Testing the homology of metazoan larvae using single cell sequencing data

Laura Piovani

UCL

Submitted for the Degree of Doctor of Philosophy

2021

Declaration

I, Laura Piovani confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Members of at least 15 animal phyla have a bi-phasic lifecycle composed of a pelagic larval stage which metamorphoses into a benthic adult. This lifecycle enables larvae to disperse in the ocean to colonise new environments. Although bi-phasic life cycles are widespread across Metazoa, it is unclear if marine larvae were present in the common ancestor or have evolved multiple times as an adaptation to a common selective pressure. Answering this question is fundamental if we are to reconstruct the lifecycle of the last common ancestor of animals.

Many marine larvae are at least superficially similar, they are small, swim through beating of ciliated bands and sense the environment with an apical organ structure. Most animals with such larvae belong to the superclade of Lophotrochozoa and this thesis will concentrate on this group. To assess the homology of larvae, authors have compared the early development and expression profile of similar larval organs across phyla and shown that, generally, these are conserved. However, these studies only compared a handful of species and genes and a more thorough approach is needed. In this work I have used single cell sequencing to characterise the cells present in two lophotrochozoan marine larvae (mollusc's trochophore larva and polyclad flatworms' Mueller's larva). I have performed in situ hybridisation and hybridisation chain reaction to identify cells belonging to different structures (including the ciliary bands and apical organ). Subsequently, I have compared the expression of orthologous genes in cell types of the two larvae. I have found that ciliary cells, neurons, muscles and proliferative cells co-express a large number of orthologous genes (>600) of which many are transcription factors (>30).

These results hint at a likely homology of lophotrochozoan larvae and set the basis to expand comparative work to more larvae across Bilateria and possibly Metazoa.

Statement of Scientific Impact

Many marine animals go through a larval stage that facilitates dispersal and colonisation of new environments. These marine ciliated larvae are widespread across the tree of animals and often look strikingly similar even though they subsequently metamorphose into morphologically distinct adults (such as sea urchins, clams or worms). This striking similarity of larvae prompts the question of whether larvae are homologous - evolved once in a common ancestor - or whether they were convergently invented multiple times as similar solutions to a common set of problems. In the case of marine invertebrate larvae many factors such as conserved features of early development, similarity in cell fates and gene expression seem to hint to a plausible common origin. This is an exciting question per se but if we can answer it, it would help us to reconstruct the lifecycle of the last common ancestor of animals.

To tackle this century old conundrum, I have optimised and applied recently developed molecular techniques such as single cell sequencing and hybridisation chain reaction to marine invertebrate larvae that had never previously been used. The data I was able to gather allowed me to compare two marine ciliated larvae from different phyla at a single cell resolution, something which had never previously been attempted. Not only has this informed us on the genetic blueprint that underlies the superficial similarity of marine ciliated larvae but has also unravelled some of the cellular complexity of these understudied organisms.

Exploring the diversity of cell types across different animals and systematically comparing them is a rapidly evolving field which will ultimately help us understand an important aspect of the evolution of the astonishing diversity of life forms we see today.

Acknowledgements

I have truly enjoyed the crazy rollercoaster that this PhD was and it wouldn't have been such a positive, creative, and thoroughly fun experience without all of you.

Firstly, I would like to thank my supervisor Max Telford for his constant support, for always having my back and being super positive about my work and my capabilities, believing in me when I made mistakes and allowing me to try again, over, and over, again.

I would also like to thank my master's supervisor Michela Sugni, who spent countless evenings on the phone with me, making sure I wouldn't lose my drive nor my positive outlook on science.

A huge thank you to all the past and present members of the Telford, Oliveri and Yang lab for putting up with my million random questions and for each teaching me something new. To Irepan for always pushing me to do better. To Jeff and Ferdi who were always interested and helped me think things through. To Paschalis who was always happy to teach me stuff and took me on the crazy *Paratomella* hunting trip. To Anne, Helen and Johannes for teaching me half of the things I know and passing on their precious worm superpowers. To Paschalia who always gives the wisest advice and is the best scientist I have ever met. To Steven who made me feel at home at UCL from day one. To Natalie, who is my sweetest friend, I simply wouldn't have done this without you, nor I would have wanted to really. To Fraser who was always there to help, who kept the animals alive, cloned my genes and pipetted my cell dissociations when my hands were shaky. To my incredible student Leisl who helped through the hardest part of my PhD and who would always show up full of enthusiasm even when we were failing hard.

Thank you to all the members of the dysfunctional happy family that is the EVOCELL network. Thank you for teaching me all there is to know about single cell sequencing and for sharing the pain that single cell sequencing is. A huge thank you to Jake who was always willing to help me, teach me and inspire me. Thank you to Luca and Kevin who sat with me through the most stressful month of my life. A big thank you to my girls Aka, Derv, Cyri and Ines for letting me be part of their cool gang.

Grazie ai miei amici di sempre, i polpi, che mi prendono in giro la maggior parte del tempo ma ogni tanto sono rimasti seri quando mi serviva una mano. Grazie per tutte le volte che mi avete chiesto del mio dottorato, di come funzionasse, di cosa stessi facendo e per aver ascoltato le mie contorte spiegazioni. Un grazie speciale ai miei migliori amici Andrea e Marta che ci sono sempre stati per tutto.

A mi gordas, que aman el mar y mirar el maravilloso que es el mundo.

Thank you to the big NLMC crew, but especially to Peter, who would hold my rope and drive me places whenever I needed a break.

Grazie alla mia famiglia, che ha sempre sostenuto e assecondato le mie pazze idee e che sa volermi bene anche da lontano.

Thank you to my cat Pesto and my dog Dinka for being my number one fans, reminding me that there's much more to life than this.

Finally, thank you to John, who was always there at the end of the good days and the bad days, who believes and loves me so much it gave me strength to start over again each day.

A mamma,

che mi ha sempre portata in laboratorio,

fatto giocare col ghiaccio secco

e disegnare su tutte le lavagne dell'univeristà.

"Never be so kind you forget to be clever,

Never be so clever you forget to be kind."

Taylor Swift

Table of Contents

List of figures

List of tables

1 Introduction

Many aquatic animal phyla have members that develop via a feeding larva that is morphologically different from the adult and typically occupies a distinct ecological niche. Interestingly, marine larvae from animals that originated before the base of the Cambrian (~540 Million years ago) share a number of characteristics. The most notable are an apical organ to sense the environment and ciliated lobes for swimming. The most striking example of this conservation of form, perhaps, are the trochophore larvae of the lophotrochozoans; trochophore larvae are found in molluscs, annelids, entoprocts and, in a possibly derived form, in platyhelminths, brachiopods, nemerteans, ectoprocts and phoronids. Despite the morphological similarities between these larvae, which will be later discussed in depth, the homology of the ciliated larvae across phyla remains debated. The aim of the research described in this thesis is to assess if the relatively dissimilar ciliated larvae of a bivalve mollusc and a polyclad platyhelminth are indeed homologous. To address this question, I will use a relatively new source of comparative data – single cell sequencing - which will be used to define the larval cell types of each larval type *a priori*. The transcriptional profile of similar cells in different species will then be compared as a mean to assess their homology.

1.1 The origin of marine larvae

Reconstructing the early evolution animals can be informed to some extent by fossils, but some aspects of life are rarely if ever fossilised. In such cases our only resort is to reconstruct features of Cambrian or Precambrian ancestors by extrapolating backwards from their living descendants. One characteristic that is very unlikely to leave a trace in the fossil record is the tiny larval stage seen in many living marine invertebrates.

Larvae of marine invertebrates are found in at least 15 animal groups and are typically considered to be an adaptation to enable dispersal. The planktonic larval stage is followed by a radical metamorphosis into a very different adult form that is usually benthic. Larvae of annelids, molluscs, flatworms, brachiopods, nemerteans, entoprocts, ectoprocts, phoronids, echinoderms, hemichordates and (to a lesser extent) cnidarians present many similar characteristics. They are small, they locomote by the co-ordinated beating of bands of ciliated cells and have a simple nervous system with (sometimes) eye spots and a sensory apical organ. These morphological similarities make us wonder whether these characters were inherited from a common larval ancestor or whether they evolved convergently many times due to physical constraints (i.e. being a small organism that can move in the water column and sense the environment to feed).

We know that biphasic lifecycles evolved more than once – it is clear that tadpoles and maggots are not homologous characteristics of amphibians and insects – but with the larvae of many marine invertebrate species the homology of larval stages is much less clear. Strikingly, if we plot the phylogenetic distribution of marine larvae we find that many (molluscs, annelids flatworms, brachiopods, nemerteans, entoprocts, ectoprocts and phoronids) belong to the major animal clade of Lophotrochozoa (which will be the focus of this work) (see figure 1, star indicates larval stages are present). Lophotrochozoa, which make up one of the three main branches of Bilateria, together with Ecdysozoa and Deuterostomia, comprises more animal phyla than any other metazoan clade (around a dozen) (for reviews see Giribet 2008; Kocot 2016) (see fig. 1). More specifically all phyla that go through a larval stage (annelids, molluscs, entoprocts, ectoprocts, phoronids, nemerteans and platyhelminths) together with the direct developing gastrotrichs make up a

subclade, which in this work will be referred to as Spiralia. To further clarify, since there is still disagreement in the field, in this thesis I will refer to the larger group containing Gnathifera as Lophotrochozoa and to the sub-clade containing annelids, molluscs, entoprocts, ectoprocts, phoronids, nemerteans, platyhelminths and gastrotrichs as Spiralia (as clarified in figure 1).

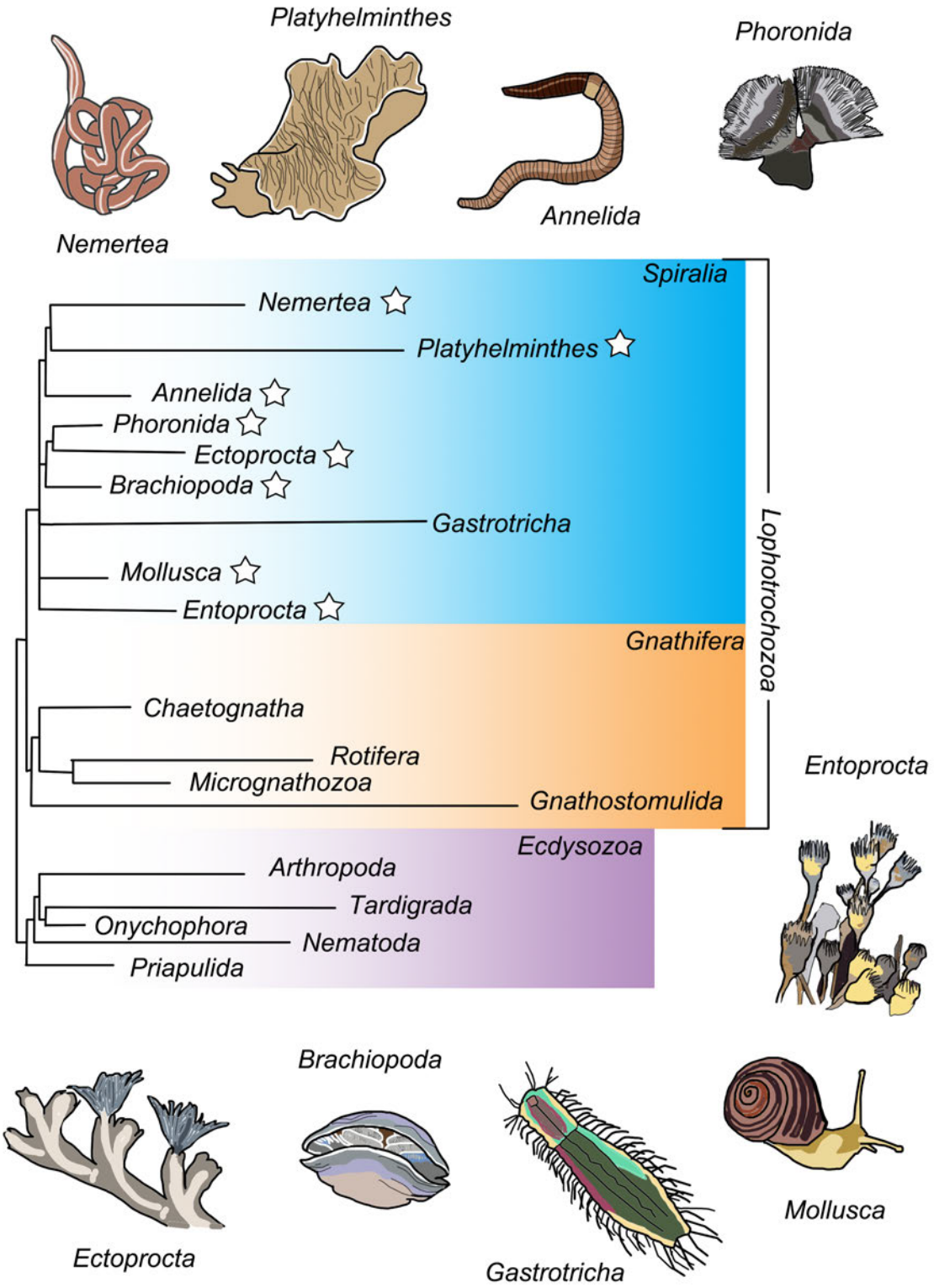


Figure 1. Our current understanding of the phylogeny of Lophotrochozoa shows a subclade called Spiralia where most phyla have larval stages (star).
Tree is simplified from Marlètaz et al, 2019. The phylogenetic tree shows a subgroup of Lophotrochozoa which I label Spiralia which includes Nemertea, Platyhelminthes, Annelida, Phoronida, Ectoprocta, Brachiopoda, Gastrotricha, Mollusca and Entoprocta. Most spiralian phyla go through spiral cleavage (except for Gastrotricha, Brachiopoda and Ectoprocta) and have ciliated larvae (indicated here with a star).

1.2 Past efforts in unravelling the origin of biphasic life cycle

So far, I have explained how marine larvae look superficially similar and how many are also closely related. So, if we were to think in parsimonious terms, it would be more likely to assume that these larvae only evolved once in the ancestor of Lophotrochozoa; however, some more distantly related larvae also share these characters (such as those of echinoderms and hemichordates). Is it hence possible that ciliated larvae are just a common solution to the similar physical problem of having to survive in the plankton?

Historically, two simplified hypotheses on the origin of larvae have been proposed: the larvae first hypothesis and the intercalation hypothesis (see figure 7). The larvae first hypothesis was initially proposed by Haeckel, and endorsed by Hatschek and Jagersten but it was Nielsen who further developed it under the name of "trochaea theory". He suggested that the last common ancestor of all metazoans was a small holoplanktonic larva with a single ciliary band that was used for swimming and to capture food particles.
This ancestral larva would have then differentiated into a trochophore-like larva in protostomes and a dipleurula-like larva in deuterostomes. Under this hypothesis the split between deuterostome and protostome likely occurred before the addition of an adult benthic stage (Haeckel, 1874; Hatschek, 1878; Jagersten, 1972; Rieger, 1994; Page, 2009). All modern-day larvae descended directly from this original larval ancestor are termed "primary larvae"; these include the protostome trochophore-like larvae dealt in this study, the dipleurula-like larvae of deuterostomes as well as the larvae of Porifera and Cnidaria but not the larvae of arthropods of chordates which evolved secondarily (Jagersten, 1972).

For a long time, this hypothesis was considered more probable as it was parsimonious to think that so many larval similarities had simply evolved once before the split of Bilateria (Rieger, 1994).

Moreover, in the 1990s many studies on sea urchin development brought mechanistic insight on how the addition of a benthic adult might have come to be. In fact, most larvae are made up of cells with a limited number of cell division that differentiate under the stimuli of individual cells. Under these restrictions only small organisms, like modern day larvae, could develop. However, in the sea urchin larva, the adult (or juvenile) develops from a number of set aside multipotent cells with a much greater mitotic potential. Moreover, during development of the juvenile/adult body, morphogenesis is driven by the patterning of broader special domains (and not individual cells) by transcription factors. Many authors suggested that the evolution of these features (i.e. the mitotic potential of set aside cells and broad special patterning) could have brought the innovation of a larger benthic adult body from the original larval ancestor (Davidson et al., 1995, 1998; Peterson et al., 1997; Cameron et al., 1998; Page, 2009).

The second scenario, depicts the first bilaterian as a holobenthic animal that secondarily intercalated a pelagic larva. This would mean that indirect development is a derived character and that larvae have potentially evolved multiple times (Jenner, 2000; Sly et al, 2003; Page, 2009). Under this scenario, which is becoming more widely accepted nowadays, larvae would have evolved secondarily from adults, meaning that most larval structures would have likely co-opted adult molecular pathways. If one considers that adult stages across Metazoa share many cell types, such as protonephridia and photoreceptors, and that all larvae face similar physical constraints it becomes less un-parsimonious to think that larval similarities could arise due to convergence.

