydrate the animals: 5 minutes in 75% MeOH/25% PTW, 5 minutes in 50% MeOH/50% PTW, 5 minutes in 25% MeOH/75% PTW.
- Rinse twice with PTW.
- Permeabilize with 100μg/ml of proteinase K diluted in PTW for 3 minutes[4].
- Stop the reaction with two washes of 2mg/ml of Glycine in PTW for 2 minutes.
- Rinse twice with PTW.
- 20 minutes of 4% PFA in PTW.
- Rinse five times with PTW.
- Block with 5% Sheep Serum in PTW for 1 hour at room temperature (RT), shaking at 450rpm.
- Add primary antibody in blocking solution over night at 4°C, shaking at 450rpm.
- Rinse five times with PTW.

[3] These are settings to get an isotropic resolution of 0.55μm/pixel for the 6dpf ProSPr. Other stages require other parameters.

[4] see footnote 1

- Block with 5% Sheep Serum in PTW for 20 minutes at room temperature (RT), shaking at 450rpm.
- Add secondary antibody in blocking solution over night at 4°C, shaking at 450rpm. (Optional: add 1μg/ml of DAPI during this step).
- Rinse five times with PTW.
- Transfer to imaging medium (DABCO/glycerol or TDE).

All antibodies are diluted 1:250.

6.4.2 EdU staining

EdU stainings were performed using the Click-iT EdU Alexa Fluor 647 Imaging Kit from Thermo Fisher Scientific® following manufacturer instructions.

6.5 SBFSEM fixation protocol

This protocol has been developed by Pedro Machado from the EMBL EM Facility, and is reproduces here with his permission.

6.5.1 Chemicals and reagents

0.3M cacodylate with 4mM $CaCl_2$ 500ml stock solution 0.3M cacodylate buffer with 4mM $CaCl_2$: - 0.22g $CaCl_2$ - 32.10g Cacodylate - 500ml ddH_2O - Adjust pH to 7.4 with HCl

1.5 % KFeCN in 0.15M cacodylate buffer with 2mM calcium chloride and 2% aqueous OsO_4: 0.15g $K_4Fe(CN)6x\ 3H_2O$ in 5ml 0.3M cacodylate with 4mM $CaCl_2$ mixed with 5ml 4% aqueous OsO_4

TCH Add 0.1mg thiocarbohydrazide to 10ml ddH_2O and place in a 60°C oven for 1h (to dissolve swirl every 10min). Filter TCH solution through a 0.22μm Millipore syringe filter.

Aspartatic acid: Aspartatic acid 0.998g in 250ml H_2O dissolves faster if pH is raised to 3.8. Check pH of lead aspartate solution several times with pH strips.

Walton's lead aspartate: 10ml 0.03M aspartic acid solution, 60°C Add 0.066g lead nitrate Adjust pH to 5.5 with 1M KOH

Durcupan ACM resin: 11.40g part A 10.00g part B 0.30g part C 0.05-0.10g part D Mix thoroughly

6.5.2 Fixation procedure

- Immerse larvae in 2.5% glutaraldehyde in 0.15M cacodylate buffer with 2mM calcium chloride, pH 7.4. Maintain at RT for 15 minutes, and afterwards on ice for 3 hours to 4 days.
- Wash 5 x 3 minutes in 0.15M cacodylate buffer with 2mM calcium chloride.
- Immerse in freshly prepared 1.5% KFeCN in 0.15M cacodylate buffer with 2mM calcium chloride and 2% aqueous OsO_4 on ice for 1 hour.
- Wash 5 x 3 minutes in ddH_2O.
- Immerse sample in filtered TCH solution at RT for 20 minutes.
- Wash 5 x 3 minutes in ddH_2O.
- Immerse in 2% OsO_4 in ddH_2O at RT for 30 minutes.
- Wash 5 x 3 minutes in ddH_2O.
- Place in 1% uranyl acetate in ddH_2O and store in fridge at 4°C overnight. (Prepare the Walton's lead aspartate fresh in the following day.)
- Wash 5 x 3 minutes in ddH_2O.
- Immerse in lead aspartate solution in a 60°C oven for 30 minutes.
- Wash 5 x 3 minutes in ddH_2O.
- Dehydrate in ice cold EtOH 6 x 5 minutes (20%, 50%, 70%, 90%, 100%, 100%).
- Immerse in glass distilled acetone.
- Immerse in 25% Durcupan:acetone mix for 2 hours.
- Immerse in 50% Durcupan:acetone mix for 2 hours.
- Immerse in 75% Durcupan:acetone mix for 2 hours.
- Immerse in 100% Durcupan overnight.
- Place in freshly prepared 100% Durcupan for 2 hours.
- Remove samples from resin, drain excess resin and polymerize in Aclar sheets at 60°C oven for 48 hours.
- Mount sample upright in stubs with very little fast-hardening silver resin and polymerize at 80°C oven for 3 hours.
- Add extra silver resin after the samples are well stuck to the stubs and polymerize at 60°C oven for 48 hours.

6.6 *In vitro* transcription of mRNAs

The mRNAs for the GCaMP6s, the ArcLight, the ChR2 and the Cas9 were produced using the mMessage mMachine High Yield Capped RNA Transcription Kits from Ambion®, following manufacturer instructions, and cleaned using the MegaClear Kit from Thermo Fisher Scientific®

6.7 Drug analysis

The behaviour of animals was monitored and annotated every second for every individual. A chamber was created to have the animals inside in a reduced volume so the entire field could be recorded, but with enough space so they could move freely. Between 10 and 20 animals were put in the chamber. The NFSW was pipetted up and down to control for the water flux. Animals were recorded during one minute. Volume was taken out and the same volume was introduced, this time containing the drugs (5-HT and NMDA, both from Sigma-Aldrich®). Pipetting was done up and down. Animals were recorded during one minute. For each drug concentration, this procedure was repeated twice. Individual animals were monitored and for those individuals for which complete monitoring was available, all the seconds were annotated according to the behaviour that the animal exhibited.

6.8 Crispr/Cas9-mediated gene knock outs

6.8.1 Design of guide RNAs

The design of the guide RNAs was done following the protocol in Hwang et al., 2013 with some modifications. Once the gene model is reconstructed, a 100-400bp region was selected to scan for possible target sites.
The online tool ZiFit (http://zifit.partners.org/) was used to generate candidate gRNAs. The specificity and SNP content of every suggested gRNA was checked using genomic and transcriptomic reads, and the best candidates were selected for cloning.

6.8.2 Cloning and transcription of guide RNAs

The oligos suggested by ZiFit were ordered from Sigma-Aldrich®, and they were annealed using a slow cool-down approach from 95°C to room temperature in Annealing buffer (3μl of 100μM oligos in 44μl of buffer). Oligos were cloned into the vector DR274 (Addgene - Plasmid 42250), previously digested with BsaI.

Correctly cloned vectors were digested with DraI to linearize. The *in vitro* transcription reaction was assembled as follows using the MAXIscript Kit from Thermo Fisher Scientific®:

- 1μg of linearized plasmid
- 2μl of 10x Transcription buffer
- 1μl of 10mM ATP
- 1μl of 10mM GTP
- 1μl of 10mM CTP
- 1μl of 10mM UTP
- 2μl of T7 enzyme
- H_2O to 20μl

Reaction was set at 37°C for 4 hours, DNA was removed with TurboDNAse, and the gRNA was purified with the MegaClear Kit from Thermo Fisher Scientific®.

6.8.3 Injection cocktail

Injection cocktail contained 12ng/μl of gRNA (volume replaced by H_2O in control), 333ng/μl of Cas9 mRNA, and 1/5 of dextran.

6.8.4 Analysis of mutants

PCR and restriction reaction

Standard nested PCRs were performed from genomic DNA using two pairs of specific primers. The nested PCR products were purified with the QIAquick Nucleotide Removal kit from Qiagen® and digested for 3 hours using the specific unique enzyme selected for genotyping (temperature dependent on the enzyme). Digested product was purified with the same kit before running it on a 3% MetaPhor Agarose gel for 3 hours at 90V.

Sequencing

The nested PCR product was cloned into the Topo TA vector, and individual clones were sent for sequencing. The alignments were performed using the software TCoffee.

Commissure count analysis of Dbx mutants

Animals were fixed at 50hpf, stained with AcTub, and mounted in 80% Glycerol. A ventral orientation of the larvae was achieved by rolling the individuals moving the coverslip. Confocal (Leica - SP8) imaging at high resolution (0.1024x0.1024x0.3357 xyz) was performed using the 63x objective. A rectangular region of interest was selected to cover the commissures (fixed area for all animals). Intensity peaks in multiple medio-lateral locations were detected to extract the mean as a representation of the commissural density for each animal.

6.9 Optogenetic control with ChR2

Animals were injected with 333ng/μl of ChR2 mRNA and 100ng/μl of H2B-GFP mRNA (kindly provided by Dr. Vopalensky). 12 hours previous to the optogenetic experiment, 3μM of All-trans-retinal (Sigma-Aldrich®) was added to the animals in NSW. 405nm wave length light was used to open the ChR2 channel and depolarize the neurons.

Appendix A

Other results

In the following sections I will summarize other advances on the set up of different neurobiological analysis techniques. These approaches are targeted to uncover some aspects of *Platynereis* nervous system, both in a systemic and a specific way. Whole nervous system approaches:

- Synaptic connectivity description. Optimization of fixation techniques to reconstruct the full neuronal circuitry of *Platynereis* at 6dpf using Electron Microscopy.
- Complete monitoring of neuronal activity. Test the use of calcium and action potential indicators, coupled with fast-imaging microscopic techniques.

Cell type specific approaches:

- Knock-down of key transcription factors. Use of Crispr-Cas9 technique to target key developmental TFs.
- Manipulation of neuronal activity. Test the use of optogenetic techniques to control the activity of neuronal populations.

A.1 Neuronal morphology and circuit connectivity in a full organism

Probably the biggest difference between neurons and any other cell type is their integration into structural and functional assemblies (circuits) capable of complex behaviours. These functional circuits are formed by physical connections (synapses) between neurons located anywhere in the animal body plan, through their cellular protrusions (axons and dendrites). Therefore, the reconstruction of these circuits is a necessary step to understand the function of nervous systems, and for that it is crucial to trace cellular protrusions and map their synapses.