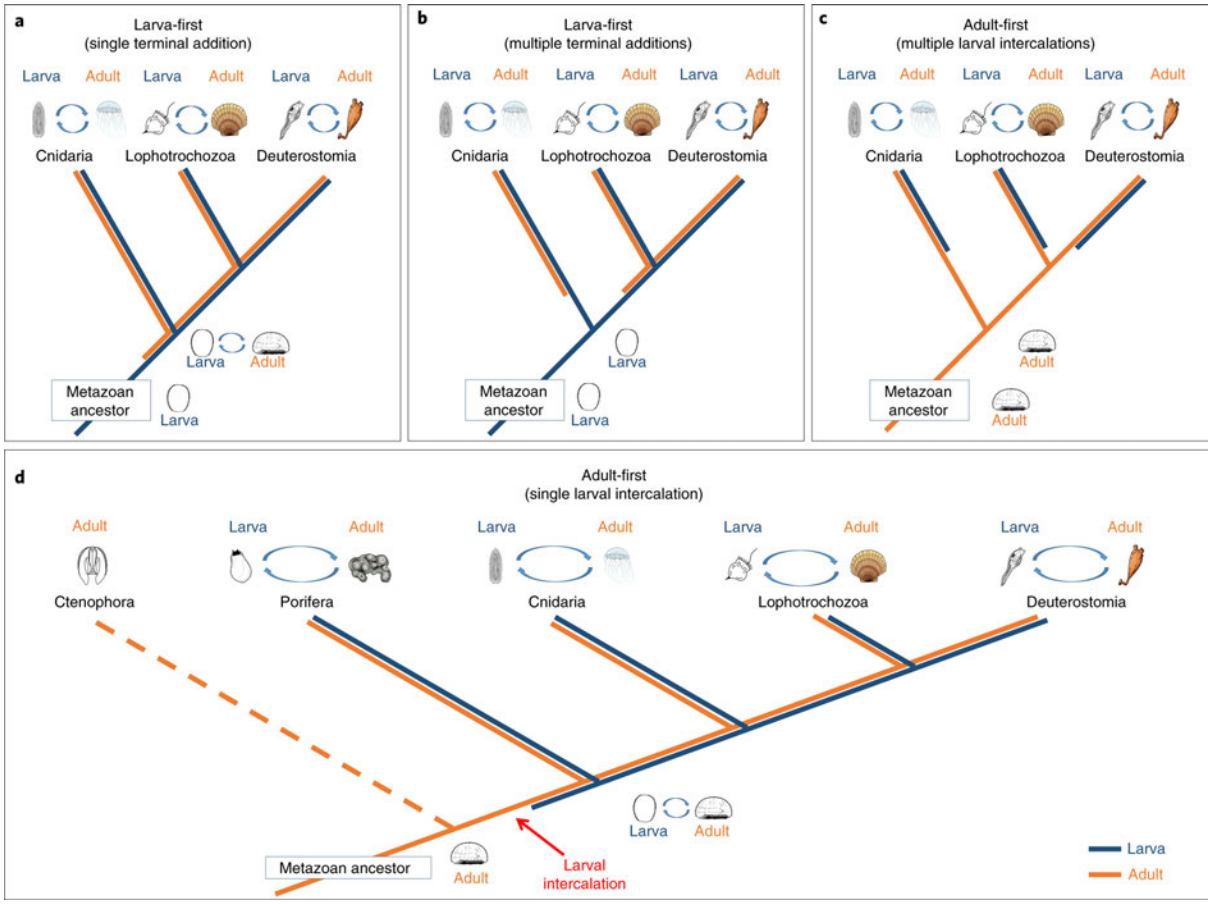


Figure 2. Different possible scenarios for the origin of larvae.

Taken from Wang et al 2020.

1.3 Phylogenetic support for larva first or adult first scenario

Early phylogenetic evidence seemed to support the adult first hypothesis as many direct developing phyla were wrongly placed at the base of the metazoan tree (Jenner, 2000; Peterson et al., 2005). In fact, for a long time acoels, which are simple direct developing worms with a reduced cluster of Hox genes, were placed as sister group to all other bilaterians. This was used as evidence that the ancestor of all bilaterians might have been an acoel-like worm, that gradually acquired a separate larval stage. A slow process of metamorphosis would likely make the animals more vulnerable and selection would favour the evolution of a more rapid and drastic metamorphosis (Page, 2009).

However, more recent phylogenies have shown that acoels are part of a phylum called Xenacoelomorpha and that they are, likely, the sister group of Ambulacraria (Philippe et al, 2011 and 2019). Moreover, many phyla with indirect development (namely annelid, molluscs, entoprocts, nemerteans, ectoprocts, phoronids and brachiopods) have been grouped together in the Lophotrochoza, suggesting that at least in this clade larval stages could be ancestral.

In the specific case of the larva-first, adult-first scenario the fossil record cannot help us untangle the order of events leading to a biphasic life stage as both larvae (which are very small) and small acoel-like worms would have not fossilized well. Moreover, there is a long-standing debate on whether animals evolved well before their fossil record as some molecular clock estimation seem to suggest. The idea that modern phyla evolved before the Cambrian without leaving any fossil trace has been used by different authors to support both the larva first and acoel-like worm first scenario, creating further confusion (Budd & Mann, 2020).

In the past few hundred years different scientists have tried to tackle the question of the origin of larvae by looking in detail at A) the developmental processes B) the morphological ultrastructure and (more recently) C) the molecular pathways that lead to the formation of these larvae.

In the next few paragraphs I will expand on these past studies to try and understand to what extent developmental, morphological and molecular similarities can help us unravel the possible homology of Lophotrochozoan larvae.

1.4 Comparison of early embryogenesis of Spiralia

Inside Lophotrochozoa, most spiralian phyla (except for brachiopods, ectoprocts and gastrotrichs) have members whose early embryos follow a stereotypical early cleavage pattern called spiral cleavage (see table 1) (which gives name to this clade). As blastomeres divide, they shift in an alternating clockwise then anticlockwise fashion creating a spiral like arrangement (Wilson, 1892). The first two divisions give rise to four cells called macromeres that divide the embryo in four quadrants (A, B, C and D). Across phyla, these macromeres usually give rise, respectively, to the lateral left (A), lateral right (C), anterior ventral (B) and the posterior dorsal (D) body tissues (Henry and Martindale, 1999). The third set of divisions, which brings the embryo from the four to the eight cells stage is characteristically unequal. This results in an embryo consisting of four larger vegetal macromeres (1Q = 1A, 1B, 1C, 1D) and four smaller animal micromeres (1q = 1a, 1b, 1c, 1d). During this unequal cleavage, the macromeres shift, giving the first sign of spiral cleavage. In some species the first shift is clockwise viewed from the animal pole, and these are named dexiotropic, in other the first shift is counter-clockwise, named laeotropic. Up until the 32-cell stage, the divisions of both micromeres and macromeres continue in alternating directions, then the cleavages become less synchronous.

This synchronicity of early divisions allows us to precisely match blastomeres across phyla and compare blastomere cell fates across spiral cleaving animals. These observations have shown, as one can see in figure 2, that the first micromere quartet (1a, 1b, 1c and 1d) gives rise exclusively to ectodermal structures (specifically head structures such as the apical organ, brain, photoreceptors and anterior ciliated prototroch). The second and third

quartets give rise to mostly ectodermal structures such as the mouth and the trunk (2nd quartet), the nervous system (2nd and 3rd quartet) and the ectodermal foregut (2nd and 3rd quartet) but also to some mesodermal structures such as muscles. Due to its origin, the latter is called ecto-mesoderm and interestingly the contribution of micromeres to the ecto-mesoderm is quite variable among species. In contrast, in the fourth micromere quartet the cell in the d quadrant (called 4d) consistently gives rise to all endo-mesoderm (Lyons & Henry, 2014). This micromere (4d) divides horizontally into two equal-sized blastomeres (4d1 and 4d2) also establishing the bilateral symmetry of the embryo (Hejnol, 2010). Finally, the endoderm usually originates from the vegetal macromeres.

Overall, the fact that most spiralians have very similar early cleavages and that the same blastomeres consistently give rise to similar structures in the larvae seems to be a strong indication that larval development is conserved. In fact, it is hard to explain why if larvae evolved independently multiple times they would develop convergently in the same way. However, precise cell tracing is not yet available for all larvae and some indeed show some plasticity that makes it harder to trace blastomeres. Moreover, when spiral cleavage is not present, as for brachiopods, ectoprocts and gastrotrichs and for radially cleaving echinoderms, cell trajectories cannot easily be compared.

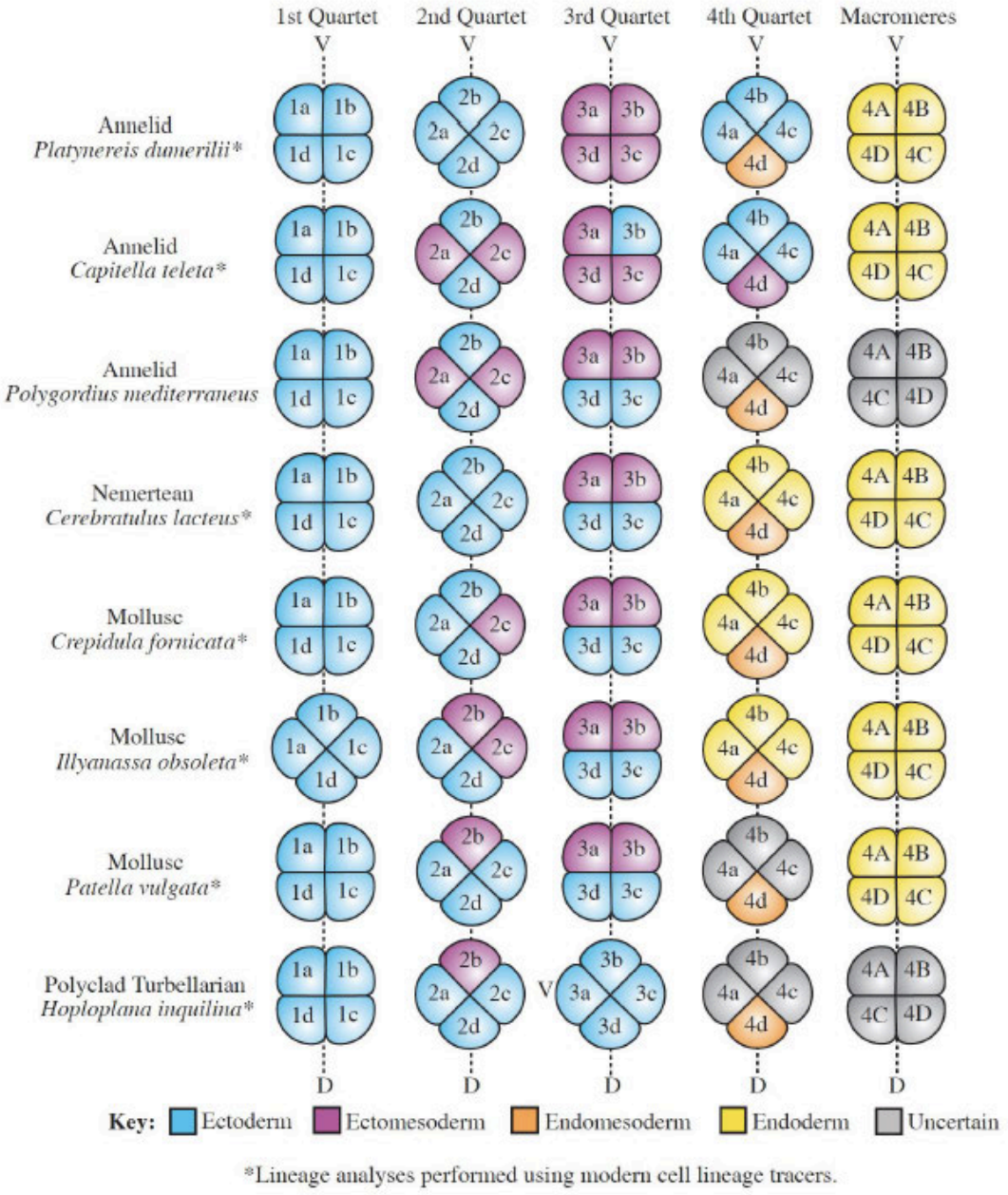


Figure 3. Fate map of early quartet cells across Spiralia show widespread conservation of cell fate, which has been interpreted in favour of larval homology.

From Lyons & Henry (2014)

1.5 Morphological comparison of lophotrochozoan larvae

Spiral cleavage and cell fate conservation are not the only characters shared by Spiralia. Lophotrochozoans in fact, owe part of their name to the characteristic larval type of annelids, molluscs and entoprocts, the “trochophore” (see fig. 3). Trochophore larvae

generally possess a preoral ciliary band (the prototroch) which approximately divides the larva in two hemispheres, the anterior region and the posterior region. In all trochophore larvae, as well as in some derived spiralian larvae which will later be discussed in details, the prototroch originates from a group of cells called the trochoblasts. The trochoblasts themselves derive from three different sets of ciliated cells: the primary, accessory and secondary trochoblasts. The primary trochoblasts, which supply most of the cells of the prototroch, derive from vegetal derivatives of the first quartet micromeres (1a2–1d2) (Henry et al, 2007). In most spiralians there are four primary trochoblasts and they only divide once or twice. As they cease to divide early on, the cells composing the prototroch are usually few (typically 20–40) and relatively large (von Dassow & Maslakova, 2017). The accessory trochoblasts derive from animal descendants of first quartet micromeres (1a1–1d1). Finally, the secondary trochoblasts originate from animal derivatives of the second quartet micromeres (2a1–2d1).

The region anterior to the prototroch, called the episphere, is typically equipped with sensory organs such as larval eyes and a long tuft of sensory cilia. The posterior part is used for locomotion and filter feeding and sometimes, in addition to the prototroch, accessory ciliary bands, - metatroch and telotroch - can be present, although their homology across phyla is not certain (Hejnol et al., 2007; Henry et al., 2007). Paired protonephridia are present in almost all trochophore larvae, and are used as excretory organs (see fig. 3).

An opposing-band feeding system, relying on the prototroch to generate a current for locomotion and feeding, and the metatroch beating in the opposite direction, was observed in several molluscs, annelids and entoprocts and prompted the idea that feeding

trochophore larvae could be ancestral to Spiralia (Strathmann, Jahn & Fonseca 1972; Nielsen, 1995). However, morphological cladistic analysis suggested that opposed-feeding, for complex as it might be, may have evolved multiple times (Rouse, 1999). These observations brought forward a more general definition of trochophore: a larva that has a prototroch (derived from trochoblasts), an apical plate and larval protonephridia (Rouse, 1999; Maslakova et al, 2004).

In the next few paragraphs I will discuss in details the different larval types of Spiralia. I will try to outline the similarities and differences that have been observed so far from a morphological point of view and highlight which larvae are currently considered homologous. I will often emphasize which features of each larva are lost during metamorphosis. These will be particularly informative for this work as larval specific features are less likely to have been co-opted independently from the adult.

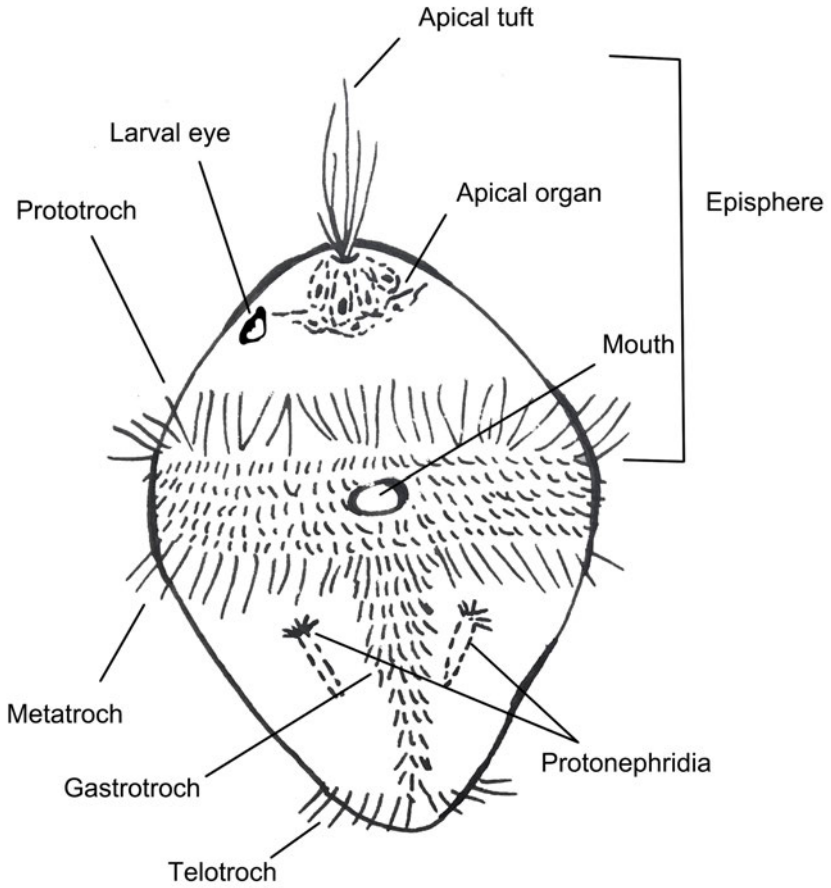


Figure 4. A typical trochophore larva features a ciliary band anterior to the mouth called the prototroch used for swimming; an apical organ with a sensory apical tuft and paired protonephridia.
Some trochophore present further ciliary bands: the metatroch, the gastrotroch and the telotroch.

1.4.1 Larvae of molluscs

Members of all mollusc classes except for cephalopods possess ciliated larvae. In caudofoveates, gastropods, polyplacophorans, scaphopods, and non-protobranch bivalves, a trochophore stage arises right after gastrulation. In some cases, the trochophore larva is free living, however, some freshwater snails and almost all terrestrial snails lay eggs in capsules. Inside the capsules the embryo goes through a ciliated phase, morphologically

similar to those of marine gastropods, which suggests that direct development is a derived character (Nielsen, 2004).

The trochophore larva of molluscs is characterised by a prototroch composed of 1 to 6 rows of pre-oral, multiciliated cells and an apical organ with a patch or tuft of cilia (Page, 2009; Nielsen, 2004). Vase-shaped cells appear to be consistently present in the apical organ (AO) of both bivalves and gastropods and are thought to have a sensory function. A single central vase shaped cell appears in the early trochophore stage and is soon followed by another two or more. Additional serotonin-like immunoreactive (-LIR) cells can be present in some molluscs as well as two round interneurons (Croll, 2009). The number of both vase-shaped cells and serotonin-LIR cells is, however, very variable between larvae (Croll & Dickinson, 2004).

Mollusc larvae often have pigmented larval eyes, which can have ciliary or rabdomeric photoreceptors and sometimes small lenses. The position and innervation of these larval eyes are, however, often different between species and hence their homology is debated (Nielsen, 2004).

A specialised dorsal epithelium makes up the mantle or shell field which is a striking character of the phylum. From the shell field the larval shell will start to form during the veliger stage, in conchiferan (shell bearing) molluscs, this region will generate at first one unpaired shell gland which, in bivalves, will divide later in two shells (Nielsen, 2004). Together with the shell, during the veliger stage the foot and one or two velar lobes, an expansion of the prototroch, start forming. During metamorphosis into the juvenile both the prototroch and the apical organ are lost. Protonephridia, which also disappear during metamorphosis, are present in larvae or embryos of solenogasters, polyplacophorans,

gastropods, bivalves and scaphopods. In gastropods transient larval features can sometimes include a larval heart and a group of shell-anchored muscles (Page, 2009).