Many efforts have been made in recent years towards this end, and the development of new techniques to identify synaptic neurons has been a very active field

of research. It is beyond the scope of this thesis to discuss and compare these technologies, but I will mention some of them and what are they based on.

Cell labelling and optic imaging:

Anterograde and retrograde tracing, based on die injections into the cells.

Brainbow, based on recombinase activity to label, genetically, neurons with different fluorescent proteins.

Synapse detection:

Viral infection, modified to infect pre-synaptic partners, and fluorescently labelled.

Grasp (GFP Reconstruction Across Synaptic Partners).

Temporal activity correlation:

Electrophysiological recordings of pairs of neurons.

Imaging of neuronal activity with high temporal resolution.

Electron-microscopy imaging:

Chemical or cryo fixation of neuronal tissue, and high-resolution sectioning and imaging.

The nature of the information extracted from these techniques if fundamentally different, as some of them detect functional synapses of a specific or pre-defined set of neurons, whereas other reveal the morphology of many neurons. In particular, electron-microscopy based approaches are very powerful because they allow to acquire, in 3 dimensions, the complete information for many critical properties of the neural tissue, like the membrane and cytoplasm of the cells, the synaptic vesicles, pre and post synaptic partners, and GAP junctions (as well as all the main cellular components like mitochondria, ER, nucleus etc). New machines have been adapting this old approach in order to achieve a high resolution in all dimensions by sectioning the tissue in very thing layers, image each of them separately, and merge them back together in 3D. This allows for the tracing of all the cellular processes, reconstructing entire neurons and identifying synaptic partners.

We decided to use electron-microscopy for these reasons, and also because the size of *Platynereis* larvae at 6dpf is ideal to reconstruct a full animal using Serial Block Face Scanning Electron Microscopy (SBFSEM). This machine incorporates a cryotome to cut thin sections of tissue after the surface of the block has been imaged with SEM. It can produce large volumes of data in a highly automatic way.

In order to acquire the full volume of a 6dpf *Platynereis* larvae at high resolution, we teamed with the EM facilities in EMBL and FMI. Their support, expertise and working time during the sample preparation and data acquisition has been crucial for the development of the project. After experimenting with different fixation protocols, we settled for a chemical fixation with glutaraldehide and osmium that achieved the better membrane-cytoplasm contrast while maintaining the staining of vesicles to help identify synapses. Figure A.1 shows the results in the TEM and SEM of different sections in the larvae.

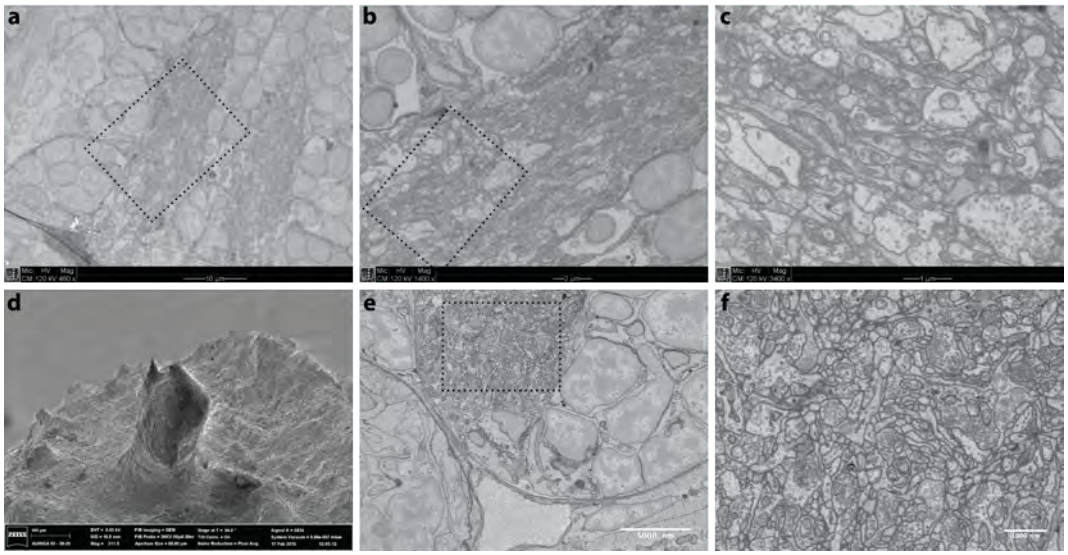

FIGURE A.1: Analysis of fixation quality of 6dpf samples for SEM.
(a-c) TEM images of a frontal section at the level of the VNC neuropil of 6dpf *Platynereis* larvae. **b** and **c** are higher magnifications of the dotted squares in **a** and **b**, respectively. **(c)** SEM image of the larvae mounted in the conductive resin (see panel d), prepared for its processing in SBFSEM. **(e),(f)** SBFSEM of a transversal plane at the level of the second segment. **f** is a magnification of the neuropil section of **e**. [Data from Pedro Machado and Christel Genoud].

After fixation (done in EMBL, Heidelberg), the samples are mounted vertically in the pins and embedded in an opaque conductive resin for their processing with SBFSEM (done in FMI, Basel) (see Figure A.1d). Several tests have given successful results, and the data is being acquired at the same time that this thesis is being written. Once the full dataset is acquired, and the sections are aligned, the neural processes need to be traced. Currently, manual tracing is still giving the best results, but this is a very time consuming process, and we have estimated that it would take ten human years to trace all neurons in the 6pdf larvae. However, given the popularity of EM approaches in neurobiology, there is a world-wide interest in developing automatic image-analysis approaches to solve this problem, and several tools to reduce the tracing time are already available. We plan to explore these tools once we have the data and can judge its characteristics, and probably settle for a combination of manual tracing and semi-automated algorithms.

A.2 Imaging neuronal activity

In the recent years, two key technical advances have made possible the recording of the activity of thousands of neurons during long periods of time. The first is the development of fluorescent indicators, located in the neurons (the delivery method can vary), that respond to neuronal activity (either calcium gradients or action potentials) by changing the fluorescent state. The second is the development of very

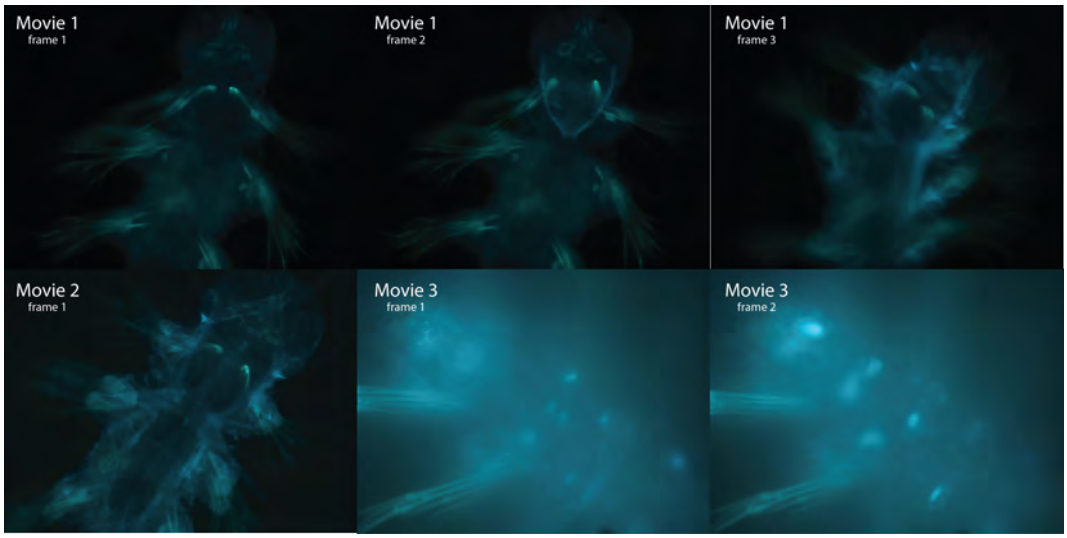

FIGURE A.2: GCaMP activity is maintained in larvae at old stages.
Single frames of 3 different movies of 6 to 10 dpf larvae injected with GCaMP6s.

fast and non-toxic imaging methods to cover large volumes of tissue at high temporal resolution. Coupling these two principles permits the imaging of large numbers of neurons visualizing and measuring their individual activity.

I aimed to used GCaMP6, a calcium indicator protein, and ArcLight, a voltage indicator protein, to record neuronal activity in *Platynereis*. Using Single Plane Illumination Microscopy (SPIM) I aimed to record the activity of all the neurons in the VNC of the 6dpf animal. This section describes the advances I made towards this goal.

A.2.1 Activity indicators in 6dpf larvae

The calcium indicator GCaMP6s

Previous research in our lab (by Dr. Maria Tosches) showed that different versions of GCaMP could be used in *Platynereis*, as the protein was correctly expressed after microinjection, and the signal could be imaged and measured (Tosches et al., 2014). As this was shown only soon after injection (up to 48hpf), I first aimed to test whether the GCaMP was still active in these larvae on later developmental stages. Figure A.2 shows the results of some recordings of nechtochaete larvae, using a normal fluorescent wide field microscope. I could detect strong signal from the GCaMP, coming from both the muscles (Movies 1 and 2) and the neurons (Movie 3).