In some molluscs, specifically in solenogastres and in protobranch bivalves, a second rarer type of larva is present and is called pericalymma. This is a trochophore-like larva with the episphere covered by a ciliated cap (called serosa); something similar is also present in some annelids. The developmental origin of the serosa is however different in different species so the homology of mollusc and annelid pericalymma larvae is unlikely (Nielsen, 2012).

To summarise, molluscs present both direct and indirect development, via either a pericalymma or, more commonly, a trochophore larva. Most authors agree that the latter is the primitive form and that trochophore larvae of molluscs are homologous to those of annelids (which will be discussed in the next section). Common larval specific characters of mollusc larvae include the prototroch, the apical organ and protonephridia and these will be investigated for their possible homology later in this work.

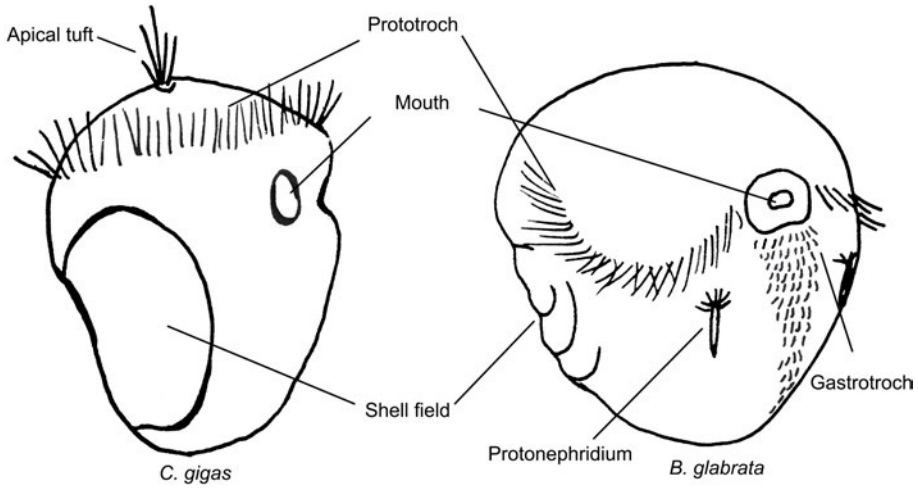


Figure 5. Trochophore larvae of the molluscs used in this study *C. gigas* and that of the gastropod *B. glabrata*.

1.4.2 Larvae of annelids

Annelids also have both direct and indirect developing species, however whenever larvae are present they usually represent a variation of a trochophore. Up to four different types of ciliary bands can be found in annelid trochophores and they are usually composed of multiciliated cells. Aside from the prototroch, considered homologous to that of other spiralian trochophores, larvae can feature a metatroch, a gastrotroch and a telotroch. Annelid trochophores present an apical organ with an apical tuft, an apical ganglion and sometimes 1 to many pigmented eyes. Eyes are made up of 1 to 2 rabdomeric sensory cells and 1 to 2 pigment cells each, and are often lost during metamorphosis. Paired protonephridia are found in most larvae. Paired protonephridia are usually present and consist of a terminal cell a duct cell and a pore cell.

Although usually annelids larvae are trochophores, a few species feature perycalymma larvae that are classified as type 2 or type 3 depending on the origin of their serosa. In type 2 perycalymma the serosa develops from a region immediately behind the metatroch whilst in type 3 perycalymma it derives from the posterior side of the first setiger. Genera that present a perycalymma larva are always within families that share classical trochophore larvae and for this reason perycalymma are considered derived.

Metamorphosis in annelids is usually very gradual and the only structures that tend to be lost are the ciliary bands and part of the apical organ (Nielsen, 2004).

Altogether annelids and molluscs trochophores share many similar structures that have long been considered homologous such as the prototroch, the apical organ and protonephridia.

However, most of these characters are also present in other spiralian larvae which are not referred to as trochophore and whose homology remains debated (Nielsen, 2004).

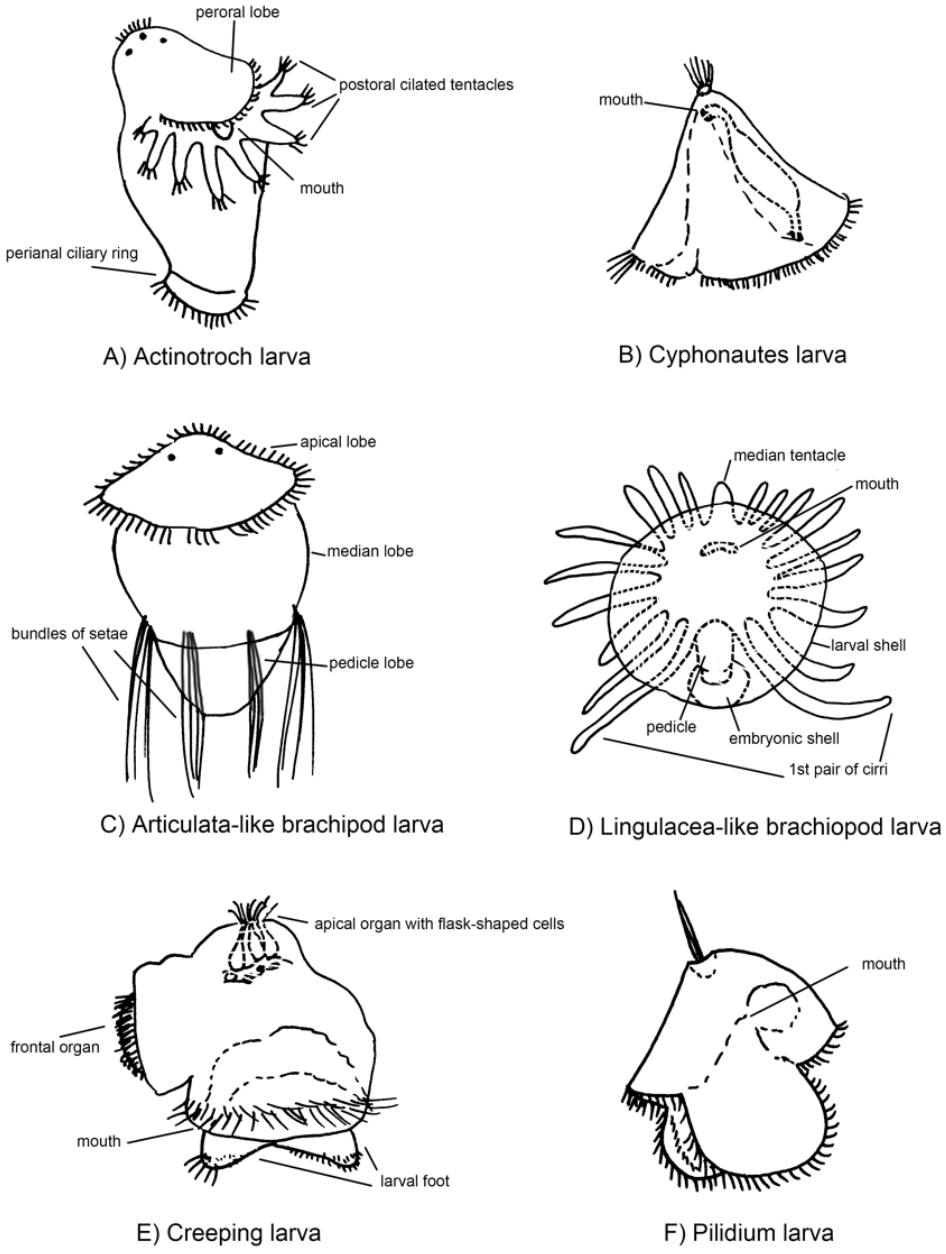


Figure 6. A representation of the diversity of larvae across Spiralia.

The pilidium larva of nemertean and the creeping larva of entoproct have been proposed as derived trochophores, the homology of the other larvae is debated (Peterson & Eernisse, 2001 and Nielsen, 2005). A) Actinotroch larva of phoronids, B) cyphonautes larva of ectoprocts, C) larva of Articulata and Craniacea brachiopods, D) Larva of Lingulacea and Discinacea brachiopods, E) creeping larva of entoprocts, F) pilidium larva of nemerteans.

1.4.3 Larvae of entoprocts

Entoprocta show two different types of larvae: a planktotrophic swimming larva and a creeping-type larva (fig. 5 E), which alternates between short periods of swimming and a creeping lifestyle. The former is more common; however, the latter is commonly considered ancestral. The swimming larvae, which can be planktonic or planktotrophic for several weeks, have a large episphere and the hyposphere is almost completely hidden by the compound cilia of the prototroch. The apical organ is comprised of an apical tuft and 3-4 serotonin positive flask cells, a paired nerve connects it to the ganglion of the frontal organ (an entoproct specific sensory organ) and then to the prototroch.

The creeping larvae possess a large larval foot, and a more complex apical organ than that of the swimming larva. The apical organ contains eight to ten serotonin-containing flask-shaped cells and surrounding peripheral non-flask cells.

Both swimming and creeping larvae possess paired protonephridia and larval eyes. The latter are located in the frontal organ and composed of a photoreceptor cell, a pigment cell, and a lens cell (Wanninger, 2015). The frontal organ as well as the apical organ and prototroch are lost during metamorphosis (Nielsen, 2002).

Altogether many features of both swimming and creeping entoproct larvae, such as the apical ganglion, protonephridia and the function and structure of ciliary bands suggests that these are derived trochophore larvae (Nielsen, 2005).

1.4.4 Larvae of nemerteans

Nemertean worms present two different larval morphologies, the so called planuliform larvae and the pilidium larvae (fig. 5 F). The pilidium larva is present only in the clade Pilidiophora and possesses a prototroch which partially originates from trochoblasts. However, differently to other Spiralia, the trochoblasts do not cease dividing early on and so the prototroch of pilidium larvae is composed of many small cells. For this reason, the homology of the pilidium ciliary band remains debated (Nielsen, 2005; Maslakova, 2010). The apical organ originates from the 1st micromere quartet, it lacks nerve cells and is lost at metamorphosis. One or two larval eyes can be present. Moreover, some authors have described a pair of ciliated ectodermal invagination in the pilidium hyposphere which have been interpreted as nephridia (Nielsen, 2005). Pilidium larvae show a very drastic metamorphosis, where the juvenile develops from a series of ectodermal invaginations (imaginal discs) and then hatches out of the larval body. A drastic metamorphosis destroys almost all larval ectodermal and mesodermal tissues (Maslakova et al, 2004).
Worms that present a planuliform larva show a considerably less dramatic metamorphosis and for this reason were termed “direct” developers. However, planuliform larvae still present transient characters, such as an apical organ, eyes, larval ectoderm and ciliary bristles or cirri that are lost in the juvenile. Studies on the planuliform larva of *Carinoma tremaphoros* have shown the presence of a “hidden” preoral belt of about 40 cells that have the same lineages as trochoblasts of other Spiralia (Maslakova et al, 2004).

Due to the phylogenetic position of clades showing planuliform and pilidium larvae, most authors consider the latter a derived form. Moreover, some authors have proposed that the nemertean larvae are derived trochophores (Peterson & Eernisse, 2001).

Altogether, larvae of annelids and molluscs are considered homologous and some authors agree that larvae of entoprocts and nemertean could also be derived trochophores (Nielsen, 2005). In these larvae in fact, the development of common larval structure such as the apical organ and the prototroch have been investigated in depth with cell tracing studies. Since these structures develop in a similar fashion and are structurally similar we are inclined to think that they were inherited from a common ancestor. However, smilar lines of evidence for the rest of spiralian larvae are lacking, although many characters appear superficially similar. To make matter worse we still haven't resolved the relationship of phyla within Lophotrochozoa and hence cannot extrapolate whether is more parsimonious to think that (at least) the trochophore-like larvae of mollusc, annelids, entoprocts and nemertean only evolved once.

In the next few paragraphs I will briefly outline what we know about the remaining larvae of spiralians and highlight studies that have approached the problem of the homology of structures such as the prototroch or the apical organ.

1.4.5 Larvae of phoronids

Both spiral, radial and bi-radial cleavage, as well as variation of these stereotypical early divisions have been observed in phoronids (Santagata, 2015a). For this reason, it is hard to identify the same blastomeres across species to infer homology of structures between the trochophore larva and the typical larva of phoronids (called actinotroch). The actinotroch

larva (fig. 5 A) possess a preoral hood or lobe over the mouth, and a post oral series of ciliated tentacles that are mostly used for feeding the “lophophore”. Moreover, the larva features a perianal ciliary ring composed of compound cilia on monociliated cells which are used for locomotion. A pair of nephridia, each with their own nephridiopores and clusters of protonephridia near the end of the nephridial ducts are present.

The apical organ of phoronid larvae has a U-shape and contains different neuronal cell types such as serotonin-like immunoreactive cells, including numerous flask-shaped cells, catecholamine-like immunoreactive cells, situated in the periphery, and a few FMRFamide-like immunoreactive cells. The latter lack cilia and do not appear to be sensory (Nielsen, 2005). When the larva is close to metamorphosis, a second sensory organ, called the frontal organ, usually develops in the hood from three median nerves emanating from the apical organ. The frontal organ usually contains a few bipolar serotonergic cells, and is thought to take part in larval settlement (Santagata, 2015a). During metamorphosis most of the larval tentacles, the whole larval nervous system and nephridia are lost (similarly to what happens in other spiralian larvae).

1.4.6 Larvae of ectoprocts

Ectoprocts present two different larval types: most species feature a lecithotrophic larva with a “corona” or ring of separate cilia, however a few species possess a planktotrophic larva called cyphonautes (fig. 5 B). Both larvae have a ring of multiciliated cells that interestingly seem to have similar cell lineage to that of prototroch cells even though ectoprocts do not undergo spiral cleavage. The apical organ consists of a core of neuronal cells surrounded in a concentric fashion by ciliated, myoepithelial and undifferentiated cells.

Eye spots are common in lecithotrophic larvae and are composed of a ciliated epithelial sensory cell surrounded by a pigment cell. A sort of modified protonephridium has only been observed in one cyphonautes larva. Almost all larval organs are lost during metamorphosis (Santagata, 2015b). Altogether larvae of ectoproct present, once again, characters that are similar to trochophore larvae (such as the apical organ and possibly the multi-ciliated corona) however their homology remains unclear.

1.4.7 Larvae of brachiopods

Brachipod larvae have no specific name, however two main morphology have been observed in different clades. Larvae of Lingulacea and Discinacea possess two shell, one semi-circular embryonic shell, which develops before the larva hatches and a larger circular larval shell (see fig. 5 D). They also present a lophophore made of pairs of cirri that bud off on either side of the median tentacle sequentially. The median tentacle itself is considered to be a sensory organ. Once the larva is ready to settle the pedicle begins to differentiate and eventually attach to the substrate to allow the larva to start burrowing.
The second type of larvae is typical of Articulata and Craniacea (see fig. 5 C). Full grown larvae present three lobes: an anterior rounded apical lobe, a middle mantle lobe and a posterior pedicle lobe. The apical lobe is overall ciliated and, in some species, it develops a prominent band of locomotory cilia. The apical lobe also features an apical tuft of immotile cilia and, sometimes, eye spots and vesicular bodies can be present (the function of the latter remains unknown). The mantle lobe, which is unciliated in mature larvae, bears four bundles of setae. The pedicle lobe, also unciliated, eventually develops an “attachment

disk" used for settling. All larvae lack a functional gut and hence do not feed (Young et al, 2002).

In general, ciliary bands of brachiopods larvae, when present, show little similarities with that of trochophore both structure, function and in cell lineage. However, brachiopods do not undergo spiral cleavage so clearly the overall development of larvae here is not conserved. Brachiopods larvae do however present an apical ganglion with cilia similar to all other invertebrate larvae, including that of echinoderms (Nielsen, 2005).

1.4.8 Larvae of platyhelminthes

In contrast to many other lophotrochozoans, most platyhelminths (or members of phylum Platyhelminthes) are direct developers and most species do not undergo spiral cleavage. However, members of the Polycladida has many members that present both spiral cleavage and indirect development via a larva.

Polyclad flatworms present three different larval morphologies: the Mueller's larvae, the Goette's larvae and the Kato's larvae (Teshirogi et al., 1981; Wang & Yu, 2008) (see fig. 6). The Mueller's larva is the most common and is present in both suborders of polyclads (Cotylea and Acotylea); it is usually characterised by eight multiciliated lobes and three simple eyes (two cerebral eyes and one epithelial eye) (Martín-Durán & Egger, 2012). The lobes have a varying number of rows of ciliary bands that are used for both filter feeding and for locomotion (Ruppert, 1978).

All larvae of polyclad flatworms are equipped with an apical organ with usually two serotonergic cells. The apical organ also contains either mono or multiciliate cells with long cilia surrounded by a ring of flask-shaped gland cells (Nielsen, 2005). Paired protonephridia have been found in both the Muller's and Goette's larva as well as in the embryos of direct developers such as the macrostomid Macrostomum (Nielsen, 2005; Rawlinson, 2010). Larval-specific characters of polyclad flatworms have been understudied due to the difficulties of raising larvae through metamorphosis, but the lobes, ciliary bands and the apical organ appear to be usually lost (similarly to other spiralian larvae).

Altogether, the apical organ and ciliated lobes of the polyclad flatworm larvae are usually not considered to be homologous to those of other spiralian larvae and it has been proposed that indirect development in flatworms has evolved secondarily (Martín-Durán & Egger 2012). Once again, the phylogenetic position of polyclads is debated and it does not necessarily explain whether spiral cleavage and bi-phasic life could be ancestral to this clade. Since this is one of the least studied larvae and it has often been considered fairly derived, one of the aims of this work will be to further characterise it in the hope of unravelling more larval similarities (or differences) and shed light on the origin of indirect development in flatworms and in Lophotrochozoa.