The imaging of the activity in the muscles is very interesting as it can be used to correlate specific movements with specific muscles, and understand the muscular units responsible for the stroke movements (see Results Section 3.3). However,

Phe	UUU	7763	Ser	UCU	8411	Tyr	UAU	5768	Cys	UGU	5596
	UUC	10832		UCC	8090		UAC	8689		UGC	5629
Leu	UUA	5319		UCA	9398	TER	UAA	1407	TER	UGA	2861
	UUG	10188		UCG	4446		UAG	1163	Trp	UGG	6680
	CUU	6991	Pro	CCU	9040	His	CAU	6486	Arg	CGU	3337
	CUC	8351		CCC	8339		CAC	7627		CGC	3809
	CUA	3738		CCA	9157	Gln	CAA	10924		CGA	4541
	CUG	11191		CCG	3933		CAG	13975		CGG	2933
Ile	AUU	9251	Thr	ACU	7856	Asn	AAU	9887	Ser	AGU	8133
	AUC	10959		ACC	7897		AAC	12741		AGC	7933
	AUA	5183		ACA	9489	Lys	AAA	14715	Arg	AGA	10205
Met	AUG	13924		ACG	5639		AAG	16755		AGG	6717
Val	GUU	7441	Ala	GCU	10274	Asp	GAU	12010	Gly	GGU	6992
	GUC	9218		GCC	10496		GAC	14423		GGC	8037
	GUA	4517		GCA	8193	Glu	GAA	16351		GGA	11617
	GUG	9851		GCG	3689		GAG	15015		GGG	4773

TABLE A.1: Codon usage in *Platynereis*.
Based on 526793 codons from highly expressed genes. [Data from Dr. Tomas Larsson].

the big dimensions of the muscles and their high brightness due to strong calcium waves, makes the signal form the neurons very dim in comparison, making it problematic for the imaging. To overcome this, Dr. Tosches and Dr. Bertucci in the lab cloned several copies of the miR-1 binding site downstream the GCaMP6s. As miR-1 is specifically expressed in the muscles acting as a repressor microRNA, the signal in the muscles was reduced considerably, although not completely eliminated. This modified version of the GCaMP6s was the one used in the experiments and results described next.

The voltage indicator ArcLight

As a way of recording direct voltage variations in the neurons, I aimed to use ArcLight, which offers higher temporal sensitivity than GCaMP6s and even single spiking events can be detected. In order to do that, we generated a nucleotide sequence coding for ArcLight, optimized for *Platynereis* codon usage. Table A.1 summarises the result of the analysis on *Platynereis* codon usage from highly expressed genes, done by Dr. Tomas Larsson.

A 1466 nucleotide sequence was generated (synthetic gene synthesis by Eurofins®) based on these results, including two unique restriction sites for subcloning it in different expression vectors. I produced the capped mRNA for ArcLight and injected it into zygotes. I used confocal imaging to record from different sections of the brain and neuropil, but no fluorescent signal was detected at 24hpf or 48hpf animals. Given that the results using GCaMP6s were more promising, I did not experiment further with the ArcLight.

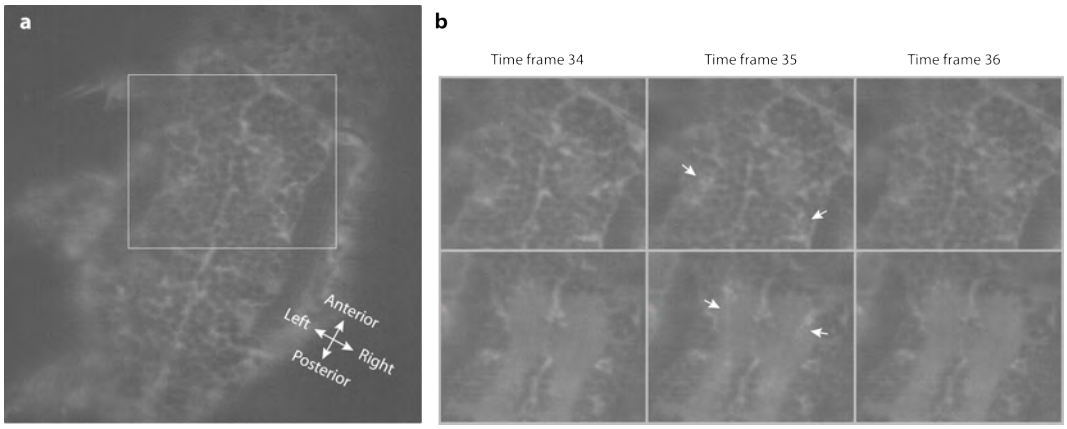

FIGURE A.3: SPIM can capture single-cell CCaMP6s-recorded activity.
(a) Single plane of one SPIM movie on the VNC of *Platynereis* at 6dpf. White rectangle indicates the area depicted in **b**. Orientation of the larvae is indicated. **(b)** Activity along three consecutive movie frames (columns). Top row: maximum projection covering the neuronal cell bodies. Bottom row: maximum projection covering the neuropil. The channels for the variation of fluorescence and the membrane signal are overlapped.

A.2.2 Using SPIM to image the VNC

In order to record the fast GCaMP6s dynamics in the entire VNC, that roughly corresponds to a block of tissue of 200x100x50 μm, and contains around two thousand neurons, we aimed to use light-sheet microscopy (SPIM), given its advantages as a fast and non-toxic approach, capable of covering this volume with enough resolution to distinguish single neurons. Towards this goal, I experimented with a couple of microscopes, one from Dr. Lars Hufnagel group in EMBL Heidelberg, for which I had the help of Gustavo Quintas Glasner de Medeiros, and the commercial SPIM from Zeiss®. GCaMP6s-injected larvae were mounted in 2% agar capillaries, and imaged in natural sea water with 500μM of NMDA and 5-HT to elicit locomotion movements (see Results section 3.4.6). The VNC was recorded at high speed (below 1 second) in several sections (8-20 depending on the video) using a ventral view[1].

I experimented with different conditions and acquired several videos that rendered terabytes of data. The analysis of this multidimensional data is not easy, especially using a discovery approach when the informative patterns to look for are unknown. One of the methods I used to find these patterns by image analysis followed the protocol in Ahrens et al., 2013, for which the images are registered across time to eliminate the movement from the video. A potential follow-up of this protocol is to use the analysis platform described in Freeman et al., 2014, that can find interesting correlations across the activity patterns of the neurons. This part of the project is still in a very preliminary stage (see also next section), but from the initial visual

[1] I had to find a compromise between speed and sampling, and experimenting with these conditions, each video had different parameters.

inspection of the SPIM videos, I made some observations that are worth mentioning to guide future experiments in this area:

- It is possible to detect single-cell activity using SPIM in *Platynereis*. Many instances of fluorescence variation associated to cell bodies were detected. In many cases these events occurred simultaneously in a bilateral fashion and signal across the neuropil was detected, indicating a specific cell type response and command (see Figure A.3).
- Very often along all the movies, a general VNC activity was recorded, and an increase of fluorescence was measured almost everywhere. This could indicate attempts of the larvae to scape from the probably uncomfortable agar embedding. These events were associated with general muscle contraction and poor performance of registration of those frames, that could affect the image analysis.
- The majority of the fluorescence variation was detected in the neuropil and not so much in the cell bodies. This could indicate local (dendritic) computations of the neurons, or could be due to the poor signal to noise ratio of these experiments. To solve this, I attempted to produce a stable transgenic GCaMP6s line in *Platynereis* (see next section).

A.2.3 Knock-in of GCaMP6 using Crispr-Cas9

In order to improve the signal to noise ratio in future experiments and eliminate the need of microinjection, I aimed to generate a stable line of pan-neuronally expressed GCaMP6s. To achieve the knock-in, I used the Crispr-Cas9 system, using long homology arms, to insert the coding DNA sequence (CDS) of GCaMP6s downstream a short piece of ElaV promoter. The strategy was to deliver the vector as a circular plasmid that would be cleaved by the same gRNA-Cas9 complex that would cleave the genome of *Platynereis* (Figure A.4c). The cleaved product, consisting on both homology arms surrounding the ElaV promoter and GCaMP6s CDS, would integrate by homologous recombination in the selected genomic place, using the regions homologous of the arms, that were adjacent to the Crispr cutting side in the genome (Figure A.4a).

I designed the Crispr guide RNA (ElaV gRNA) to be directed far downstream but near the endogenous ElaV locus, hoping for the inserted promoter to have access to the correct transcription machinery and epigenetic modifications. I made sure that no coding sequence of any other gene was near the insert. In order to test the efficiency of the ElaV gRNA I sequenced different clones as described in Results section 3.4.2. Figure A.4 shows that the efficiency ElaV gRNA is considerably lower than that seen for Dbx gRNA.

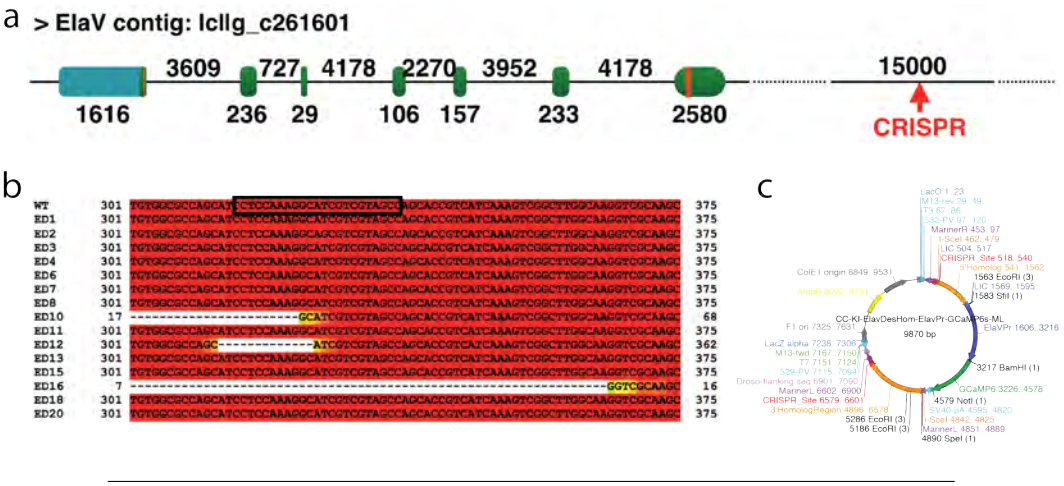

FIGURE A.4: Design strategy for ElaV::GCaMP6 Knock-In.
(a) Gene model of ElaV locus in *Platynereis* genome. The red arrow indicates where the ElaV gRNA directs the Cas9 to cut the DNA and recombine the insert. **(b)** Sequencing results to test the efficiency of ElaV gRNA. Black rectangle indicates the binding site. **(c)** Plasmid map of the vector used to deliver the ElaV::GCaMP6s construct.

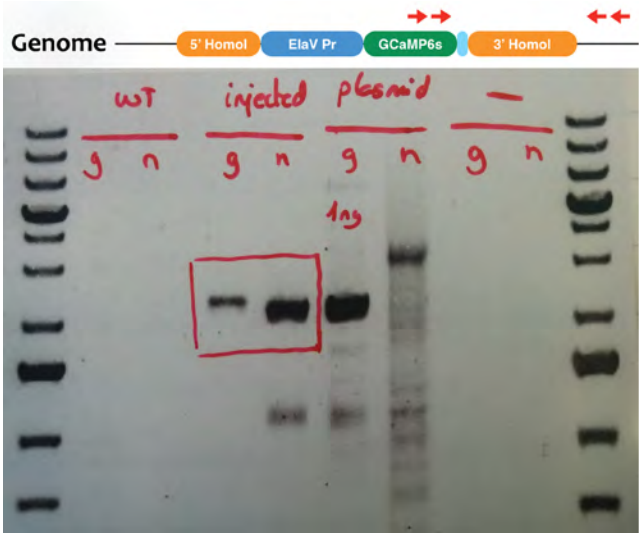

FIGURE A.5: ElaV::GCaMP6 Knock-in genotype.
On the top, illustration of the correct insertion of the construct in *Platynereis* genome, and the location of the primers used to genotype. On the bottom, a picture of the PCR gel, showing the results of the two PCRs (genomic and nested), on the WT and injected animals, and the plasmid and blank controls.

Nevertheless, as some mutants were observed, I tried to do the knock-in by microinjecting the ElaV gRNA, the Cas9 mRNA, and the circular vector in oocytes. After 48 hours, I extracted the genomic DNA from the larvae and amplified a 1800bp region using primers located in the genome (outside the homology arms) and in the GCaMP6s CDS. In principle, only correctly-inserted copies would produce the expected PCR product. Figure A.5 illustrates the result of this genotyping, showing a clear 1800bp band in the injected animals. In the PCR results for the plasmid control (PCR on 1ng of vector), there is a band of similar size on the genomic PCR, which cannot be explained by misalignment of primers on the vector.

These results were encouraging as they indicated that the knock-in was successful using the Crispr-Cas9 system. However, I could not reproduce the genotyping results in further attempts, and after growing the injected animals and crossing them

with wild types, no fluorescence was detected in the F1 generation. We are currently testing other Crispr-Cas9 knock-in strategies using higher efficiency gRNAs and shorter homology arms.

A.3 Crispr/Cas9 mediated knock down of other TFs

Besides of investigating the role of Dbx1 in *Platynereis* (see Results section 3.4), I aimed to assess the functional role of the neurons specified by those transcription factors (TFs) that in vertebrates specify the major locomotor-related cell types (see Introduction section 2.2.2). These transcription factors were selected as follows:

- Hb9 (Motoneuron population)
- Eve (V0v interneuron population)
- En (V1 interneuron population)
- Gata1/2/3 (V2b interneuron population)
- Sim (V3 interneuron population)

The transcription factor Chx10, that in vertebrates specifies the V2a population, very interesting for its rhythmic properties, was not included in the study because the gene model of this TF was not reconstructed due to poor sequencing data for its locus.

For each of these five TFs, I designed a Crispr-Cas9 gRNA, and tested their efficiency by the digestion method explained in Results section 3.4.1. Figure A.6 shows the results of the genotyping. None of the five gRNAs performed as efficiently as Dbx gRNA, but they exhibited different percentages of mutation. For instance, En and Gata1/2/3 nested PCRs showed medium and high resistance to the digestion, respectively, and Hb9 showed a shorter band in the PCR indicating the presence of deletions (Figure A.6a). The diffused band of Sim could also indicate mutations, but the band was digested by the enzyme.

Nevertheless, I sequenced some clones for each experiment, as genotype by digestion can under-represent the real mutation rate given that the digestion site can potentially be intact in a mutated locus. Indeed, higher mutation rates were detected for all the different TFs. Given the observed results, I examined the relative location of the restriction site with respect to the gRNA for each of the gRNAs used. I noticed a correlation of this position and the efficiency detected in the PCR digestion gel. The restriction enzymes sites for En, Gata1/2/3 and Dbx1, which are the ones showing higher efficiency, are located closer to the gRNA loop than the ones for Hb9, Eve and Sim, that showed higher digestion resistance, possibly caused by a mutation bias of the Cas9. This fact can be used in further gRNA designs to improve the mutant detection by PCR and digestion.

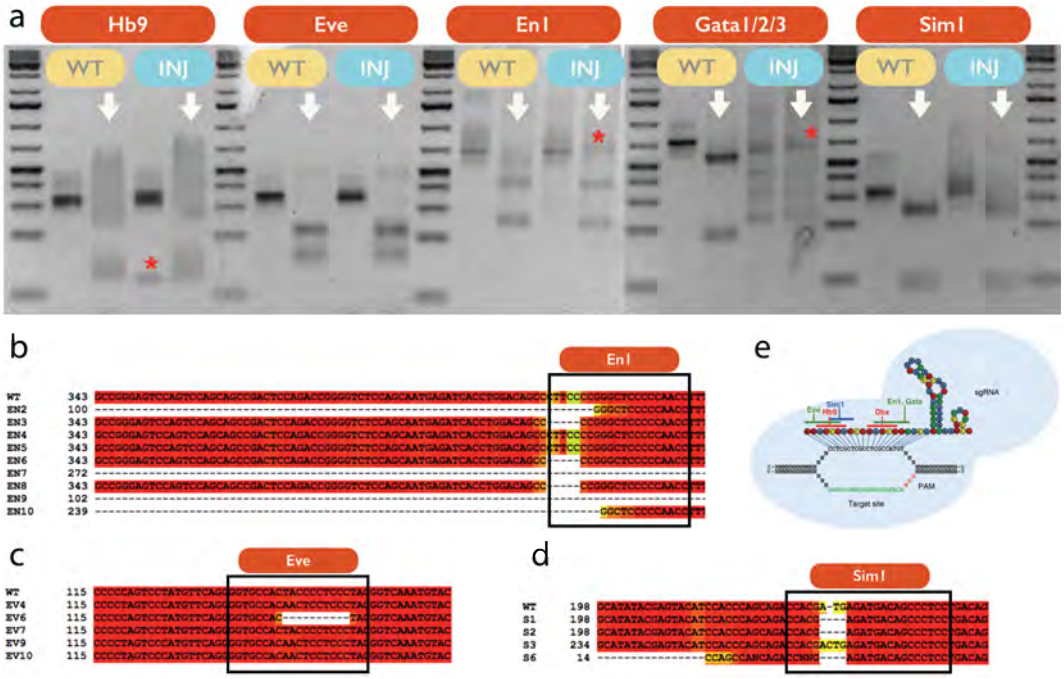

FIGURE A.6: Efficiency of Crispr-Cas9 for the TF specifying main vertebrate locomotor cell types.
(a) PCR and digestion genotyping results. PCR band and digested product showed for WT and injected animals for 5 different TFs. **(b-d)** Example of sequencing results for some mutants. **(e)** Position of the restriction sites with respect to the gRNA. [Illustration of Cas9 and gRNA reproduced with permission from Hwang et al., 2013].

The analysis of the phenotype of these mutants was not carried out because of a lack of antibodies for these transcription factors (currently in preparation) that could aid in the detection of any defects. Using PduLoco combined with individual genotyping to analyse any behavioural defect, is something planned for the near future.

A.4 Modulating neuronal activity

In order to understand the functional properties of nervous system, manipulating the activity of neuronal populations and even individual neurons represents a powerful approach, as one can gather evidence of how the precise activity patterns of specific components influences the output of the system. Modern advances towards this end have resulted in the development of transmembrane channels whose three dimensional structure can be reversibly modified by specific wave length, allowing the control of ions flux between neurons and extracellular space (and therefore depolarization and/or hyperpolarization of the membrane) by the simple application of light. Several of these proteins (activators and inhibitors of neural activity responding to several wave lengths) are nowadays available.

To test whether these approaches could be used in *Platynereis*, I cloned the ChannelRhodopsin (ChR2), a neuronal activator, in a expression vector and generated mRNA that I microinjected into fertilized oocytes. As ChR2 requires the light-isomerizable chromophore all-trans-retinal in order to work, I administered it in the natural sea water. I tested different concentrations of all-trans-retinal, in order to find a working condition that was not toxic for the animals, as proper development is affected at high concentrations. After determining the conditions, I tested how the illumination with 405nm wave length (blue) laser affected the behaviour of animals. The results were very encouraging.

Trochophores usually swim describing small spirals. This is a "default" state, and they only stop by closing the cilia, which requires active cholinergic excitation of the ciliary cells. Immediately after blue light illumination, injected animals stopped swimming, and remained still as long as blue light was on. Turning off the blue light resulted in the opposite response, and animals started swimming immediately (Figure A.7a). To directly investigate the behaviour of the ciliary cells, I immobilized one animal and recorded ciliary beating at high temporal resolution while switching the blue light on and off for random periods of time. Figure A.7 b and c show the high correlation between the ciliary movement and the ChannelRhodopsin activation, clearly indicating that the optogenetic manipulation of neurons in *Platynereis* is indeed possible and allows a very high precision in temporal control of neuronal activity.