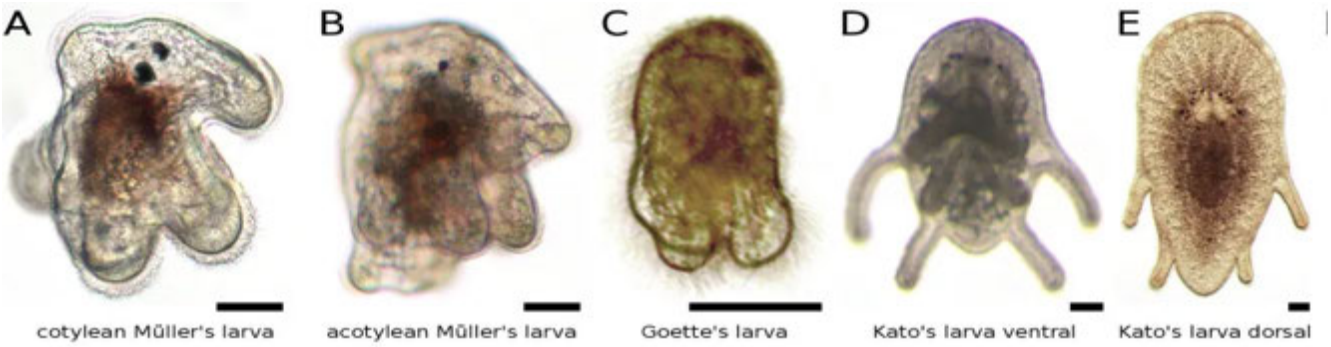


Figure 7. Polyclad flatworms are the only member of Platyhelminthes to present both spiral cleavage and a larval stage.

Four main larval morphology can be found in polyclad flatworms: A) Müller's larvae of a cotylean (*Prosthiostomum siphunculus*); B) Muller's larvae of an acotylean species (*Planocera multitentaculata*); C) Goette's larva of the acotylean *Imogine mediterranea* (four lobes and one cerebral and one epidermal eye); D-E) Kato's larva of the acotylean *Planocera reticulata* (eight lobes and 12 eyes) (from Martin-Duran & Egger, 2012)

Table 1. Shared characters across Spiralia, adapted from Marletaz et al (2019) and Nielsen (2005)

phylum	larval type	protonephridia	prototroch	apical serotonergic cells
Annelida	trochophore, perycalimma	Y	Y	Y
Brachiopoda	brachipods larvae	N	N	Y
Ectoprocta	chyphonautes, coronata	?	?	Y
Entoprocta	swimming, creeping	Y	Y	Y
Gastrotricha	N	Y	N	N
Mollusca	trochophore, perycalimma	Y	Y	Y
Nemertea	planuliform, pilidium	Y	Y	N
Phoronida	actinotroch	Y	N	Y
Plathyleminthes	Muller's, Goette's, Kato's	Y	Y	Y

1.5 Moving beyond ontogeny and morphological comparison of larval structures

In the previous sections I have described in details the incredible variety of larval forms present across Spiralia highlighting their similar characters. In summary, most of these larvae possess some form of ciliary bands (with often shared cell lineages across phyla), an apical organ (with vase or flask shaped cells and serotonergic cells) and paired protonephridia (for a list of shared features see table 1). However, some larvae such as the trochophore of molluscs and annelids and the larvae of entoprocts and nemertean are widely considered homologous (Peterson & Eernisse, 2001 and Nielsen, 2005) while others, such as the Mueller's larvae of flatworms, the actinotroch of phoronids and the cyphonautes of ectoprocts remain debated. Ultimately in all cases where we see A) plasticity in development (i.e. spiral cleavage is not the norm) or B) variance in developmental modes (indirect development is rare or many larval types are present) it becomes harder to clarify what characters are ancestral and which are derived and to homologise different structures in the larvae. Moreover, we often cannot rely on phylogenies to assess what characters are more likely to be synapomorphic since relationship within and across phyla of Lophotrochozoa remain largely unresolved. A striking example of this is, for instance, the lack of consensus on the relationship between the trochophore bearing molluscs and annelids.

1.5.1 Shared molecular blueprint of larval organs

As mentioned in detail in previous sections, morphological observation of marine larvae highlighted many shared structures including transient larval features such as ciliated lobes, apical organs and protonephridia. However, a part from a few exceptions (such as the trochophore larva of mollusc and annelids) homology of these superficially similar structure remains unclear. Moreover, cell tracing studies have shown that across spiralian larvae at least, cell fate is usually approximately conserved (Nielsen, 2004; Nielsen, 2005; Henry et al, 2007). But when spiral cleavage is not present, as for some spiralian larvae as well as radially cleaving echinoderms, it becomes hard to match blastomeres.

Altogether, it appears impossible to establish with certainty the homology of larval structures on morphological or developmental grounds alone. And for this reason, in the last few decades authors have tried to approach this problem by looking at the expression of developmental regulatory genes to see whether there is a conserved molecular blueprint in structures that are thought to be homologous (Arendt et al, 2001; Arenas-Mena et al, 2007; Marlow et al, 2014). The idea behind these studies is that if two similar structures arose via convergent evolution they would likely deploy different molecular pathways. These studies have hinted that many genes could have conserved roles in larval patterning across larvae of protostomes and deuterostomes. To give a few examples, *brachyury*, *goosecoid*, *nk2.1* and *foxA* are all expressed in the foregut of larvae, *otx* is expressed in the oral ciliary band, and *nodal* is used to establish left and right symmetry (Arendt et al, 2001; Dunn et. al 2007; Grande & Patel 2008; Raff, 2008). Although very promising, these studies present a few weaknesses. Firstly, sampling is usually low, with a few genes compared between just two species (sometimes only one protostome and one deuterostome).

Secondly, if we imagine that larvae secondarily evolved from a benthic ancestor then many genes would have likely been co-opted from the adult (Raff, 2008). Their expression would then be similar because it was co-opted separately by adults that inherited it from their common ancestor. To give an example, early on after their appearance larvae would have need to evolve a gut for feeding. It is more parsimonious to think they would co-opt genes involved in gut formation in the adult rather than re-invent gut formation from scratch. In the sea urchin at least, there is evidence that genes expressed in the larval gut are also expressed in the adult gut, which could mean that they were co-opted from the adult (Love et al, 2008).

1.5.2 Bulk RNA-seq comparison of larval stages

More recently, three studies have tried to compare full transcriptomic dataset from different ontogenetic stages across many species, looking both at similarities in gene expression and at the age of genes expressed in each developmental stage (Xu et al, 2016; Wu et al, 2019; Wang et al, 2020). This new approach allowed the authors to look at many more genes and species than previous studies presented. However, these had the main pitfall of losing the information of where each gene is expressed (so for instance one could not compare the specific expression of structures that are thought to be homologous).

Both studies seem to agree on two points. Firstly, they found the highest level of novel genes expression at the trochophore stage, meaning that likely trochophore larvae are "younger" than other developmental stages and hence (at least) secondarily evolved from adults. Secondly, they found that for most of the genes that were highly expressed at the

trochophore stage the deepest clade that contained a homolog was the Metazoa. Wang and colleagues interpreted these findings by hypothesizing that larvae evolved only once before the split of metazoan clades. Wu and co-authors argued that the ecology of the larvae may play a role in this result. Finally, Xu and colleagued hinted that the evolutionary history of larvae might be more complex than we think.

1.5.4 Limitation of molecular comparisons of larvae

Ultimately it seems that by looking at different sets of genes in different animals, authors can sometimes overestimate similarities. At the same time, bulk RNA-seq (which allows us to look at all genes in one animals) doesn't have the resolution necessary to compare similar structures across species. Furthermore, species trees continue to change and some relationship (such as those within Lophotrochozoa) are so uncertain that we cannot use them to establish what is more likely to be ancestral in a parsimonious scenario.

However, whether it happened once or many times, the shift from direct to indirect development must have prompted many adaptive challenges in that ancestral animal. And if any traces of this shift are left after how many million years of evolution we would expect to find them in the molecular underpinnings of larval and adult organs.

In this study we have used a more precise way of comparing those molecular blueprints using the full transcriptional profiles of single cells to compare larvae across phyla. This allowed us to look at all genes expressed without losing the resolution of the structures we are more interested in (such as ciliary bands, apical organs and protonephridia). More

specifically, we have used single cell sequencing to assign cell types *a priori* and assess what cell types make up the "classical" trochophore larva of mollusc and the "derived" Muller's larva of polyclad flatworms. We were interested in knowing if organs of the larvae which have been hypothesized to be homologous, such as the ciliary lobes or the apical organ, are actually made up of cells with shared transcriptional profiles. This will allow us to predict if transient larval structures are indeed homologous and were present in the ancestor of Lophotrochozoa. Ultimately, by comparing our results with those of others we hope to shed light on the origin of larvae and cell types in general.

1.6 Studying cell type evolution using single cell sequencing

Similar to what happened in species phylogenies, cell types have been historically compared based on morphological characters. Cell ultrastructure was described using light and electron microscopy and would sometimes allow the identification of homologous cell types within closely related species but, as for species phylogenies, it was harder to span across longer evolutionary distances (Salvini-Plawen & Mayr, 1977). The first step towards a more complete comparison of cell types across phyla came from immunohistochemistry and *in situ* hybridisation techniques. This helped to start to characterise cells from a molecular point of view but was only feasible for looking at a few genes at the time.

A further, more substantial, advance was brought on by single cell sequencing, a technique that allows us to look at many genes expression profile across multiple cells. This technique finally enables us to assign cell types based on their complete molecular fingerprint rather than their structure or a few genes only (Arendt, 2008).

Using the complete transcriptional profile of cells, we can first identify larval specific cell types, such as cells of the apical organ, protonephridia, ciliary bands cells and larval eyes and then compare them across species. This will inform us as to whether some of these larval cell types share regulatory elements that were inherited by their common last ancestor or if indeed they evolved several times.

1.6.1 Droplet based single cell sequencing

Since the first single cell sequencing paper was published many new techniques have been developed to isolate and label single cells. In general, single cell sequencing requires first to dissociate tissues or whole organisms into single cells. Single cells need to then be placed into separated reaction chambers where cells are lysed and the RNA is labelled. The main difference between different methods is how cells are isolated: this can happen by using micromanipulation, either by hand or using serial dilution, with the use of automated flow cytometry sorting (FACs) or with microfluidics (for review see Shapiro et al, 2013; Gawad et al, 2016).

For this project we will use a droplet-based method which isolates cells with the use of a microfluidic chip where cells are separated into single aqueous droplets inside an oil emulsion. Each drop hosts a hydrogel carrying combinatorially barcoded primers, and this allows the mRNA of different cells to be labelled inside each droplet. Then the emulsion is broken and the mRNA from different cells can be processed together (see figure 8). Droplet based methods have the advantage of rapidly capturing a high number of cells, since cells are not sorted into physical wells and so the number of reaction “chambers” is not limited (Klein et al, 2015).

Using this technique, I'll be able to separately capture cells from the trochophore stage of a molluscs (the pacific oyster *Crassostrea gigas*) and the Müller's larva of the polyclad flatworm *Prosthecereaus crozieri* (previously *Maritigrella crozieri*). For each species I will then cluster together cells with similar transcriptional profiles, try to assign cell type identity and localize them in the larvae using *in situ* hybridization and immunohistochemistry. Finally, I will look for orthologous genes across the two species and use this to detect similar cell types in the two larvae. I would be particularly interested to see whether A) structures thought to be homologous in the two larvae (ciliary bands, apical organ cells) share many orthologous genes and B) whether I can detect any larval specific cells and whether these are shared across the two animals (such as a specific neuronal subset, or protonephridia). Should transcriptional profiles appear too divergent we could either conclude that these larvae are not related or that they have had too much time to diverge and traces of this distant relative are lost.

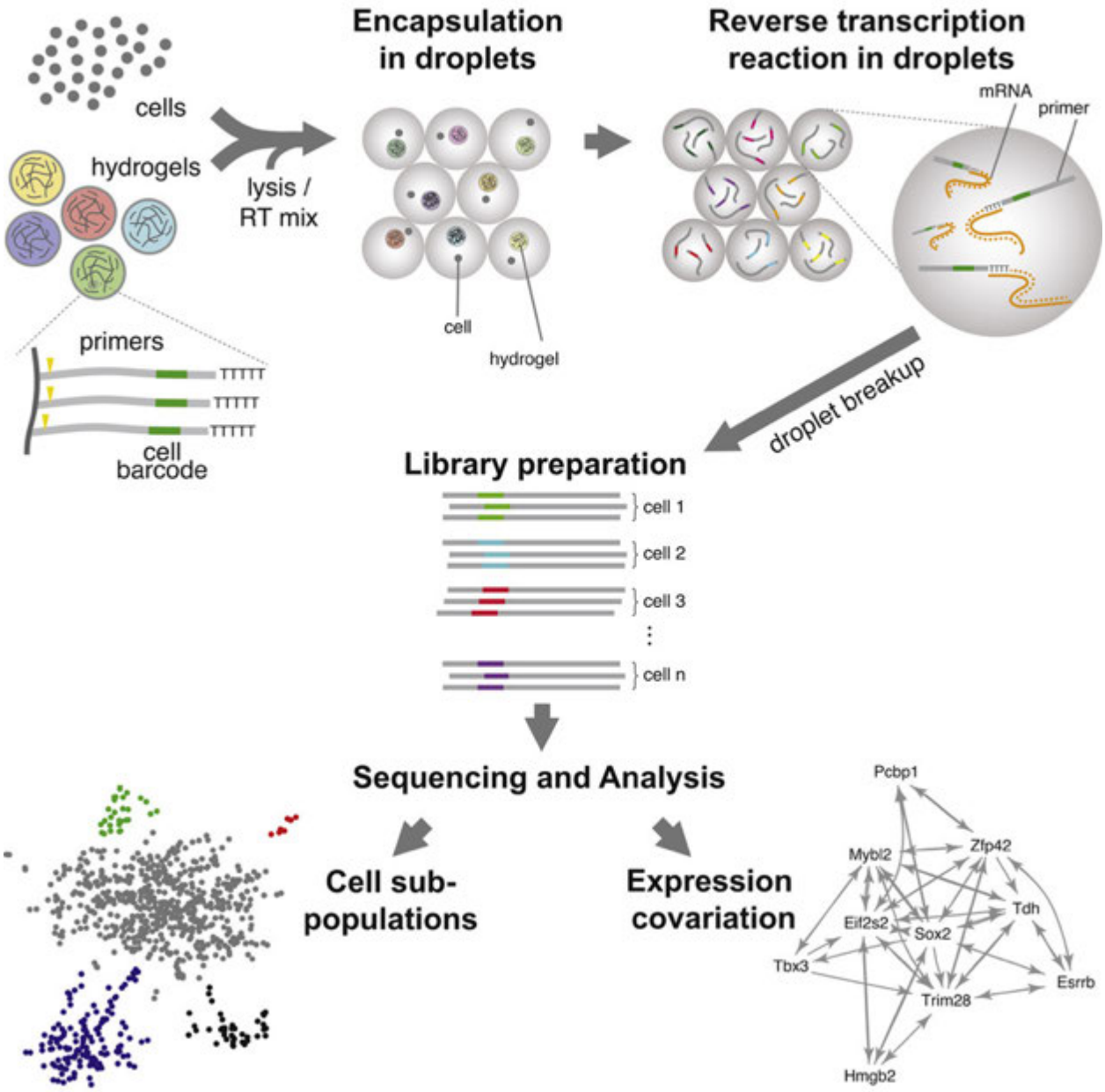


Figure 8. Droplet-based method for single cell sequencing allows sequencing of thousands of single cells in one experiment (from Klein et al., 2015).

In droplet-based methods cells are isolated into droplets contatining lysis buffer, reverse-transcription mix, and one hydrogel microsphere carrying barcoded primers each. Primers are released after encapsulation, then, during reverse transcription the cDNA inside each droplet is tagged with a barcode. Droplets lysate and barcoded cDNA from all cells is linearly amplified before sequencing.

2 Materials and methods

2.1 Animal husbandry

Crassostrea gigas individuals were bought during spawning season (May to August 2018 and 2019) from Richard Haywards Oysters at Borough Market in London, UK. The animals sold there are raised in farms in Salcott Creek Essex, UK. In the lab, oysters were kept at 16 C in running artificial sea water and fed three times a week with Spirulina powder and invertebrate food supplement.

Prosthecereaus crozieri adult specimens were collected in coastal mangrove areas in the Lower Florida Keys, USA in October 2019. Animals were found either on the ascidian *Ecteinascidia turbinate* or on the seafloor beneath them as previously described (Lapraz et al., 2013). Once the animals were taken back to London they were kept at room temperature (~21 C) in plastic boxes filled with artificial sea water. The water was changed daily for the first two weeks and then once every 2-3 days. The animals cannot be fed in the lab as they only eat ascidians, and so they were kept starved. Whenever eggs were found they were placed in separate containers and daily checked for hatching larvae.

2.2 Embryo and Larvae culture and fixation

Male and female *Crassostrea gigas* embryo were shucked, gametes were stripped and put in glass beakers containing filtered sea water (FSW). Eggs were left in ASW for about 1 hour to improve synchronicity then a dilution of sperm was added. After 5 minutes the water was tipped onto a 20µm filter mesh and washed several times to ensure fertilization and avoid

polyspermy. Fertilised eggs were then collected from the mesh and placed in beakers of FSW at either 20^0C or 25^0C in an incubator. Trochophore larvae were collected on a 20µm mesh after 16h/20h and either fixed for 1h or 20' (for HCR only) in 4% formaldehyde (diluted from 16 % paraformaldehyde: 43368 EM Grade, AlfaAesar) in 0.1 M phosphate buffered saline (PBS) or put in 1ml of trizol awaiting RNA extraction. Fixed embryos were dehydrated in a scale of MeOH-PBS (1:4,1:1,4:1) and then placed in 100% MeOH at -20 for storage. For phalloidin staining only larvae were not dehydrated and put in PBS with Sodium Azide.

Larvae of *P. crozieri* were collected daily once they started hatching. They were transferred into a filter and washed several times in filtered sea water. The larvae were then relaxed in 7.14% $MgCl_2$ * 6H2O in a glass vial and once they had settled the solution was substituted with 4% formaldehyde in 0.1M PBS for 30'. The larvae for WMISH were then washed 4 times in PBS and were dehydrated in a scale of MeOH-PBS (1:4,1:1,4:1) and then placed in 100% MeOH at -20 for storage. Larvae to be used for immunohistochemistry were washed 4 times in PBS and then transferred into PBS with Sodium Azide. Some larvae were placed in trizol awaiting RNA extraction.

2.3 Cloning and antisense probe synthesis

2.3.1 RNA extraction *Crassostrea gigas*

For RNA extraction 200µL of chloroform were added to all the samples in TRIZOL. Tubes were vortexed for one minute and then centrifuged (10 minutes, 4^0 C, 13000 RPM). The

upper aqueous layer was transferred to a new tube and an equal volume of of Isopropanol was added. Samples were centrifuged for 30 minutes at 4^0 C at 13000 RPM. The supernatant was then discarded and 1mL of freshly prepared 70% ethanol was added to wash the pellet. Samples were centrifuged for 15 minutes at 4^0 C at 13000 RPM. Ethanol was removed, and the pellet was air dried then the pellet was resuspended in 30µL of (NF- H_2O).
RNA was extracted using the RNeasy®mini kit (50) according to the manufacturer's instructions. The mRNA concentrations were measured using a NanoDrop™ and 1µL of 100ng/µL of each RNA sample was run on an electrophoresis gel to check for RNA degradation. To generate cDNA the Invitrogen SuperScript®III First-Strand Synthesis System for RT-PCR kit was used to transcribe the extracted mRNA into DNA. 0,5 to 1 ug of RNA were used and manufacturer's instructions were followed.

2.3.2 RNA extraction *Prosthecereaus crozieri*

Mixed larvae were placed in 1mL of TRIZOL, they were then either left at -80^0C awaiting extraction or directly vortexed until dissolved. 0.2mL of chloroform were then added and the tube was shaken for15 seconds. Samples were then transferred into a Phase Lock tube and manual instruction were followed. The aqueous phase was then pipetted and placed into a new tube. To the tube 0.25mL of RNAse-free isopropanol and 0.25 mL of high salt precipitation solution (0.8 M sodium citrate and 1.2 M NaCl) were added. The high salt solution is useful to remove proteoglycan and polysaccharide as it keeps them soluble whilst the RNA precipitates. The tube was then mixed by inverting it 10/20 times and left to sit at RT for 10 minutes. To precipitate the RNA the samples were then centrifuged at 4^0C for 15 minutes at 12000G. The supernatant was then removed and 1mL of ice cold 75% EtOH in

nuclease free water was added to the pellet. The tube was again centrifuged at 4^0C for 5 minutes at 7500G, the supernatant was removed and then the EtOH wash was repeated 3 more times. After the final wash the pellet was left to airdry at RT for 5 minutes and then 20µL of nuclease-free water were added to the RNA pellet. The tubes were then vortexed, spun down and placed on ice. RNA concentration and purity were assessed using a nanodrop and then the RNA was stored at -80^0C.

2.3.3 Primer Design and Cloning

Primer3 (http://primer3.ut.ee/) was used to design suitable primers for the ORFs of the genes selected. Parameters were changed as follows: Max 3' stability was changed to 8.0and max poly-x was changed to 3.0. Product range was adjusted to the size of the ORF. Primers were ordered from MWG Eurofins and stock solutions were prepared according to the manufacturer's instructions. The stock solution was then diluted to a 10pmol/µL concentration for PCR. In case a sequence could not be amplified with the first set of primers nested primers were ordered. For primer list see table S1.

2.3.4 Cloning Polymerase Chain Reaction (PCR)

Genes were cloned using either Red Taq polymerase or Q5 high fidelity polymerase PCR following manuals instructions. Pcr products were then run on a 1% agar gel in 1% TAE and fragment of the right size were cleaned using NucleoSpin Gel 8 PCR -Clean up kit following

the kit's instructions. Samples were eluted in 20µL of 70°C TE (pH 8) elution buffer to increase yield. The concentration of each sample was measured using the Qubit® Fluorometer dsDNA assay using manufacturer's instructions. Sequences amplified using Q5 polymerase were A tailed as Q5 polymerase produces PCR product with blunt ends that cannot be ligated efficentely. The A-tailed PCR product was ligated into a pGEM-T Easy vector (Promega) according to manufacturer's instructions. The ligation was incubated at 37°C for 2 hours or at 14°C overnight (or over weekend) before it was transformed into IBA Stargate Top 10 competent cells following manufacturer's instructions. Cells were then plated onto pre–warmed LB agar plates (with Ampicillin+X-Gal +IPTG). Plates were left to incubate overnight at 37°C. Ten white colonies per plate were selected and picked with a sterile 10µL pipette tip. The tips were briefly dipped in a PCR tube containing 25µL of RedTaq PCR master mix (prepared as for kit's instructions) and then placed into culture tubes containing 25ml of LB broth (+Amp) for later use.

A colony PCR program was run and products were run on a gel. For the reactions that contained an insert of the correct size, the cells placed in LB broth previously were incubated overnight in a shaker at 225 rpm at 37 °C. Plasmids were purified from cells using the Qiagen Plasmid midi kit according to the manufacturer's instructions and 1µg of the purified plasmid was sent for sequencing at Source Bioscience to confirm the insertion of the correct fragments and to check the orientation of the insertion in the vector. To amplify the inserted fragment, a template PCR was then performed on the purified plasmid using M13F and M13R primers and the Red Taq polymerase kit. The template DNA was then purified, concentration tested, ran on a gel electrophoresis and stored at -20° C. The anti-sense transcript was transcribed with either T7 or Sp6 enzyme (20U/µ l; Roche) and incubated for 3 (T7) to 6 (Sp6) hours. After incubation, 1µ l of DNase RNase-free (10U/µ l;

Roche) were added and incubated for 15 minutes at 37° C to remove the DNA template. Following this process, 30µ l DEPC-treated water and 25µ l 7.5M LiCl were added to precipitate the transcript overnight at -20° C. The sample was then centrifuged for 10 minutes at maximum speed, supernatant was removed and the pellet was washed in 200µL of 80% ethanol. The supernatant was removed, the pellet air dried for a maximum of 10 minutes and then re-suspended in 50µ l of DEPC-treated water. Aliquots of a working solution of 50ng/µL were prepared and stored at -80° C.

2.4 *In situ* Hybridisation

Whole embryo chromogenic *in situ* hybridisations for *C. gigas* were carried out following different protocols for optimisation purposes. The protocols trialled for chromogenic *in situ* followed these three papers: Grande and Patel 2008, Wood et. al 2018 and Osborne et. al 2018. Since the protocol from Osborne et al yielded the best results this was used for single fluorescent *in situ* up until the antibody blocking step. After that the protocol from Wood et al. 2018 for fluorescent *in situ* was followed with the only modification of using PTw (1x PBS with 0.5% Tween-20) as a buffer instead of TBST (0.2M Tris pH 7.5, 0.15M NaCl, 0.1% Tween-20). All ISH were performed in 96- well "U" bottom plate placed in a petri dish.

Many different whole-embryo chromogenic *in situ* hybridisations protocols were tried for *P. crozieri*: those from Wood et. al (2018), the UREA-based protocol from Sinigaglia (2018s), the modified protocol from Osborne and colleagues used for the oyster, and a modified protocol for Capitella provided by Kate Rawlinson. The latter is the protocol used for the two ISH shown in chapter 5 (r-opsin and troponin-T). Briefly, samples were rehydrated in

Ptw (0.1% Tween-20 in 1xDEPC-PBS) then they were put in proteinase-K (0.01mg/ml in Ptw) for 3 minutes. The digestion was stopped with two washes of Ptw with 2mg/mL glycine. Samples were incubated for 5' in 1% triethanolamine in Ptw with 1.5 µL acetic anhydride per 500µL then another 500µL of 1% triethanolamine in Ptw with 3 µL acetic anhydride were added (another 5'). Samples were washes in Ptw and re-fixed in 3.7% formaldehyde in Ptw for 60'. Samples were washed again in Ptw then put at 80^{o}C for 10' to kill endogenous alkaline phosphatase activity. Then, they were pre-hybridised in 500µL hybridisation buffer (see below) for 10' at RT, the hybridisation buffer was changed and they were left to pre-hybridise at 37 ^{o}C overnight. The following day probes were denatured at 85C for 10', diluted to a final concentration of 0.05ng/µL and added to the samples. Samples were hybridised over weekend then they were washes in hybridisation buffer. They were washed gradually into 2xSSC then into 0.05X SSC and finally in Ptw. Samples were washed into PBT (1xPBS, 0.2% Triton X-100, 0.1% BSA) then blocked in 1X Boheringer-Mannheim blocking buffer in maleic acid buffer (100mM maleic acid, 150mM NaCl, Ph 7.5) for 1hr at RT. They were incubated overnight at 4 ^{o}C in 1:5000 anti-Dig/AP in 1X Boheringer-Mannheim blocking buffer in maleic acid buffer. On the final day samples were washed in PBT, quickly rinsed in AP buffer (see below) and staining was developed at RT in 500 µL of AP buffer with 4µL of NBT/BCIP ready mix solution (Roche).

HYBE BUFFER	Final concentration
Formamide	50%
20X SSC (pH 4.5)	5X
20 mg/mL heparin	50ug/mL

20% Tween20	0.1%
10% SDS	1%
10mg/mL salmon sperm DNA	50 ug/mL
DEPC H_20	

AP bufffer	Final concentration
1M NaCl	100mM
1M MgCl2	50mM
Tris pH 9.5	100mM
20% Tween-20	0.5%
dH_20	

2.5 *In situ* hybridization chain reaction (HCR)

Probes were designed using the probe generator devised by Ryan Null from the Ozoplat lab (https://github.com/rwnull/insitu_probe_generator) and then ordered from IDT (Integrated DNA technologies), amplifiers were bought from Molecular Instruments and buffers were prepared following the recipe below. HCR experiments were carried out following the protocol established by Luca Santangeli (Arendt Lab). Briefly, samples were re-hydrated into PTW-DEPC (), pre-hybridized in 200µL of hybridization buffer for 30' at 37°C and then 50µL of hybridisation buffer with 8nM each of probe were added. Samples were incubated

overnight at at 37°C in a thermoblock shaking at 750rpm. The following day samples were washed four times in 0.5 mL HCR probe wash solution for 10’ at 37 then three times in 1mL 5X SSCT(DEPC) for 5’ at RT. Samples were pre-amplified in 100µL of amplification buffer for 30' at RT. At this point 2µL of each hairpin (for three probes experiments B1-H1, B1-H2, B2-H1, B2-H2, B3-H1, B3-H2) per experiments were placed in different PCR tubes. Heated in a PCR Thermocycler for 1:30' at 95°C, quickly spun down, and let to cool at RT for 30' in the dark. Then all hairpins were pooled in one tube with 50 µL per experiment of amplification buffer. Then 50µL of amplification buffer and hairpin mix was added to all tubes with the samples (final concentration of 40nM Hairpin).

Samples were incubated overnight at 25°C in a thermomixer shaking at 750rpm. The following day samples were washed three times in 1 ml 5X SSCT for 10' at RT, then stained with DAPI (final conc. = 5ug/ml) in 500µL PTW for 15’. They were washed again twice in 500µL PTW and then transferred to 2,2’-Thiodiethanol for imaging. The amplifiers used for the HCR experiments were B1-647, B2-594 and B3-488 and they were imaged using a Zeiss LSM-800 confocal microscope.

30% HCR probe hybridization buffer For 40 mL of solution

30% formamide	12 mL formamide
5× sodium chloride sodium citrate (SSC)	10 mL of 20× SSC
9 mM citric acid (pH 6.0)	360 µL 1 M citric acid, pH 6.0
0.1% Tween 20	400 µL of 10% Tween 20
50 µg/mL heparin	200 µL of 10 mg/mL heparin
1× Denhardt's solution	800 µL of 50× Denhardt's solution
10% dextran sulfate	8 mL of 50% dextran sulfate

Fill up to 40 mL with ultrapure H2O

30% HCR probe wash buffer For 40 mL of solution

30% formamide	12 mL formamide
5× sodium chloride sodium citrate (SSC)	10 mL of 20× SSC
9 mM citric acid (pH 6.0)	360 µL 1 M citric acid, pH 6.0
0.1% Tween 20	400 µL of 10% Tween 20
50 µg/mL heparin	200 µL of 10 mg/mL heparin

Fill up to 40 mL with ultrapure H2O

HCR Amplification buffer For 40 mL of solution

5× sodium chloride sodium citrate (SSC)	10 mL of 20× SSC
0.1% Tween 20	400 µL of 10% Tween 20
10% dextran sulfate	8 mL of 50% dextran sulfate

Fill up to 40 mL with ultrapure H2O

5× SSCT For 40 mL of solution

5× sodium chloride sodium citrate (SSC)	10 mL of 20× SSC
0.1% Tween 20	400 µL of 10% Tween 20

Fill up to 40 mL with ultrapure H2O

2.6 Immunohistochemistry

Larvae and embryos were rehydrated from methanol to 0.1% Triton X-100 in 0.1 M phosphate-buffered saline (PBST) by four PBST washing steps, each reducing the

concentration of methanol in PBST by 25%. Larvae and embryos were then left for 2-hour blocking step in 1% bovine serum albumin diluted in PBST (BSA solution). Primary antibody (1:250 monoclonal Mouse anti-Acetylated Tubulin antibody from Sigma, which labels stabilised microtubules and ciliated cells) and a secondary antibody (1:500 Alexa Fluor® 568 Goat anti-Mouse from Invitrogen™) were diluted in BSA solution. Primary antibody incubation took place at 4°C overnight in the dark, followed by several washes of PBST. Then secondary antibody incubation took place at 4°C overnight in the dark, followed by several washes of PBST. Additionally, 0.1 µm of the nuclear stain SytoxGreen (Invitrogen) or 1:4000 of DAPI stock solution was added during the final wash to specimens for 15 min and rinsed with PBST several times.

2.7 Cell dissociation *C. gigas*

Crassostrea gigas samples at the right developmental stage (16h, 17h and 22h for 25C) were collected on a 20µm filter mesh and collected in Low binding tubes. Samples were first spun down at maximum speed for 30 seconds, water was quickly removed and substituted with no Calcium no Magnesium seawater (NoCaNoMg-ASW).

This passage doesn't cause embryos to stop swimming as it does in other species. Animals were then spun down at 3.8 rpm for 30 seconds NoCaNoMg-ASW was removed and fresh NoCaNoMg-ASW was added. Animals were spun down again this time at 3.2 rpm for 30 seconds and most of the water was removed. Animals were placed in a 4x4 well and were left in the solution for 3-5 minutes. After this time 300 µL of 0.5% Pronase (Roche cat # 10165921001) and 1% sodium thioglycolate (Sigma T0632) in Low CaNoMg-ASW seawater (LowCaNoMg-ASW) were added and the solution was gently pipetted up and down to mix.

After 3 minutes 10 µL of 5mg/mL Liberase (Roche, cat # 05401119001) were added. The solution was mixed by gently pipetting up and down. After another 2 minutes, very gentle manual trituration was initiated with a 200 µL pipette set to 150µL. Dissociating embryos were forced through a very small gap in between the pipette tip and the well bottom several times until most cells were dissociated. The trituration was stopped after 20' since the first enzyme was added and the solution was transferred to a new low binding tube.

2.8 Cell dissociation *P. crozieri*

1-day old larvae of *P. crozieri* were collected in a 40µm filter and washed several times with filtered ASW. After cleaning the larvae were washed several times with NoCaNoMg-ASW to prepare for dissociation. Larvae were collected in the centre of the mesh and transferred to a plastic cell culture petri dish and most of the NoCaNoMg-ASW was removed by pipetting. 300µL of 1:100 solution of Prot14 (3.5u/mg; Sigma P5147) in LowCaNoMg-ASW previously activated at 37C for 1h were added. The solution was pipetted gently for 5-10' until most of the larvae were dissociated. After this time, a few small orange structures were left undissociated, possibly the larval gut. These were collected with the pipette and gently triturated until most dissolved. The whole dissociation process usually lasted around 15 minutes. Samples were resuspended in 1 mL of elution solution (see table below for details) then the cells were spun down for 3' at 3.2 rpm.

Sample Id	Cg1-Cg6	Cg?-Cg7	Cg8-Cg9-Pc1-Pc4
Elution Solution	LowCaNoMg-ASW	NoCaNoMg-ASW	NoCaNoMg EDTAfree-ASW

The supernatant was removed and 500µL of elution solution were added. The pellet was gently resuspended and then the cells were spun down again for 3' at 3.2 rpm. The supernatant was removed and cells were resuspended in 60-200 µL of elution solution depending on the cell concentration. The pellet was resuspended gently. The solution was then filtered twice through a 20µm filter mesh (for *C. gigas*) or a 40µm filter mesh (*for P. crozieri*) to remove big clumps of undissociated tissue. 10µL of the solution were stained with 1:500 of 11 µm of Fluorescein Diacetate (e.g. Sigma F7378) and 1mg/ml of Propidium Iodide (e.g. Sigma P4170) and cell viability was quickly assessed under an AxioImager.M1 microscope. An additional 10µL of cells were counted on a hemocytometer to assess the solution concentration. About 30'000 cells were loaded into a 10x chip following manual instructions. Using the 10x Chromium controller and Chromium single cell 3' Kit v2, v3 or v3.1 (Cat #120237, 10x Genomics, USA) (see table below).

Sample Id	Cg1-Cg6-Cg8-Cg9	Cg?-Cg7	Pc1-Pc4
Kit version	v2	v3	v3.1

Different elution solutions were used to load the samples to try and trouble shoot salt concentration (see table below) the result of these different elutions are described in chapter 5.

Sample Id	Cg1	Cg2-Cg6	Cg?-Cg7	Cg8-Cg9
Final elution solution	LowCaNoMg-ASW eluted in water	LowCaNoMg-ASW	NoCaNoMg-ASW	NoCaNoMg EDTAFree-ASW

Sample Id	Pc1-Pc3	Pc2-Pc4		
	NoCaNoMg EDTAFree-SW	NoCaNoMg EDTAFree-SW Eluted in water		

cDNA synthesis and library preparation were carried out according to manufacturer's recommendation.

Post library quality control was determined on the Qubit Fluorometer with the Qubit dsDNA HS Assay kit and a 1:10 or 1:5 sample dilution was run on Agilent 4200 TapeStation system with the High sensitivity D1000 ScreenTape and High Sensitivity D1000 reagents. Post-library quantification was performed with Illumina Library Quantification Kit. Single cell libraries were sequenced on an Illumina NextSeq500 using a 2x75 paired-end kit.

NoCaNoMgSW: 495 mM NaCl, 9.7 mM KCl, 26.6 mM $NaHCO_3$, 5mM EDTA, 50mM Tris-HCl in H_20. Adjust final pH to 8.0 and filter through 0.22 µm filter.