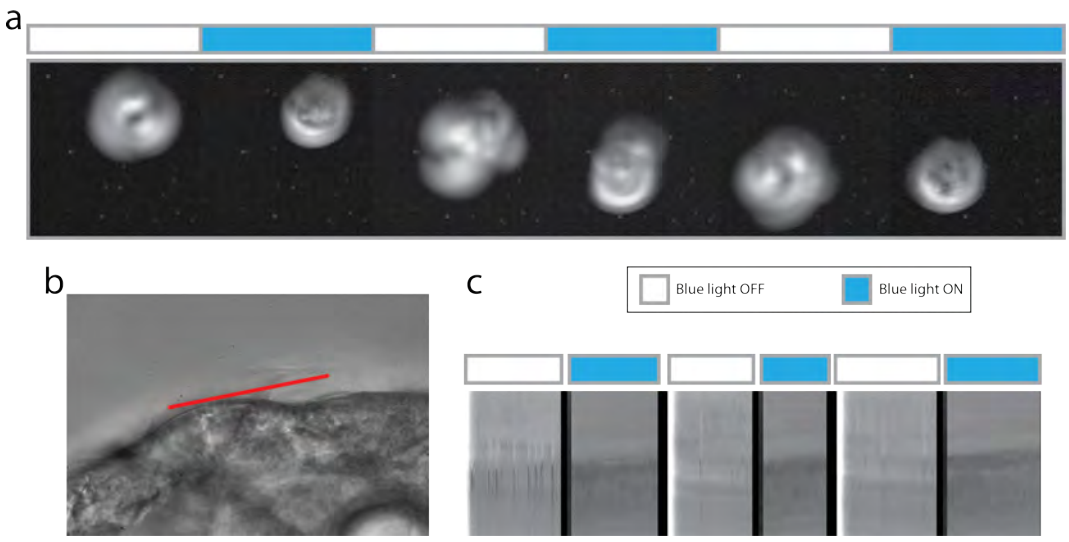

FIGURE A.7: Manipulating neuronal activity with ChR2 in *Platynereis*.
(a) Series of maximum projections showing the standard deviation of a subset of consecutive frames during a recording of 48hpf larvae swimming. Total duration of the movie: 18 seconds. **(b)** Frame of a high time resolution movie of the ciliary beating. Red line is used to plot the kymograph in **c**. **(c)** Kymograph showing the ciliary arrest when blue light is shined onto the larvae. Total time of the movie: 70 seconds.

A.4.1 Functional applications to cell types

In order to gain functional insight into specific neuronal populations, genetic drivers are needed to selectively express proteins like ChR2 in a group of neurons of known molecular profile. This allows the functional characterization of cell types by analysing their role within specific neuronal circuits. While genetic manipulations using specific promoters, knock-in constructs, and genetically-modified animal lines are routinely performed in established model organisms (*Caenorhabditis elegans*, *Drosophila*, mice and zebrafish), non-model organisms, common in the EvoDevo field, are far behind with respect to the implementation of these useful techniques. *Platynereis* is not an exception to this rule, and despite it is indeed possible to create stable transgenic lines (Backfisch et al., 2013), the lack of genetic drivers[2] makes difficult the functional assessment of cell-types.

With the aim of gaining a genetic access to locomotor-related cell types in *Platynereis*, I attempted to clone the promoters of these TFs whose orthologues specify the main CPG neuronal populations in vertebrates (see Section 2.2.2). Using the genomics and transcriptomics resources (set up by Dr. Larsson and Dr. Simakov), I managed to reconstruct the gene models for *hb9*, *sim*, *gata1/2/3*, *dbx1*, *en* and *eve*. Together with David Puga, we cloned from 2 to 5 kb of the region upstream the ATG for the majority of these genes.We used different constructs to test both the specificity and strength of these sequences in driving the expression of fluorescent proteins.

[2]The difference between established and non-established laboratory models in the availability of resources is mainly caused by the difference in the amount of scientist working with these organisms.

Despite we have been able to label specific neurons (data not shown), these constructs tend to not give any expression in the big majority of animals injected. This is probably due to the complex transcriptional regulation that these developmental TFs might be subject to, and therefore the small pieces cloned are not sufficient to recapitulate the endogenous expression. We are currently testing the possibility of using Crispr-Cas9-directed Knock-in (see Section A.2.3) to label the endogenous proteins, but this is still work in progress.

Appendix B

Commissural neurons in bilaterians

B.1 Location and morphology of commissural neurons

In the CNS, a neuron is referred as commissural if its axon crosses the midline (floor plate) of the neural tube (or neural plate in case of invertebrates), projecting contra-laterally. In case the axon of a neuron projects longitudinally in the same side that the soma of the cell is, the neuron is referred as ipsilateral. In vertebrates, all commissural neurons are interneurons. In invertebrates, motoneurons can be also commissural.

The soma of commissural neurons are widely distributed dorso-ventrally in the neural tube (Wentworth, 1984a). Some are located very close to the lateral funiculus, and others among ventral root motoneurons. They differentiate, at the beginning, from alar and basal plates. As other cells in the neural tube, they start to differentiate radially within the ventricular (inner) layer. As the somas reach the basal (outer) region of the neural tube, they become unipolar, retracting their apical protrusion. The basal protrusion converts into the axon of the neuron (Wentworth, 1984a; Wentworth, 1984b).

Together with ipsilateral neurons, commissural axons first project ventrally towards the floor plate. In this process, some commissural axons cross the dendrites of ventromedial motoneurons. The crossing of the first commissural neurons takes places during E10 and E11 stages in mice (Wentworth, 1984a). In a second step, ipsilateral and commissural axons respond differently to floor plate cues in order to cross it (commissural) or to project longitudinally (ipsilateral). After midline crossing, commissural axons are directed towards three different medio-lateral longitudinal paths by a combinatorial code of receptors (see below), and their axons turn and project only rostrally in some species (zebrafish, mouse, rat), and in both directions (rostrally and caudally) in others (chicken and *Xenopus*). In *Drosophila*, where there are two commissures per segment, a subset of neurons project ipsilaterally and pioneer the longitudinal fascicles scaffold by interactions with the glia. This initial

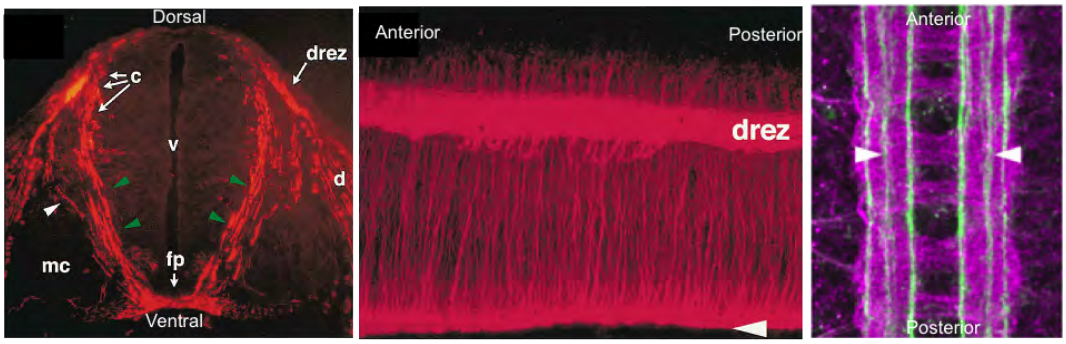

FIGURE B.1: Location of Tag-1$^+$ cells in the vertebrate neural tube.
On the left, cross section at the brachial level of the mice neural tube at stage E11.5 stained with 4D7 antibody to label Tag-1 commissural neurons and their axons (c). Green arrowheads point to commissural axons growing towards the floor plate. On the middle, sagital view of the spinal cord of similar embryos as before. fp, floor plate; v, ventricle; mc, motor column; d, dorsal root ganglion; drez, dorsal root entry zone. On the right, Dorsal view of stage 16 *Drosophila* neuropil (purple), counterstained with anti-FasII to visualize both the two commissures per segment and the three longitudinal fascicles. [Reproduced with permission from (Serafini et al., 1996) (mice images) and (Spitzweck, Brankatschk, and Dickson, 2010) (*Drosophila*)].

establishment of the paths is essential for later ipsilateral and commissural neurons to project longitudinally. In vertebrates, commissural neurons are located all along the anterio-posterior axis of the neural tube (Figure B.1).

The guidance of the growth cone is mediated by actin dynamics, and the elongation of the axon by tubulin polymerisation. Cytoskeletal regulation is house-keeping for every neuron, but signialling cues and receptor/cascade machineries are specific to neuron types, allowing them to respond and migrate differently (Sánchez-Soriano et al., 2007) (see below).

In vertebrates, a well characterized set of commissural neurons is specified in the most dorsal region of the neural tube. The projections of this population of interneurons (dl1) predominantly cross the midline (see Helms and Johnson 1998 in (Nawabi and Castellani, 2011)). Dl1, which is Lhx2+/Lhx9+, is derived from Math1 progenitors. At least some dl1 commissural neurons migrate to the dorsal deep horn (Saba, Nakatsuji, and Saito, 2003). There are several subtypes of commissural neurons, including GABAergic in rodents and glycinergic in the frog (Guan and Rao, 2003).

B.2 Commissural axon guidance

Commissural neurons has been a population of cells widely used in many organisms for studying the process of axon guidance. As a consequence, many molecules have been identified regulating the navigation of commissural axons. The process

that drives this axons to reach their final connectivity points is defined by a series of steps that requires a complex regulation of molecular mechanisms. First, axons need to travel ventrally and reach the midline. Second, they need to cross it. And third, they need to elongate towards their targets by finding the proper paths. Growth cones are the navigation systems of the axons, and are guided by attraction/repellent signals to reach checkpoints in their trip. Depending on the step, different molecules are present in the membrane of growth cones in order to properly respond to specific signalling cues present in the tissue. Transcriptional, post-transcriptional and post-translational regulation mechanisms have been demonstrated to play an important role in regulating the presence of guidance receptors in the membrane of growth cones.

Guidance cue	**Receptor**
Attractants	
Netrin	DCC, neogenin
Shh	Boc, Patch/Smo
NrCAM	Tag-1
Repellents	
Slit	Robo family
Ephrin-B3	EphB1-EphB3
Sema3B	Nrp2/PlexA1
BMP7	BMPRI/II

TABLE B.1: Compilation of known attractant and repellent signals and their respective receptors.

BMP7, expressed by the roof plate, dorsally to commissural neurons, might mediate their initial ventral projection, as netrin molecules have not been shown to be required for this step (Guan and Rao, 2003), but as short-range attractants (Figure B.2A).

In both vertebrates and *Drosophila*, Robo receptors in the growth cones play an important role in regulating the invasion and exit of and from the floor plate. There are three known Robo receptors, and they respond differently to Slit signalling emanating from the floor plate. Generally, in pre-crossing growth cones, Robo receptors are sequestered from the membrane so a Slit-mediated repulsion does not occur (Figure B.2B and C). Their regulation and activity, however, differ in vertebrates and flies.

In *Drosophila*, Robo1 is the key receptor for the decision of midline crossing. It is expressed both in ipsilateral and commissural neurons, but degraded only in commissural by Commissureless (Comm). The down-regulation of Robo1 in commissural axons causes their unresponsive behaviour to midline Slit, allowing them to cross the floor plate (Spitzweck, Brankatschk, and Dickson, 2010). Robo1 prevents midline crossing whereas Robo2 promotes it in commissural neurons and helps preventing the crossing of ipsilateral axons. Robo3 is expressed later and important for medio-lateral specification of longitudinal fascicles together with Robo2 (see below).

In vertebrates, splicing regulation of Robo3 mRNA produces two isoforms (Robo3.1 and Robo3.2) with opposite activities. Robo3.1 is expressed at the pre-crossing stage and induces (through a yet unknown mechanism) the blockage of the Slit-mediated repulsion (Chen et al., 2008), whereas Robo1, Robo2 and Robo3.2 are expressed the postcrossing stage (Nawabi and Castellani, 2011; Long et al., 2004). Comm and Robo3 (Robo3.1) might regulate the initial levels of Robo protein without shutting them completely, and signals from the floor plate could further regulate (block completely) Robo levels, in a positive feedback loop (Dickson and Gilestro, 2006).

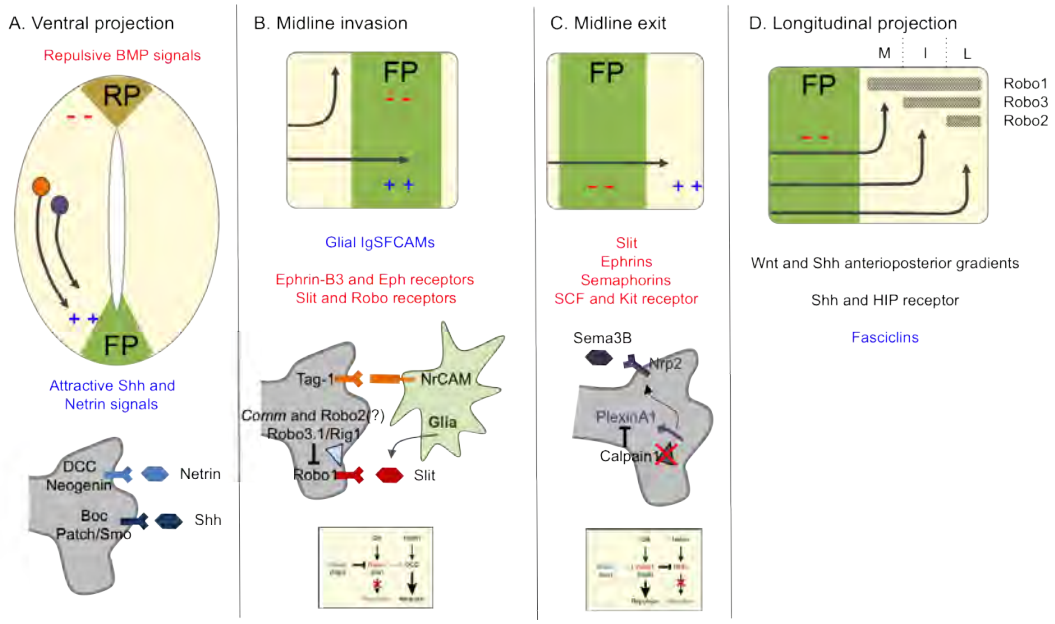

FIGURE B.2: Summary of guidance signals for commissural neurons.
Illustration of the different steps during commissural axon guidance and the interaction between the known molecules orchestrating this complex process. In blue, attractive signals. In red, repulsive signals. RP, roof plate; FP, floor plate. [Adapted from (Nawabi and Castellani, 2011), (Dickson and Gilestro, 2006), (Sánchez-Soriano et al., 2007) and (Spitzweck, Brankatschk, and Dickson, 2010)].

As many reports point out, Robo are not the only receptors involved in floor plate exit. An additional molecular mechanism to regulate attractive/repulsive responses

is through DCC (Frazzled in *Drosophila*), the Netrin receptor. During the crossing step, DCC is silenced by forming a complex with Robo1 in the presence of Slit midline signalling (Figure B.2B and C). Response to Class 3 Semaphorins (ex. Sema3B) is mediated by holoreceptors composed by a member of Neuropilins (ex. Nrp2), which bind the ligands, and a member co-receptor of the Plexin-A family (ex. Plexin-A1), which triggers the signal. Cell Adhesion Molecules (CAM) (ex. NrCAM) in the floor plate play an active role as contact signals and checkpoints for specifying the timing and regulation of the Plexin-A receptors in order to trigger a repulsive response. Specifically, precrossing commissural axons remain unresponsive to Sema3B cues due to the lack of Plexin-A1 receptor in the membrane, which is expressed and translated but cleaved by the calcium dependent protease calpain-1. It's co-receptor Nrp2 is not prevented from exposure. During the floor plate crossing step, glial NrCAM is involved in the repression of calpain-1 activity, enabling the accumulation of Plexin-A1 in the membrane therefore triggering the response to Sema3B and preventing midline recrossing (Nawabi et al., 2010) (Figure B.2C).

In *Drosophila*, contralateral motoneurons RP1 and RP3 depend on the contralateral homologue somata as a checkpoint in their path towards the neuropil exit point (Sánchez-Soriano et al., 2007). Semaphorins have been shown not to be required for commissural axon guidance in *Drosophila* (Long et al., 2004). Other cues involved in preventing commissural axons from floor plate multi-crossing are the Ephrins and the B class Eph receptors expressed in post-crossing segments (Long et al., 2004).

Once the axons have crossed the floor plate, they need to find the correct path to project longitudinally (fascicle). Both in vertebrates and flies, axons are sorted into three medio-lateral areas (Figure B.2D). In *Drosophila*, axons that express only Robo1 project in the most medial zone. Those co-expressing Robo1 and Robo3 are located in the intermediate zone. Those expressing the three receptors project more laterally. Robo1 is expressed in the three areas, but it does not contribute to the formation of these three discrete zones. (Spitzweck, Brankatschk, and Dickson, 2010). The definition of the medio-lateral zones was long thought to be mediated by the combinatorial membrane expression of Robo receptors and their response to Slit (Long et al., 2004; Sánchez-Soriano et al., 2007; Dickson and Gilestro, 2006). However, Spitzweck, Brankatschk, and Dickson, 2010 demonstrated that the fascicle choice was only dependent on gene expression patterns and/or total amount of Robo protein but not on structural differences between the Robo receptors. Fasciclins have also been shown to play a role as contact-dependent fasciculation signals. Both long and short range (adhesion molecules) interactions are involved in driving commissural and ipsilateral axons to their targets by specifying the medio-lateral arrangement of longitudinal fascicles.

B.3 Specification of commissural neurons

Expression domains in the dorsal neural tube are specified by bHLH transcription factors Ngn1, Math1 and Mash1. Math1 progenitors gives rise to interneurons expressing Lhx2-A/B, Lhx9 and Brn3a. Ngn1 progenitor domain specifies cells that co-express Lim1/2 and Brn3a (Gowan et al., 2001). Dl1 cells are specified by the HLH transcription factor Math1. Some commissural neurons must be specified somewhere else, as they are still present in Math1-/- mice (Gowan et al., 2001). Downstream Math1 acts Mhb1, which is a repressor necessary and sufficient for the specification of commissural interneurons (Saba, Johnson, and Saito, 2005). The genes repressed by Mhb1 have not been identified. Its repression activity promotes the expression of Tag1 and DCC. Some commissural neurons are generated downstream of Ngn2 (Simmons et al., 2001). LIM transcription factors (Lhx2 and Lhx9) control the expression of receptor Robo3 (also known as Rig1) in dl1-commissural neurons (Wilson et al., 2008). Lbx1 is a repressor of commissural neurons since Lbx1-/- mice generates an excess of them because of mis-specification of dorsal interneurons (Gross, Dottori, and Goulding, 2002).

Bibliography

Achim, K. et al. (2015). "High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin." *Nat. Biotechnol.* 33.5, pp. 503–9.

Ackermann, C., A. Dorresteijn, and A. Fischer (2005). "Clonal domains in postlarval Platynereis dumerilii (Annelida: Polychaeta)". *J Morphol* 266.3, pp. 258–280.

Ahrens, M. B. et al. (2013). "Whole-brain functional imaging at cellular resolution using light-sheet microscopy." *Nat. Methods* 10.5, pp. 413–20.

Alaynick, W. A., T. M. Jessell, and S. L. Pfaff (2011). "Snapshot: Spinal Cord Development". *Cell* 146.

Arendt, D and K Nübler-Jung (1999). "Comparison of early nerve cord development in insects and vertebrates." *Development* 126.11, pp. 2309–25.

Arendt, D. (2008). "The evolution of cell types in animals: emerging principles from molecular studies." *Nature Reviews Genetics* 9.11, pp. 868–82.

Arendt, D., M. Tosches, and H. Marlow (2016). "From nerve net to nerve ring, nerve cord and brain — evolution of the nervous system". *Nature Reviews Neuroscience* 17.1, pp. 61–72.

Arendt, D. et al. (2008). "The evolution of nervous system centralization." 363.1496, pp. 1523–8.

Asadulina, A. et al. (2012). "Whole-body gene expression pattern registration in Platynereis larvae." *Evodevo* 3.1, p. 27.

Backfisch, B. et al. (2013). "Stable transgenesis in the marine annelid Platynereis dumerilii sheds new light on photoreceptor evolution." *PNAS* 110.1, pp. 193–8.

Bagnall, M. W. and D. L. McLean (2014). "Modular organization of axial microcircuits in zebrafish." *Science* 343.6167, pp. 197–200.

Barrio, M. et al. (2007). "A regulatory network involving Foxn4, Mash1 and delta-like 4/Notch1 generates V2a and V2b spinal interneurons from a common progenitor pool". *Development* 134.19, pp. 3427–3436.

Batista, M. F., J. Jacobstein, and K. E. Lewis (2008). "Zebrafish V2 cells develop into excitatory CiD and Notch signalling dependent inhibitory VeLD interneurons." *Dev. Biol.* 322.2, pp. 263–75.

Bem, T. et al. (2003). "From swimming to walking: a single basic network for two different behaviors". *Biol. Cybern.* 88.79.