LowCaNoMg-SW: 460 mM NaCl, 10mM KCl, 1mM $CaCl_2$, 10mM HEPES. Bring the pH to 7.6 and filter through 0.22 µm filter.

NoCaNoMgEDTAFree-SW: in 1 liter of distilled water add 31g NaCl, 0.8g KCl, 0.29g $NaHCO_3$, 1.6g Na_2SO_4. Bring the pH to 8.

2.9 Demultiplexing and mapping of single cell reads

Individual 10x sample libraries were multiplexed using Cell Ranger Makefastq v3.0.2 with default settings. Reads for each sample were mapped and demultiplexed by cell barcode and UMI using Cell Ranger Count v3.0.2.

2.10 QC and Clustering of cells using Seurat v3

After loading the single cell count matrix of the different samples, we checked distribution of UMI, gene reads and mitochondrial gene content across cells. We performed clustering using Seurat v3 and visualize clusters with the UMAP function.

2.11 Re-annotation of *Crassostrea gigas* genome

Existing gene annotations of *Crassostrea gigas* lacked sufficient UTR annotations (Zhang et al, 2012), such that many reads of the 10x Genomics runs could not be mapped to genes. Therefore, a combination of bulk RNAseq and the 10x Genomics single-cell RNAseq data was used to re-annotate the *C. gigas* genome, this worked was carried out by Daniel Leite from the Telford Lab. Single ended bulk RNAseq data were collected from Zhang *et al* (2012) (SRR334222, SRR334223, SRR334224, SRR334225, SRR334226, SRR334227, SRR334228, SRR334229, SRR334230, SRR334231, SRR334232, SRR334233, SRR334234, SRR334235, SRR334236, SRR334237, SRR334238, SRR334239, SRR334240, SRR334241, SRR334242, SRR334243, SRR334244, SRR334245, SRR334246, SRR334247, SRR334248, SRR334249, SRR334250, SRR334251, SRR334252, SRR334253, SRR334254, SRR334255, SRR334256, SRR334257, SRR334258, SRR334259). This data and the 10x Genomics samples were trimmed with Trimmomatic v0.39 (Bolger et al, 2014) using the following settings, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:35. Prior to mapping reads, the soft-masked version of the *C. gigas* genome (RefSeq assembly: GCF_000297895.1) (Zhang et al, 2012) was modified, as the mitochondrial genome was identified, fragmented in three parts, within Scaffold 161. To amend this mis-assembly the mitochondrial genome of *C. gigas*

(NC_001276.1) was aligned to scaffold 161 with BLASTN v2.8.1 (Altschul et al, 1990). Aligned regions were removed from Scaffold 161 and the remaining scaffold fragments were renamed. Finally, the mitochondrial genome (NC_001276.1) was added to the modified genome. Trimmed reads were then mapped to the modified *C. gigas* genome using STAR v2.5.3a (Dobin et al, 2013) with default parameters and the --outSAMtype BAM Unsorted option. The BAM alignment output was then sorted with Samtools v1.9 (Li et al, 2009). The aligned and sorted BAM file was then used as the input for *ab initio* gene prediction with BRAKER v2 (Hoff et al, 2016; Lomsadze et al, 2014; Barnett et al, 2011; Stanke et al, 2006, 2008; Camacho et al, 2009; Li et al, 2009; Altschul et al, 1990) with the following options, --UTR=on, --crf, --softmasking, --gff3 and --rounds 15. All gene predictions made by BRAKER for the mitochondrial genome were removed and replaced by mitochondrial gene annotations made with MITOS v2 (Bernt et al, 2013).

3 Single cell sequencing in the pacific oyster *Crassostrea gigas*

3.3 Single cell sequencing in the pacific oyster *Crassostrea gigas*

As discussed in chapter 1 the aim of my project is to compare the transcriptional signatures of the cells of Lophotrochozoan larvae. Specifically, I am going to compare the classical trochophore larva of a mollusc, the pacific oyster (*Crassostrea gigas*), with the (probably) more derived Muller's larva of a polyclad flatworm (*Prosthecereaus crozieri*) to try and understand if A) common characters of larvae (such as the ciliary bands or the apical organ) share similar transcriptional signatures and/or B) they have other similar cell types and ultimately C) whether they are homologous.

The first step in this process is to gather transcriptional data for the cells of the two larvae which I obtain using single cell sequencing technology. After performing single cell sequencing data for both larvae, I had firstly to map the reads onto the reference genomes, then analyse the mapped read data to make sure it was of good quality and finally to use the transcriptional profiles of single cells to build groups of cell types (cell clusters) to compare across the two animals. This chapter will concentrate on the initial mapping of the single cell sequencing data onto the genome and on the quality assessment carried out on the different oyster single cell sequencing experiments (runs or captures). In total I performed nine single cell sequencing experiments which will be referred to, from now on, with the initial of the species (Cg) and a chronological number (Cg1-Cg9).

3.3.1 Mapping of single cell reads to *Crassostrea* genome

To gather information about gene expression in different cells the first step is to map the single cell reads to the reference genome. This allows us to match the short reads from the 3' ends of the transcripts that were barcoded in each cell with the genes they belong to. Reads from samples Cg1-Cg6, which were the first ones we obtained, were initially mapped against the *Crassostrea gigas* genome published by Zhang et al in 2012. The results of this mapping showed that between 45-65% of the reads mapped correctly to the reference genome (see figure 9). This relatively low overall mapping could be due to poor genome quality, high polymorphism in the different populations of oysters as well as to reads mapping to more than one location, and hence being discarded. However, what was really worrying was that as few as 3% of reads mapped to exonic regions and a considerable number of reads (up to 35%) were mapping to intergenic regions. This result was very problematic because any read mapping to an intergenic region gets automatically discarded and cannot be used for downstream analysis.

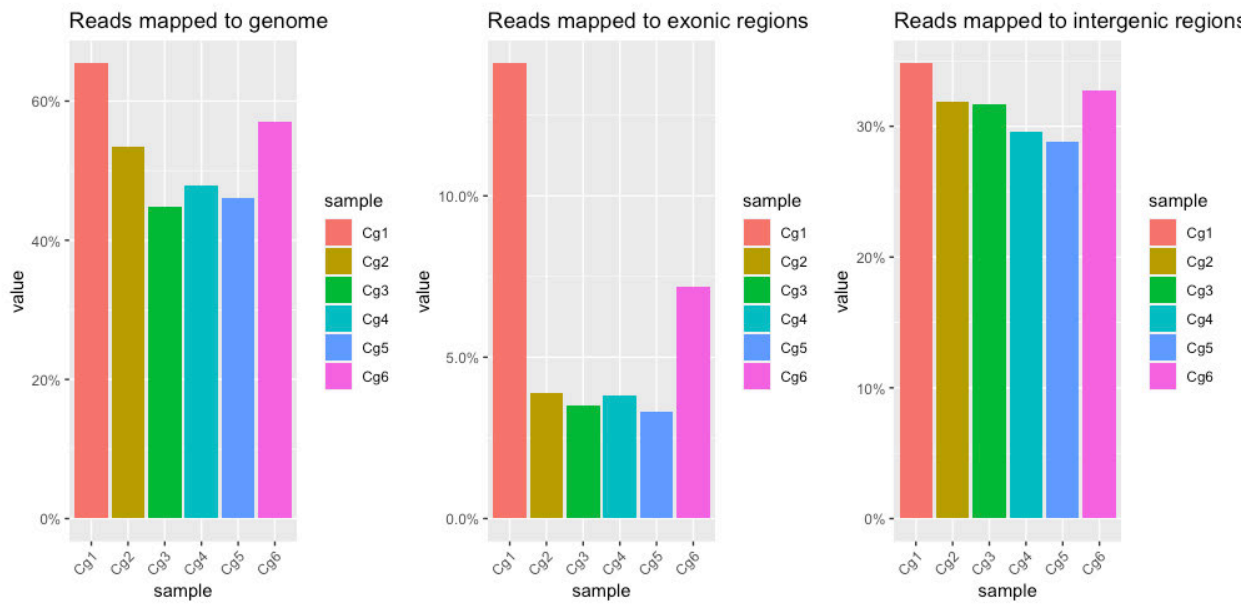


Figure 9. scRNAseq reads mostly map outside exonic regions.

Percentage of reads from different single cell sequencing runs mapping to genome, exonic and intergenic regions. Overall genome mapping ranges from 45% to 65% however reads mapping to exonic regions are very low (maximum to 15%). This is problematic since only reads mapping to exons can be used for downstream analysis.

I thought that this low proportion of reads mapping to exons could be caused by the fact that we used a 3' 10x Genomics kit which only captures the 3' end of a gene, specifically the three prime untranslated region (3'-UTR) which is the region of a gene after the translation termination codon. I hypothesized that the 3' UTR of some genes could be missing from the Zhang et al. (2012) annotations as this would cause reads mapping to these areas to end up in intergenic regions. To test this idea, Daniel Leite in our lab re-annotated the genome using not only transcriptomic data but also our single cell reads which would specifically target the 3' UTR regions of the genes (for more details see 2.12).

Figure 10 shows how that this re-annotation drastically improved mapping, specifically boosting the overall mapping of reads to the genome to above 70% for most samples, bringing the mapping to exonic regions up to 50% and reducing mapping to intergenic

region to less than 5%. This allowed us to recover many more reads that would have previously been discarded and consequently this improved the total number of cells recovered, total number of genes recovered as well as the median number of genes and UMIs per cell (see figure 11). What was incredibly striking to see was that most of these values (i.e. total cell number, total gene number, median genes per cell and median UMIs per cell) roughly doubled, which meant we got double the amount of data than we initially thought we had.

To confirm that the improvements in mapping had indeed been caused by re-annotation of the 3' UTRs Daniel Leite checked the overall size of exons, and lengths of proteins both of which showed an overall increase (see figure 12A and 12B). Moreover, we were able to see throughout the genome that the new annotation captured 3'UTRs previously missed (see figure 12C).

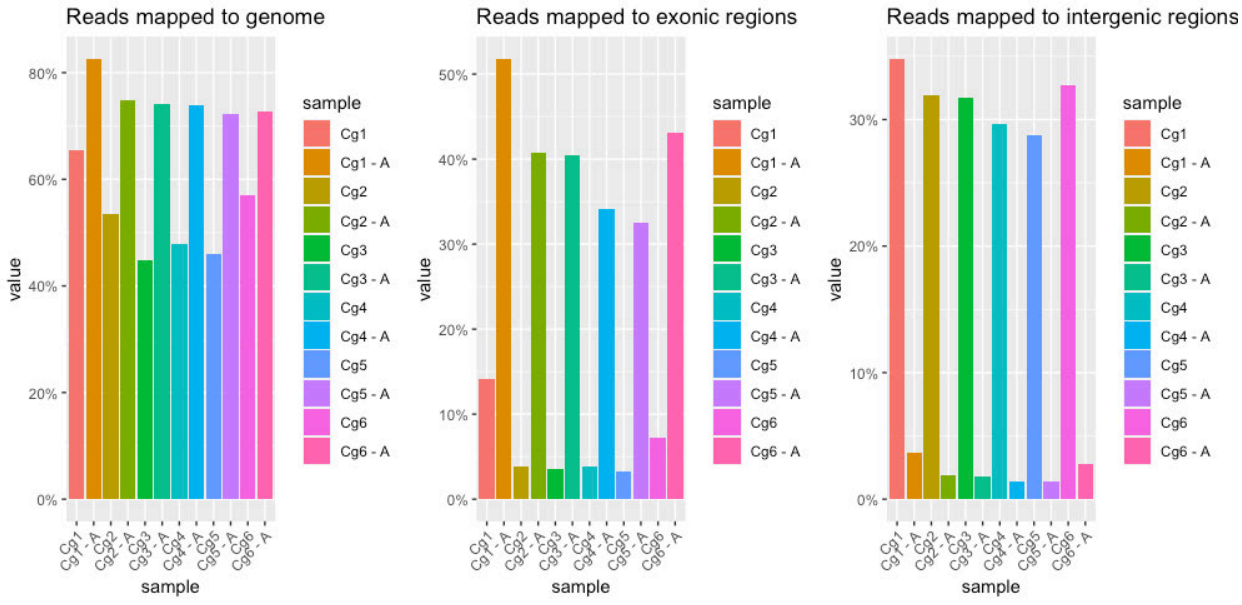


Figure 10. Genome re-annotation helps mapping reads to exonic regions.

Re-annotation using single cell sequencing 3' data (Cg1-A, Cg2-A etc..) vs previous genome annotation (Cg1,Cg2 etc…) improves overall mapping to genome (up to 80%), increases

mapping to exonic regions (up to 50%) and decreases mapping to intergenic regions (<5%). This means many reads that were previously discarded (not mapped to exons) can now be used for downstream analysis.

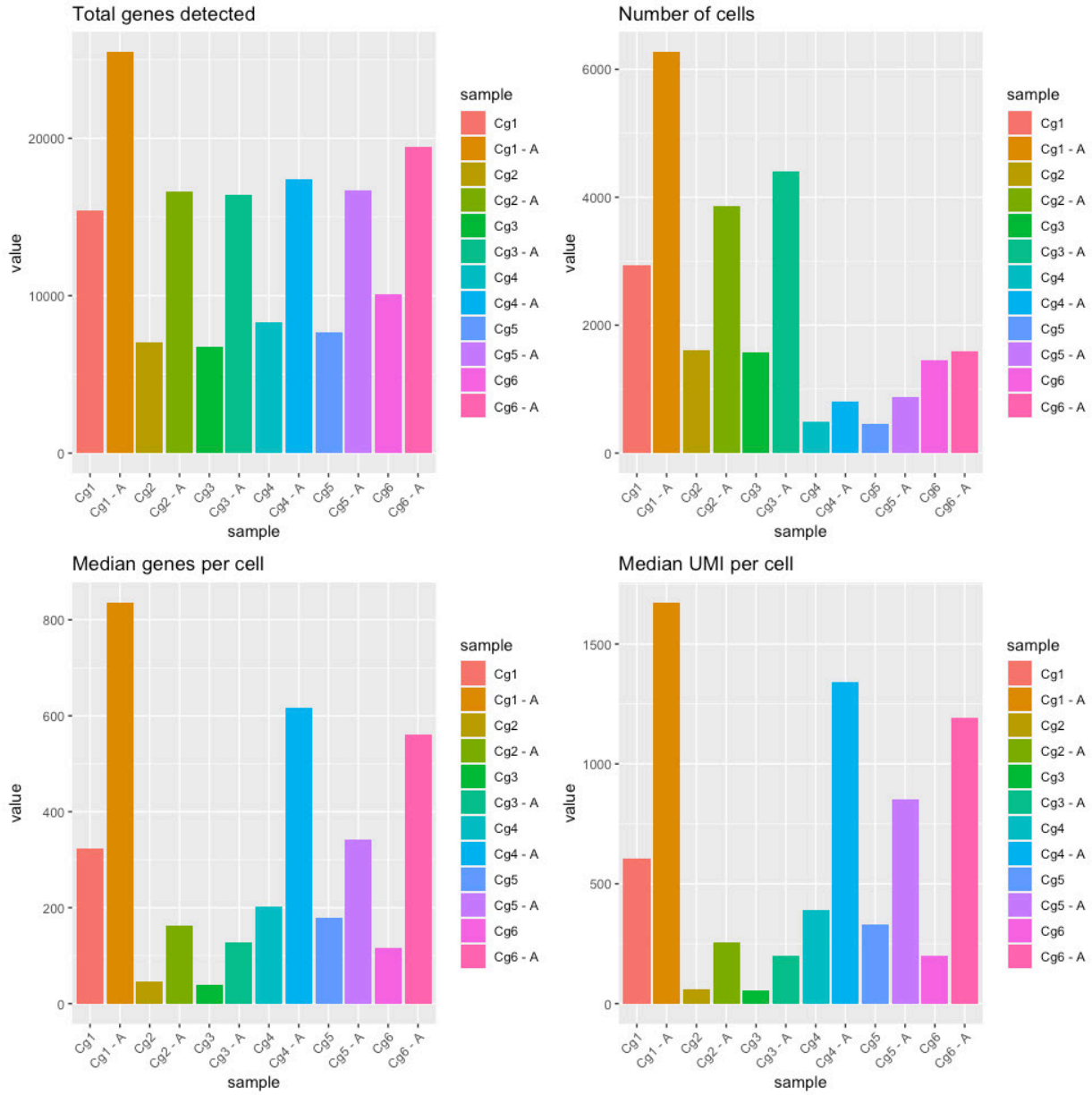


Figure 11. Genome re-annotation improves number of gene and cell recovered as well as quality of cells.

Re-annotation using single cell sequencing 3' data (Cg1-A, Cg2-A etc..) vs previous genome annotation (Cg1,Cg2 etc…) also improves total number of genes recovered (up to 25K), number of cells detected (up to 6K), median genes per cells (up to 800) and median UMI per cell (up to 1500).

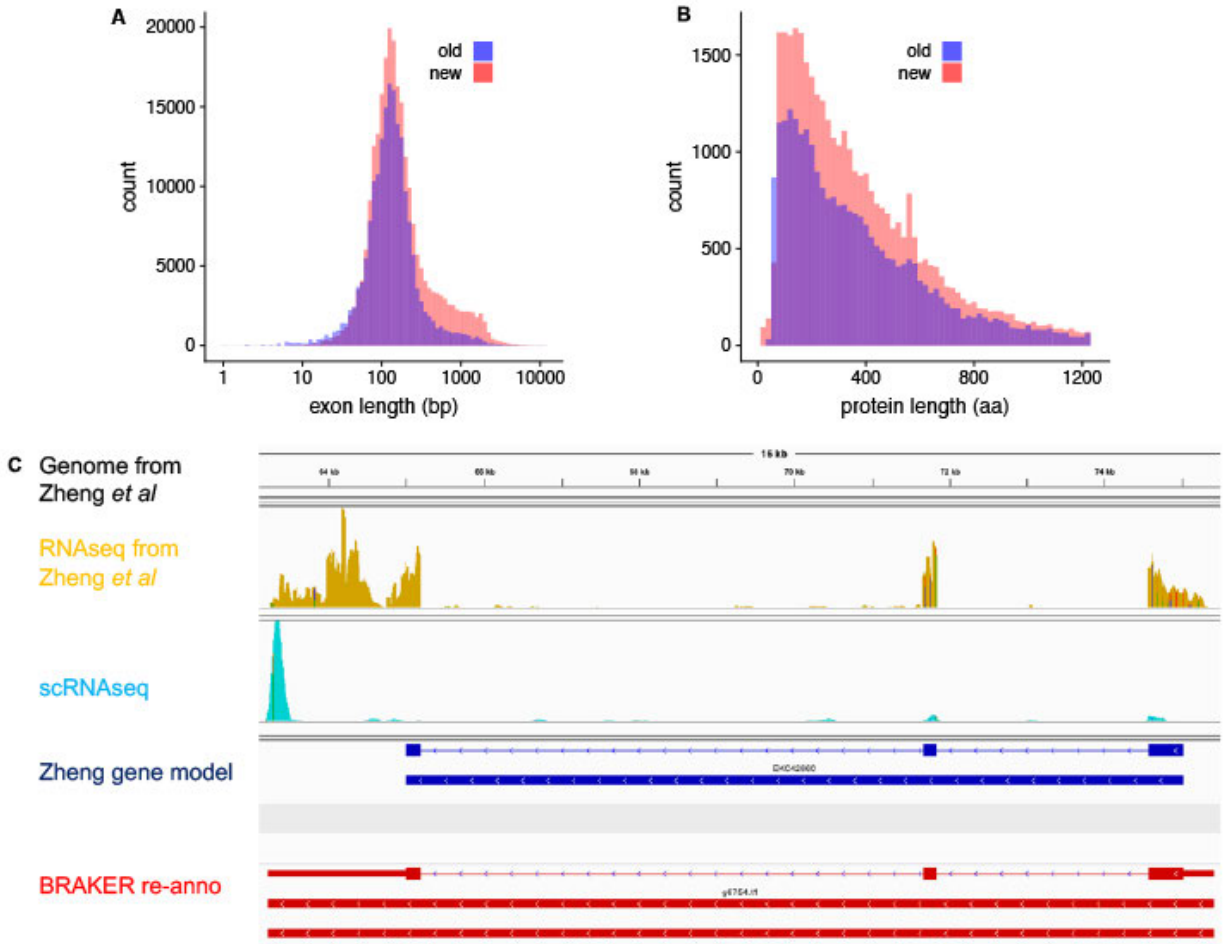


Figure 12. Re-annotation using single cell sequencing 3' data produces longer exons (A) longer proteins (B) and captures 3' UTRs of gene that were previously missing (C)

A shows the number of exon (y axis) ordered by exon length (x axis) in the old annotation (blue) vs new annotation (red). B shows the number of proteins (y axis) ordered by protein length (x axis) in old annotation (blue) vs new annotation (red). C shows the alignment to the reference genome of the RNAseq reads from Zhang et al (2012) in yellow, the scRNAseq in light blue, the old annotation in blue and the new annotation in red.

3.3.2 Quality control among different batches

Altogether, the re-annotation using 3' biased single cell data greatly improved the read mapping however, it also highlighted big differences across samples, with Cg1 showing substantially better results than all the other runs (see figure 10, 11 and 13). This difference in quality across samples will be discussed in the depth in the rest of this chapter. Before

entering the details of the quality differences between the repeats I will explain the main technical differences among them. Briefly, Cg1 was the first single cell capture performed from larvae that had developed at 25ºC for 15h. The dissociation, capture and library preparation of this sample were carried out during a course with the help of members of the Arendt lab (EMBL) as well as of the EMBL Gene Core facility. Samples Cg2 and Cg3 were similarly developed at 25ºC for 15h however, due to a cell counting mistake, 5 times more cells than usual (we usually load 30,000) were loaded onto the microfluidics chip. Samples Cg4 and Cg5 were developed at 25ºC for 16h and captured to make up for the mistakes made for Cg2 and Cg3. Cg2 and Cg3 in fact show a higher number of cells captured (as opposed to Cg4 and Cg5), however they might be expected to have an increased number of multiplets (droplets that contain more than one cell) according to 10x chromium instructions. Cg6 was developed at 25ºC but captured at a later point in development (22h) to try to compare later larval stages with the earlier data. The preparation of sequencing libraries for samples Cg2-Cg6 was performed by members of the Arendt lab. Since we noticed varying qualities across samples, at a later date I captured samples Cg7-Cg9 (in the summer of 2019). All dissociations and captures were carried out in the Arendt lab, developed at 20ºC for 20h, as this seemed to improve synchronicity of the larvae, and cells were eluted in different buffers before loading (NoCaNoMg-ASW, LowCaNoMg-ASW with or without) in an attempt to optimise cell viability.

As previously mentioned, although re-annotation of 3' UTRs greatly improved mapping and increased cell numbers and genes recovered per cell, in all of these samples we still noticed a very high-quality difference across them (see figure 10, 11 and 13). In particular sample Cg1 showed the highest number of cells (~6000) and genes (~25000) detected as well as the

highest number of median genes (~800) and median UMIs (~1500) per cell. As for the other experiments, Cg2 and Cg3 show quite a high number of cells recovered (~4000) but low median number of genes (<200) and low median number of UMIs (<250). Cg4 and Cg5 on the other hand show very low cell number (<1000) but higher median genes and UMI content with Cg4 looking overall better than Cg5 despite the two being technical replicates (same cell dissociation loaded into separate wells of the microfluidics chip). Cg 7 and Cg9 show very poor quality overall with low cell numbers (<1000), low median genes (<100) and low median UMIs per cell (<200). Finally, Cg8 has approximately 4000 cells with ~400 median genes and ~500 median UMIs.

A combination of factors could have led to this incredible variability amongst the experiments some of which I will explore here.

1) Technical errors in library preparation. During the preparation of the cDNA library for samples Cg2-Cg6 the PCR run for the library preparation failed. For this reason, our colleagues in the Arendt lab had to prepare a new library with the cDNA that was left (>3uL instead of the suggested 35uL). To see how much this had impacted on the quality of the samples I decided to count the number of UMIs (or unique molecular identifiers) that we recovered per sample. This is a good measure of how much RNA was actually captured and sequenced per sample because UMIs are not impacted by PCR bias. As we can see in figure 13 Cg1 has about 10 times more total UMIs than all of the other samples. Since the software we use for mapping and to assign the barcodes to different cells uses the UMI content to tell the difference between a droplet that contained a cell and an empty droplet, we only retain cells with more than a minimum number of UMIs. This result can explain the higher recovery of cells,

genes and UMI as well as genes per cell in Cg1. Moreover, this shows that the difference in cDNA input greatly impacted the quality of data. However, we can clearly see that even samples that did not suffer the PCR run failure (i.e. Cg7, Cg8 and Cg9) still have substantially lower UMIs than Cg1 which made us reflect on what else could be causing the differences we see.

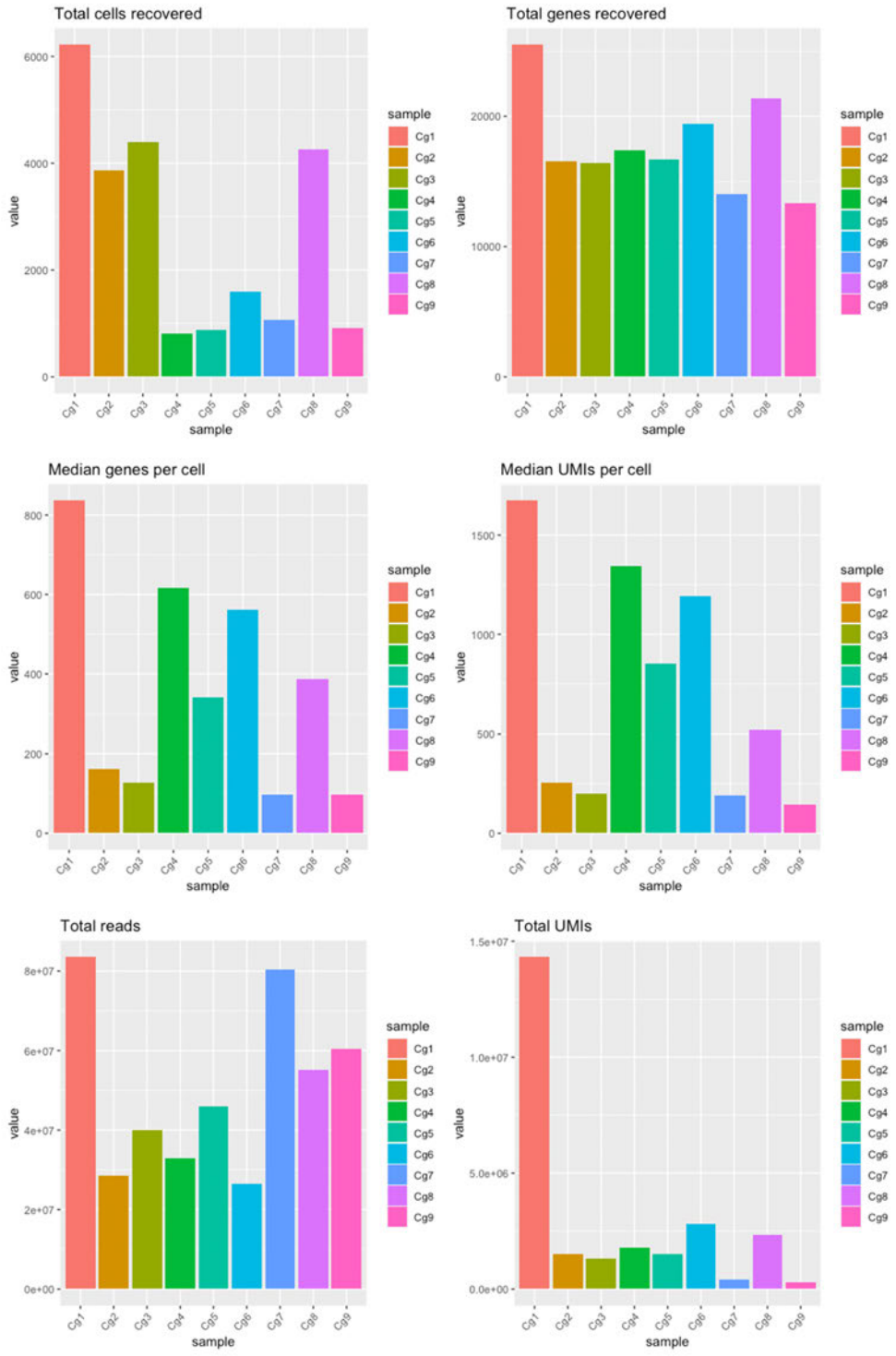


Figure 13. Different scRNAseq samples show very different quality.

Bar graphs showing total number of 1) cells and 2) genes recovered, 3) median genes and 4) UMIs, 5) total reads and 6) total UMIs per sample highlights differences in qualities between different repeats with sample Cg1 showing highest values overall (~6K cells, ~25K genes, ~800 median genes per cell, >1500 UMIs per cell…)

2) Experimental conditions resulting in low cell viability. The second possible cause for variability across samples is cell viability. Dying cells with ruptured membranes lose most of their cytoplasmic content causing an over-representation of mitochondrial genes normally inaccessible for sequencing (since they would be stored inside the mitochondria). Calculating the percentage of mitochondrial reads per cells allow us to assess cell viability *a posteriori*. It is worth noting that with the original Zhang et al genome, it was not possible to correctly measure this variable since the mitochondrial genome was only partially included and was fragmented in the genome. It was hence necessary first to manually extract all the fragments of mitochondrial genes from the genome and then re-annotate them separately (carried out by Daniel Leite). Figure 14 shows the percentage of mitochondrial genes across cells, and, as expected, this appears to be very variable among batches. More specifically samples Cg2 and Cg3 show almost no cells with mitochondrial gene content lower than 5% and Cg4 and Cg5 show many cells with mitochondrial percentage higher than 80%. This difference was probably caused by overly rough mechanical trituration of the samples Cg2-6. Cell dissociation for sample Cg1 was prepared during a Single cell sequencing course and was carried out with the help of Paola Bertucci from the Arendt Lab who has great experience with cell dissociation protocols. Cg7 and, more so, Cg9 show a group of cells with low mitochondrial gene content and a second group with varying percentages. Cg8 looks overall similar to Cg1 with most cells showing less than 10% of mitochondrial gene content. By looking at two other parameters - genes and UMIs content per cell (shown in figure 7,8) - we can see that samples with high mitochondrial gene content (i.e. Cg2, Cg3, Cg7 and Cg9) tend to have lower nuclear gene content and lower UMIs content. Moreover,

figure 13 clearly shows that cells with high mitochondrial gene content also have lower UMIs per cell confirming once again that high mitochondrial gene content corresponds to poor overall cell quality.

To correct for the difference in cell viability I carried out all further clustering analysis after first removing all cells that had more than 10% of mitochondrial gene content (see fig. 15).

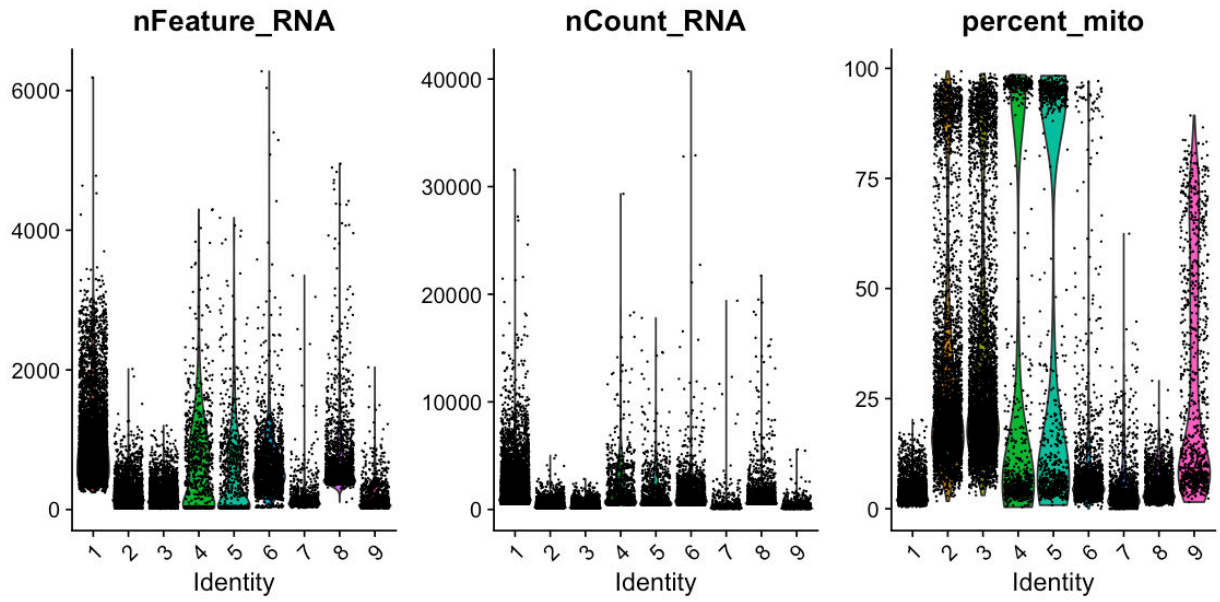


Figure 14. High Mitochondrial gene content correlates with poor outcomes.

Violin plots of genes (NFeature_RNA), UMI (nCount_RNA) and percentage of mitochondrial genes (percemt_mito) per cell in samples Cg1-9 shows that A) there is high variability in mitochondrial percentage across cells and batches B) samples with high mitochondrial gene content (Cg2, Cg3, Cg7 and Cg9) also tend to have lower gene an UMI content per cell. 1-9 correspond to Cg1-Cg9.

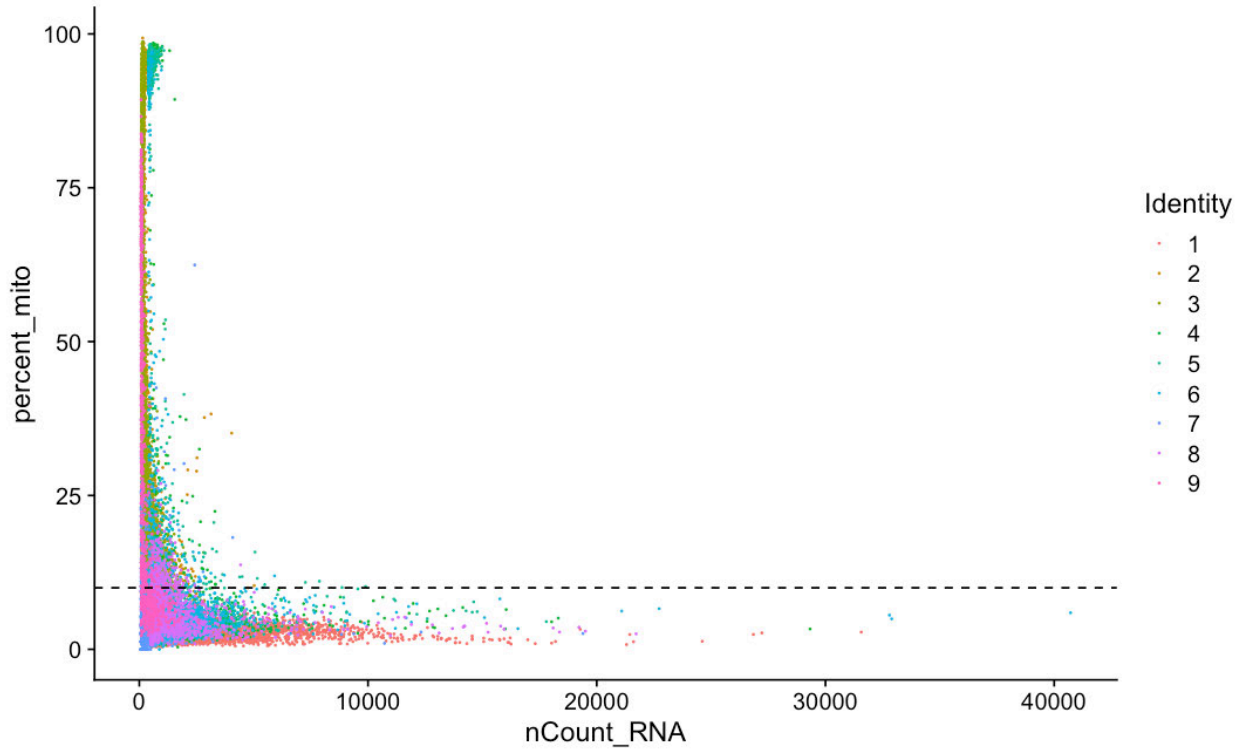


Figure 15. High quality cells have lower mitochondrial gene content.