Bikoff, J. B. et al. (2016). "Spinal Inhibitory Interneuron Diversity Delineates Variant Motor Microcircuits". *Cell* 165.1, pp. 207–19.

Briscoe, J. et al. (2000). "A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube". *Cell* 101.4, pp. 435–445.

Büschges, A. et al. (2008). "Organizing network action for locomotion: Insights from studying insect walking". *Brain Res Rev* 57.1, p. 162171.

Cacciatore, T., R Rozenshteyn, and W. Kristan (2000). "Kinematics and modeling of leech crawling: evidence for an oscillatory behavior produced by propagating waves of excitation." *The journal of Neuroscience* 20.4, pp. 1643–55.

Cang, J. and W. Friesen (2002). "Model for intersegmental coordination of leech swimming: central and sensory mechanisms." *J. Neurophysiol.* 87.6, pp. 2760–9.

Cardona, A. et al. (2012). "TrakEM2 software for neural circuit reconstruction". *PLoS ONE* 7.6, e38011.

Chen, Z. et al. (2008). "Alternative splicing of the Robo3 axon guidance receptor governs the midline switch from attraction to repulsion." *Neuron* 58.3, pp. 325–32.

Christodoulou, F. et al. (2010). "Ancient animal microRNAs and the evolution of tissue identity." *Nature* 463.7284, pp. 1084–8.

Crone, S. et al. (2008). "Genetic ablation of V2a ipsilateral interneurons disrupts left-right locomotor coordination in mammalian spinal cord". *Neuron* 60, pp. 70–83.

Daly, J. M. (1973). "Behavioural and Secretory Activity during Tube construction by Platynereis dumerilii Aud & M. Edw. [Polychaeta: Nereidae]". *J. mas. Bio. Ass. UK* 53.03, p. 521.

Dasen, J. S. et al. (2008). "Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1." *Cell* 134.2, pp. 304–16.

Demilly, A. et al. (2011). "Coe genes are expressed in differentiating neurons in the central nervous system of protostomes." *PLoS ONE* 6.6, e21213.

Demilly, A. et al. (2013). "Involvement of the Wnt/-catenin pathway in neurectoderm architecture in Platynereis dumerilii." *Nat Commun* 4, p. 1915.

Denes, A. S. (2007). "The evolution of dorsoventral patterning and of neuron types in the trunk nervous system of bilaterian animals". PhD thesis. University of Heidelberg.

Denes, A. S. et al. (2007). "Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria." *Cell* 129.2, pp. 277–88.

Dickson, B. J. and G. F. Gilestro (2006). "Regulation of commissural axon pathfinding by slit and its Robo receptors." *Annu. Rev. CellDev. Biol.* 22, pp. 651–75.

Dougherty, K. J. et al. (2013). "Locomotor rhythm generation linked to the output of spinal shox2 excitatory interneurons." *Neuron* 80.4, pp. 920–33.

Fetcho, J. R. and D. L. McLean (2010). "Some principles of organization of spinal neurons underlying locomotion in zebrafish and their implications." *Ann Ny Acad Sci* 1198, pp. 94–104.

Finci, L. I. et al. (2014). "The crystal structure of netrin-1 in complex with DCC reveals the bifunctionality of netrin-1 as a guidance cue." *Neuron* 83.4, pp. 839–49.

Fischer, A. (2010). "Mesoderm formation and muscle development of Platynereis dumerilii (Nereididae, Annelida)". PhD thesis. University of Berlin.

Fischer, A. H., T. Henrich, and D. Arendt (2010). "The normal development of Platynereis dumerilii (Nereididae, Annelida)." *Front Zool* 7, p. 31.

Francius, C. et al. (2015). "Genetic dissection of Gata2 selective functions during specification of V2 interneurons in the developing spinal cord". *Dev Neurobiol* 75.7, pp. 721–737.

Freeman, J. et al. (2014). "Mapping brain activity at scale with cluster computing." *Nat. Methods*.

Friesen, O. W. and W. B. Kristan (2007). "Leech locomotion: swimming, crawling, and decisions". *Curr. Opin. Neurobiol.* 17.6, p. 704711.

Garcia-Campmany, L., F. J. Stam, and M. Goulding (2010). "From circuits to behaviour: motor networks in vertebrates." *Curr. Opin. Neurobiol.* 20.1, pp. 116–25.

Gomez-Marin, A. et al. (2012). "Automated tracking of animal posture and movement during exploration and sensory orientation behaviors." *PLoS ONE* 7.8.

Goulding, M (2012). "Motor Neurons that Multitask." *Neuron* 76.4, pp. 669–70.

Goulding, M. (2009). "Circuits controlling vertebrate locomotion: moving in a new direction." *Nature Reviews Neuroscience* 10.7, pp. 507–18.

Gouti, M., V. Metzis, and J. Briscoe (2015). "The route to spinal cord cell types: a tale of signals and switches". *Trends Genetics* 31.6, pp. 282–289.

Gowan, K et al. (2001). "Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons." *Neuron* 31.2, pp. 219–32.

Gray, P (2008). "Transcription factors and the genetic organization of brain stem respiratory neurons". *J Appl Physiol* 104.5, pp. 1513–1521.

Gray, P (2013). "Transcription factors define the neuroanatomical organization of the medullary reticular formation". *Frontiers in Neuroanatomy* 7.7.

Gribble, S. L., B. O. Nikolaus, and R. I. Dorsky (2007). "Regulation and function of Dbx genes in the zebrafish spinal cord". *Developmental Dynamics* 236.12, pp. 3472–3483.

Grillner, S. (2003). "The motor infrastructure: from ion channels to neuronal networks." *Nature Reviews Neuroscience* 4.7, pp. 573–86.

Grillner, S. and T. M. Jessell (2009). "Measured motion: searching for simplicity in spinal locomotor networks." *Curr. Opin. Neurobiol.* 19.6, pp. 572–86.

Gross, M., M. Dottori, and M. Goulding (2002). "Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord." *Neuron* 34.4, pp. 535–49.

Guan, K. L. and Y. Rao (2003). "Signalling mechanisms mediating neuronal responses to guidance cues." *Nature Reviews Neuroscience* 4.12, pp. 941–56.

Guertin, P. A. (2009). "The mammalian central pattern generator for locomotion." *Brain Res Rev* 62.1, pp. 45–56.

Guhmann, M et al. (2015). "Spectral Tuning of Phototaxis by a Go-Opsin in the Rhabdomeric Eyes of Platynereis". *Curr. Biol.* 17.31, pp. 2265–2271.

Hägglund, M. et al. (2013). "Optogenetic dissection reveals multiple rhythmogenic modules underlying locomotion." *PNAS* 110.28, pp. 11589–94.

Heckscher, E. et al. (2014). "Atlas-builder software and the eNeuro atlas: resources for developmental biology and neuroscience". *Development* 141.12, p. 25242532.

Helmstaedter, M., K. L. Briggman, and W. Denk (2011). "High-accuracy neurite reconstruction for high-throughput neuroanatomy." *Nature Neuroscience* 14.8, pp. 1081–8.

Helmstaedter, M. et al. (2013). "Connectomic reconstruction of the inner plexiform layer in the mouse retina." *Nature* 500.7461, pp. 168–74.

Hirth, F. et al. (2003). "An urbilaterian origin of the tripartite brain: developmental genetic insights from Drosophila." *Development* 130.11, pp. 2365–73.

Holland, L. Z. et al. (2013). "Evolution of bilaterian central nervous systems: a single origin?" 4.1, p. 27.

Holland, N. D. (2016). "Nervous systems and scenarios for the invertebrate-to-vertebrate transition". 371.1685, p. 20150047.

Hwang, W. Y. et al. (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." *Nat. Biotechnol.* 31.3, pp. 227–9.

Joshi, K. et al. (2009). "LMO4 Controls the Balance between Excitatory and Inhibitory Spinal V2 Interneurons". *Neuron* 61.6, pp. 839–851.

Juárez-Morales, J. et al. (2016). "Evx1 and Evx2 specify excitatory neurotransmitter fates and suppress inhibitory fates through a Pax2-independent mechanism". *Neural Dev* 11.1.

Kammermeier, L. et al. (2001). "Differential expression and function of the DrosophilaPax6 genes eyeless and twin of eyeless in embryonic central nervous system development". *Mechanisms of development* 103, pp. 71–78.

Kicheva, A. and J. Briscoe (2015). "Developmental Pattern Formation in Phases". *Trends in Cell Biology* 25.10, pp. 579–591.

Kiehn, O. (2006). "Locomotor circuits in the mammalian spinal cord." *Annual Reviews Neuroscience* 29, pp. 279–306.

Kiehn, O. (2011). "Development and functional organization of spinal locomotor circuits." *Curr. Opin. Neurobiol.* 21.1, pp. 100–9.

Kimura, Y., C. Satou, and S. Higashijima (2008). "V2a and V2b neurons are generated by the final divisions of pair-producing progenitors in the zebrafish spinal cord." *Development* 135.18, pp. 3001–5.

Kimura, Y. et al. (2013). "Hindbrain V2a neurons in the excitation of spinal locomotor circuits during zebrafish swimming." *Curr Biology Cb* 23.10, pp. 843–9.

Kozlov, A. et al. (2009). "Simple cellular and network control principles govern complex patterns of motor behavior." *PNAS* 106.47, pp. 20027–32.

Kreshuk, A. et al. (2014). "Automated detection of synapses in serial section transmission electron microscopy image stacks." *PLoS ONE* 9.2, e87351.

Kristan, W. B., R. L. Calabrese, and O. W. Friesen (2005). "Neuronal control of leech behavior". *Progress in Neurobiology* 76.5, p. 279327.

Kulakova, M. et al. (2007). "Hox gene expression in larval development of the polychaetes Nereis virens and Platynereis dumerilii (Annelida, Lophotrochozoa)." *Dev Genes Evol* 217.1, pp. 39–54.

Kutejova, E. et al. (2016). "Neural Progenitors Adopt Specific Identities by Directly Repressing All Alternative Progenitor Transcriptional Programs". *Developmental Cell* 36.6, pp. 639–653.

Lacin, H et al. (2014). "Transcription factor expression uniquely identifies most postembryonic neuronal lineages in the Drosophila thoracic central nervous system". *Development* 141.5.

Lacin, H. et al. (2009). "dbx mediates neuronal specification and differentiation through cross-repressive, lineage-specific interactions with eve and hb9." *Development* 136.19, pp. 3257–66.

Lanuza, G. M. et al. (2004). "Genetic Identification of Spinal Interneurons that Coordinate Left-Right Locomotor Activity Necessary for Walking Movements". *Neuron* 42.3.

Lauri, A. et al. (2014). "Development of the annelid axochord: Insights into notochord evolution". *Science* 345.6202, pp. 1365–1368.

Ljunggren, E. et al. (2014). "Optogenetic activation of excitatory premotor interneurons is sufficient to generate coordinated locomotor activity in larval zebrafish". *The journal of Neuroscience* 34.1, pp. 134–139.

Long, H. et al. (2004). "Conserved roles for Slit and Robo proteins in midline commissural axon guidance." *Neuron* 42.2, pp. 213–23.

Lu, D. C., T. Niu, and W. A. Alaynick (2015). "Molecular and cellular development of spinal cord locomotor circuitry". *Front Molecular Neuroscience* 8.

Maco, B. et al. (2014). "Semiautomated correlative 3D electron microscopy of in vivo-imaged axons and dendrites." *Nat Protoc* 9.6, pp. 1354–66.

Marlow, H. et al. (2014). "Larval body patterning and apical organs are conserved in animal evolution". *BMC Biology* 12.1, p. 7.

McCrea, D. A. and I. A. Rybak (2008). "Organization of mammalian locomotor rhythm and pattern generation." *Brain Res Rev* 57.1, pp. 134–46.

Montgomery, J. E. et al. (2016). "Intraspinal serotonergic neurons consist of two, temporally distinct populations in developing zebrafish". *Developmental Neurobiology* 76, pp. 673–687.

Morlot, C. et al. (2007). "Structural insights into the Slit-Robo complex". *PNAS* 104.38, pp. 14923–14928.

Morris, S. C. and J. S. Peel (2008). "The earliers annelids: Lower Cambrian polychaetes from the Sirius Passet Lagerstatte, Peary Land, North Greenland". *Acta Palaeontologica Polonica* 53.1, pp. 137–148.

Mulloney, B. and W. M. Hall (2007). "Local and Intersegmental Interactions of Coordinating Neurons and Local Circuits in the Swimmeret System". *J Neurophysiology* 98.1, pp. 405–413.

Nawabi, H. and V. Castellani (2011). "Axonal commissures in the central nervous system: how to cross the midline?" *Cellular and Molecular Life Sciences* 68.15, pp. 2539–53.

Nawabi, H. et al. (2010). "A midline switch of receptor processing regulates commissural axon guidance in vertebrates." *Genes Development* 24.4, pp. 396–410.

Orlovsky, G., T. Deliagina, and S Grillner (1999). *Neuronal Control of Locomotion. From Mollusc to Man*. Oxford University Press.

Petracca, Y et al. (2016). "The late and dual origin of cerebrospinal fluid-contacting neurons in the mouse spinal cord". *Development* 143.5, pp. 880–891.

Pfeifer, K., A. W. Dorresteijn, and A. C. Fröbius (2012). "Activation of Hox genes during caudal regeneration of the polychaete annelid Platynereis dumerilii." *Dev Genes Evol* 222.3, pp. 165–79.

Pierani, A et al. (2001). "Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1." *Neuron* 29.2, pp. 367–84.

Pinheiro, P., M. Gering, and R. Patient (2004). "The basic helix-loop-helix transcription factor, Tal2, marks the lateral floor plate of the spinal cord in zebrafish." *Gene Expr Patterns Gep* 4.1, pp. 85–92.

Raible, F. and D. Arendt (2004). "Metazoan Evolution: Some Animals Are More Equal than Others". *Curr. Biol.* 14.3, R106–R108.

Raible, F. et al. (2005). "Vertebrate-type intron-rich genes in the marine annelid Platynereis dumerilii." *Science* 310.5752, pp. 1325–6.

Randel, N et al. (2014). "Neuronal connectome of a sensory-motor circuit for visual navigation". *Elife* 3, e02730.

Randel, N. et al. (2015). "Inter-individual stereotypy of the Platynereis larval visual connectome." *Elife* 4, e08069.

Reichert, H. (2005). "A tripartite organization of the urbilaterian brain: developmental genetic evidence from Drosophila." *Brain Res Bull* 66.4-6, pp. 491–4.

Roberts, A., W. Li, and S. R. Soffe (2012). "A functional scaffold of CNS neurons for the vertebrates: the developing Xenopus laevis spinal cord." *Developmental Neurobiology* 72.4, pp. 575–84.

Roberts, A., W. C. Li, and S. R. Soffe (2010). "How neurons generate behavior in a hatchling amphibian tadpole: an outline." *Front Behav Neurosci* 4, p. 16.

Rybak, I., K. Dougherty, and N. Shevtsova (2015). "Organization of the Mammalian Locomotor CPG: Review of Computational Model and Circuit Architectures Based on Genetically Identified Spinal Interneurons". *eNeuro* 2.5.

Saba, R., J. E. Johnson, and T. Saito (2005). "Commissural neuron identity is specified by a homeodomain protein, Mbh1, that is directly downstream of Math1." *Development* 132.9, pp. 2147–55.

Saba, R., N. Nakatsuji, and T. Saito (2003). "Mammalian BarH1 confers commissural neuron identity on dorsal cells in the spinal cord." *The journal of Neuroscience* 23.6, pp. 1987–91.

Sánchez-Soriano, N. et al. (2007). "Drosophila as a genetic and cellular model for studies on axonal growth." *Neural Dev* 2, p. 9.

Sarvestani, I. et al. (2012). "A computational model of visually guided locomotion in lamprey." *Biol. Cybern.*

Saudemont, A. et al. (2008). "Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid Platynereis." *Dev Biol* 317.2, pp. 430–43.

Serafini, T et al. (1996). "Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system." *Cell* 87.6, pp. 1001–14.

Seredick, S. et al. (2014). "Lhx3 and Lhx4 suppress Kolmer–Agduhr interneuron characteristics within zebrafish axial motoneurons". *Development* 141.20, pp. 3900–3909.

Simionato, E. et al. (2008). "atonal- and achaete-scute-related genes in the annelid Platynereis dumerilii: insights into the evolution of neural basic-Helix-Loop-Helix genes." *Bmc Evol Biol* 8, p. 170.

Simmons, A. et al. (2001). "Neurogenin2 Expression in Ventral and Dorsal Spinal Neural Tube Progenitor Cells Is Regulated by Distinct Enhancers". *Dev Biol* 229.2, pp. 327–339.

Smarandache, C., W. M. Hall, and B. Mulloney (2009). "Coordination of rhythmic motor activity by gradients of synaptic strength in a neural circuit that couples modular neural oscillators." *The journal of Neuroscience* 29.29, pp. 9351–60.

Smarandache-Wellmann, C., C. Weller, and B. Mulloney (2014). "Mechanisms of coordination in distributed neural circuits: decoding and integration of coordinating information." *J. Neurosci.* 34.3, pp. 793–803.

Spitzweck, B., M. Brankatschk, and B. J. Dickson (2010). "Distinct protein domains and expression patterns confer divergent axon guidance functions for Drosophila Robo receptors." *Cell* 140.3, pp. 409–20.

Steinmetz, P. R. et al. (2007). "Polychaete trunk neuroectoderm converges and extends by mediolateral cell intercalation." *PNAS* 104.8, pp. 2727–32.

Steinmetz, P. R. et al. (2011). "The segmental pattern of otx, gbx, and Hox genes in the annelid Platynereis dumerilii." *Evolution Development* 13.1, pp. 72–9.

Strausfeld, N. and F. Hirth (2013). "Deep homology of arthropod central complex and vertebrate basal ganglia." *Science* 12.340, pp. 157–161.

Talpalar, A. E. et al. (2013). "Dual-mode operation of neuronal networks involved in left-right alternation." *Nature* 500.7460, pp. 85–8.

Tessmar-Raible, K. and D. Arendt (2003). "Emerging systems: between vertebrates and arthropods, the Lophotrochozoa." *Curr. Opin. Genet. Dev.* 13.4, pp. 331–40.

Thelie, A et al. (2015). "Prdm12 specifies V1 interneurons through cross-repressive interactions with Dbx1 and Nkx6 genes in Xenopus". *Development* 142.19, pp. 3416–3428.

Tomer, R. et al. (2010). "Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium." *Cell* 142.5, pp. 800–9.

Tosches, M. et al. (2014). “Melatonin Signaling Controls Circadian Swimming Behavior in Marine Zooplankton”. *Cell* 159.1.

Wanner, A. A. et al. (2016). “Dense EM-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb.” *Nat. Neurosci.* 19.6, pp. 816–25.

Wentworth, L. (1984a). “The development of the cervical spinal cord of the mouse embryo. I. A Golgi analysis of ventral root neuron differentiation.” *J. Comp. Neurol.* 222.1, pp. 81–95.

Wentworth, L. (1984b). “The development of the cervical spinal cord of the mouse embryo. II. A Golgi analysis of sensory, commissural, and association cell differentiation.” *J. Comp. Neurol.* 222.1, pp. 96–115.

Wilkinson, M. (2015). *Restless Creatures - The story of life in ten movements*. New York: Basic Books.

Williams, E., M. Conzelmann, and G. Jékely (2015). “Myoinhibitory peptide regulates feeding in the marine annelid Platynereis”. *Front Zool* 12.1, p. 1.

Wilson, S. et al. (2008). “A molecular program for contralateral trajectory: Rig-1 control by LIM homeodomain transcription factors.” *Neuron* 59.3, pp. 413–24.

Wyart, C. et al. (2009). “Optogenetic dissection of a behavioural module in the vertebrate spinal cord”. *Nature* 461, pp. 407–410.

Zhang, J. et al. (2014). “V1 and V2b Interneurons Secure the Alternating Flexor-Extensor Motor Activity Mice Require for Limbed Locomotion”. *Neuron* 82.1, p. 138150.