Scatterplot of mitochondrial gene content (percent_mito) versus UMI (nCount_RNA) per cell across different samples showing that high quality cells (with high UMIs number) have lower mitochondrial gene content. Dashed line indicates 10% mitochondrial gene content which will be set as a threshold for all downstream analysis. 1-9 correspond to Cg1-Cg9.

3) Capture buffer containing EDTA affecting PCR steps. The PCR failure as well as the differences in viability likely caused differences in quality across the different experiments. However, these causes do not fully explain the differences across samples. A final possible cause of the observed differences in quality is the use of different buffers for cell elution before loading. Samples Cg7-Cg9 were processed in the summer of 2019, one year later than samples Cg1-Cg6. Since we had noticed a high variability in cell viability, I tried to optimise trituration as well as modifying the buffer used to elute cells before capture. Samples Cg2-Cg6 were eluted in LowCaNoMg-ASW, however I realised that eluting cells in NoCaNoMg-ASW improved cell viability. For this reason, sample Cg7 was eluted in NoCaNoMg-ASW.

Surprisingly, this led to a very low cDNA yield and an overall low UMI count and so a small cell number (see figure 13). I believe that this result was caused by the presence of EDTA in the NoCaNoMg-ASW which impedes the retro transcription of mRNA to cDNA. For this reason, for samples Cg8 and Cg9 I changed the buffer to NoCaNoMg-EDTA-free-ASW. This gave us better results for sample Cg8 although not for sample Cg9.

Altogether, I think that many different factors can play a role in the success of a single cell sequencing experiments and having the possibility to repeat some of these results with changing condition proved very interesting. It is worth noting how none of these factors fully explain the incredible quality of sample Cg1 which I like to think was mostly down to beginner's luck.

At this point, before applying any cut-off, I had approximately 21500 cells expressing a total of ~38000 genes across 9 different samples with very different qualities. In the next sections I'll discuss what filtering parameters I used to select good quality cells across samples and how I performed clustering analysis.

3.3.3 Sample integration

Since I have shown in the previous paragraphs how mitochondrial gene content seems to be inversely correlated with overall sample quality I decided to discard all cells with more than 10% mitochondrial gene content. I also only selected cells that contained at least 200 genes. After these initial filtering steps, I had 10505 remaining cells expressing a total of ~38k genes and I processed the data using a standard Seurat pipeline with the following steps.

First, to account for different sequencing depths across cells I normalized each gene count per cell by the total gene count for that cell. I then calculated the 2000 most variable genes and went on using only those for downstream analysis to reduce dimensionality. I scaled the data to shift the expression of each gene so that the mean expression across all cells was 0 and scaled the expression of each gene so that their variance across all cells was 1. This helps to avoid highly expressed genes dominating the clustering. Finally, I regressed out UMI counts and mitochondrial gene content thus inhibiting the effects of these biases on clustering.

To analyse cells' distributions in the expression space I needed to further reduce the 2000 dimensions (from the 2000 genes with most variable expression across cells) by performing principal component analysis (PCA). The idea of a PCA is to identify sets of linearly related genes (Principal Components or PCs) to condense the information brought about by our 2000 most variable genes. Once PCs are calculated it is important to decide how many of these components bring meaningful information, this makes downstream computational analysis quicker and also reduces noise. To decide how many PCs are relevant I used the JackStraw function which randomly permutes a subset of the data (i.e. creates a smaller expression matrix using a subset of genes) and then calculates the projected PCA scores for each "random" gene. It then compares these "random" scores with the real observed PCA scores to determine statistical significance. The final result is a p-value for the association between each gene and each principal component, as one can see in fig. 16, the first 50 PCs all have significant p-values so I decided to use all of them, we could have possibly used more but it would be unusual for this type of analysis.

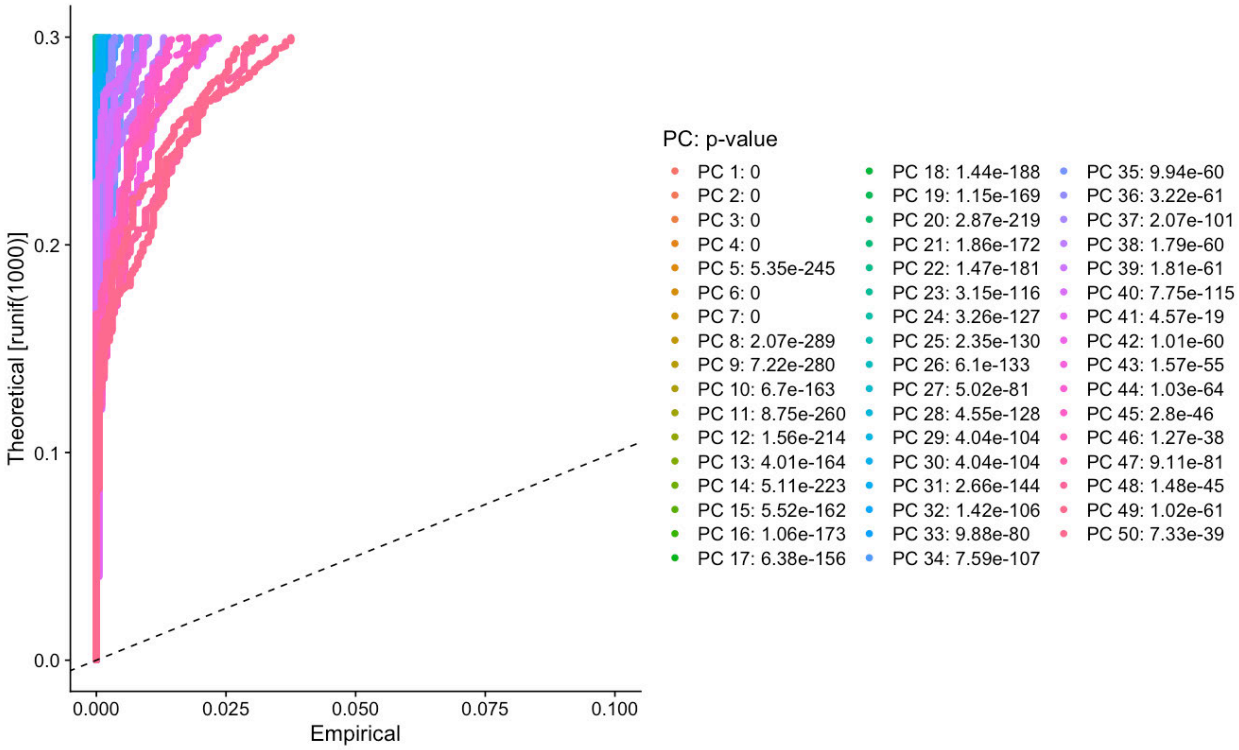


Figure 16. JackstrawPlot comparing the distribution of p-values for all genes across each PC, compared with a uniform distribution (dotted line), showing that all 50 PCs are relevant.

Once I set the number of PCs that are significant (in this case 50) we can start to explore this reduced multidimensional space to find out how cells are grouped. In Seurat v3 this is done by the function `FindNeighbors` which constructs a Shared Nearest Neighbor (SNN) graph of cells based on the selected PCs. Once I have computed the SNN graph I can identify clusters by partitioning it into local neighborhoods based on the degree of connectivity among the cells. This is done by the function `FindClusters` which has a parameter called 'resolution' that allows the user to change the number of clusters obtained. I initially set the resolution to 1 which gave me 31 clusters. I could then produce a UMAP plot which is a 2D representation of the multidimensional space (with its 50 PCs), in this graph, every dot is a cell and each colour shows a different cluster (see fig. 16 A). However, what I realised when

colouring the cells by their experiment of origin is that cells derived from different batches (Cg1, Cg2, Cg3, Cg4, Cg5, Cg6, Cg7, Cg8 and Cg9) appear to cluster separately (see fig. 16 B).

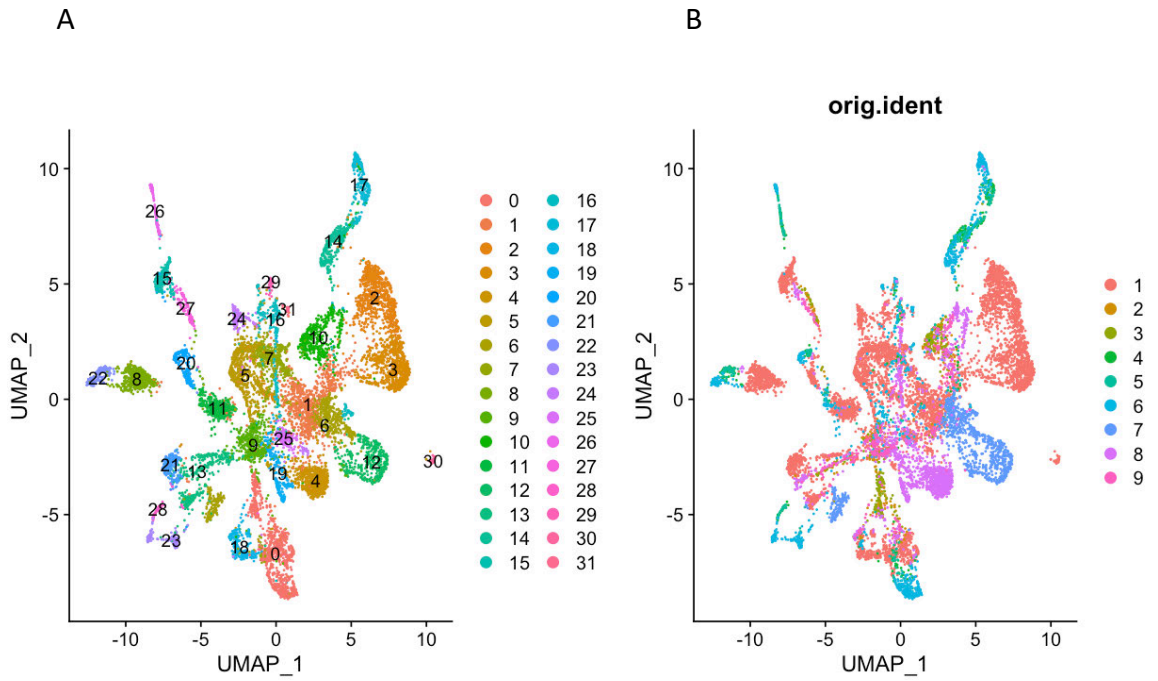


Figure 17. Cells separate out by sample of origin.

UMAP of single cells from samples Cg1-9 coloured by cell clusters showing 31 distinct clusters for resolution 1 and by sample of origin showing that cells cluster separately by sample. A- colouring represents different clusters (numbered 0-31) B- colouring represents different samples: 1= Cg1, 2=Cg2 etc.

The fact that each sample clusters separately is not entirely surprising considering the differences in quality among the batches, however we do not expect our batches to contain different cell types, apart from possibly sample Cg6 which was captured at a later developmental time point. For this reason, I decided to apply the integration pipeline of Seurat v3, which has been developed to integrate cells across individuals, technologies, and experiments. Briefly, the samples are treated individually as described before, up until the stage at which the variable genes are calculated, then the function `FindIntegrationAnchors` is used to find correspondences ("anchors") across single-cell datasets using the Mutual Nearest Neighbors (MNN) algorithm. It is worth noticing that these anchors are single cells,

so this method assumes that the different samples have at least some common cells. Then the list of anchors is passed to the `IntegrateData` function that will return a `Seurat` object with the "integrated" expression matrix (see fig. 18)

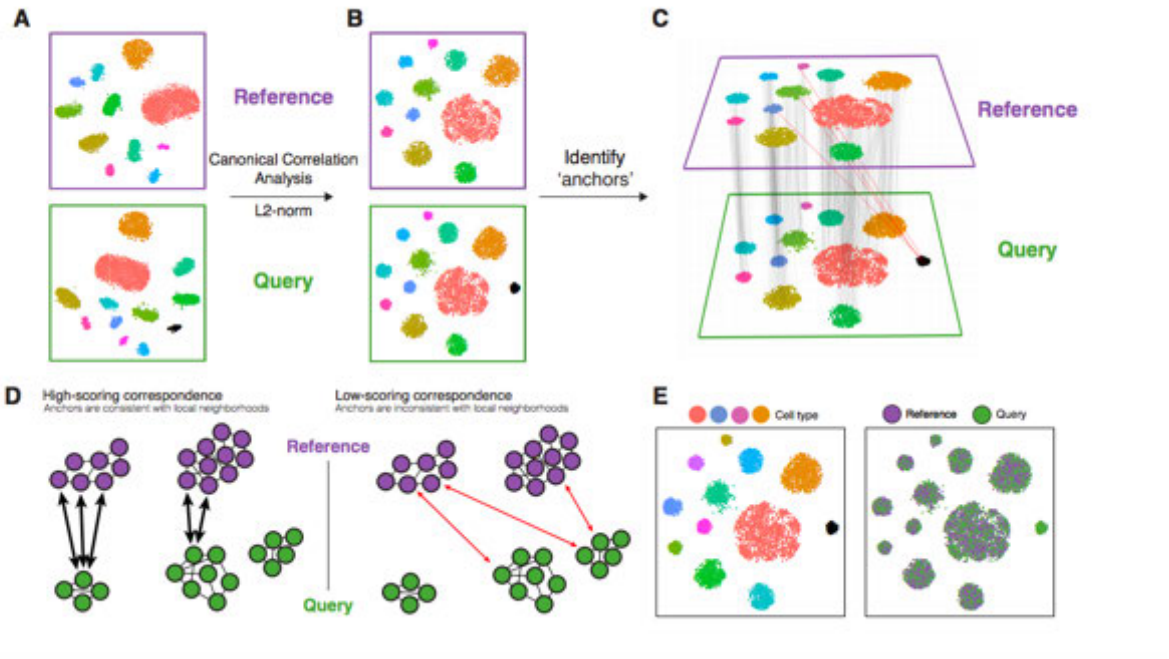


Figure 18. Schematic overview of the integration workflow in Seurat v3 from Stuart et al. 2019.

(A) UMAP representation of two SCS datasets (reference and query) from separate single-cell experiments. The two datasets share all cells clusters except for the black (only present in the “query). (B) Projection of the two datasets into a shared subspace defined by shared correlation structure after CCA and normalization of the canonical correlation vectors. (C) Identification of anchors between the two datasets. Most connections are given by biologically meaningful connection (grey lines) but sometimes there are “incorrect” anchors at low frequency (red lines). (D) For each anchor pair a score based on the consistency of anchors across the neighbouring cells of each dataset is given (here red anchors give lower scores). (E) Anchors and their scores are used to compute “correction” vectors for each query cell, transforming its expression so it can be analysed in the integrated space.

After integration the UMAP plot shows that most cluster are made up of cells from all 9 samples (fig. 19), moreover the clusters are very similar to those I obtained for Cg1 alone (see fig. 20). In fact, if I take the integrated clustering, filter it to only keep cells from Cg1 and then use the colours from the clustering I obtain from Cg1 sample alone almost all cells

that belonged to the same cluster stick together. This could either mean that the integration works well or that, since most cells belong to Cg1, they somehow drive the clustering (see figure 21). To test between these two possibilities, I re-ran the integration using just 1400 randomly selected cells from Cg1. As one can see in figure 22, clusters remain very similar after subsampling of Cg1, showing that the difference in cell number among samples is not driving the clustering during integration.

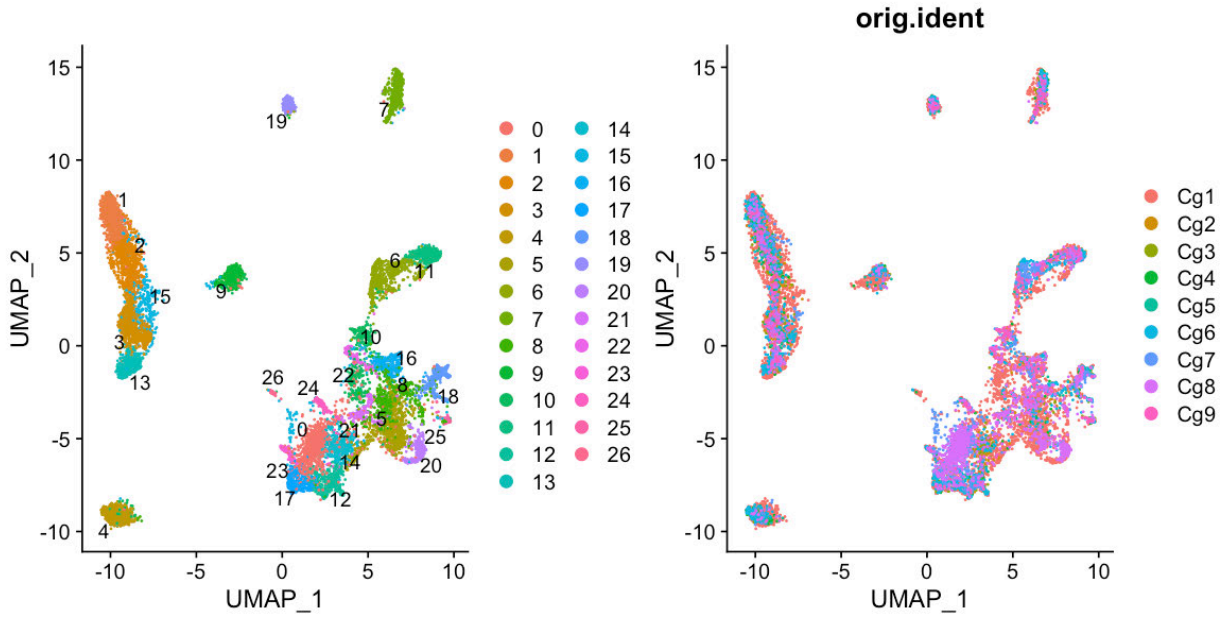


Figure 19. After integration cells from different batches cluster together.

UMAP plot of Cg1-Cg9 integration coloured by clusters (left) and by batch of origin (right) shows that cell do not cluster by batch.

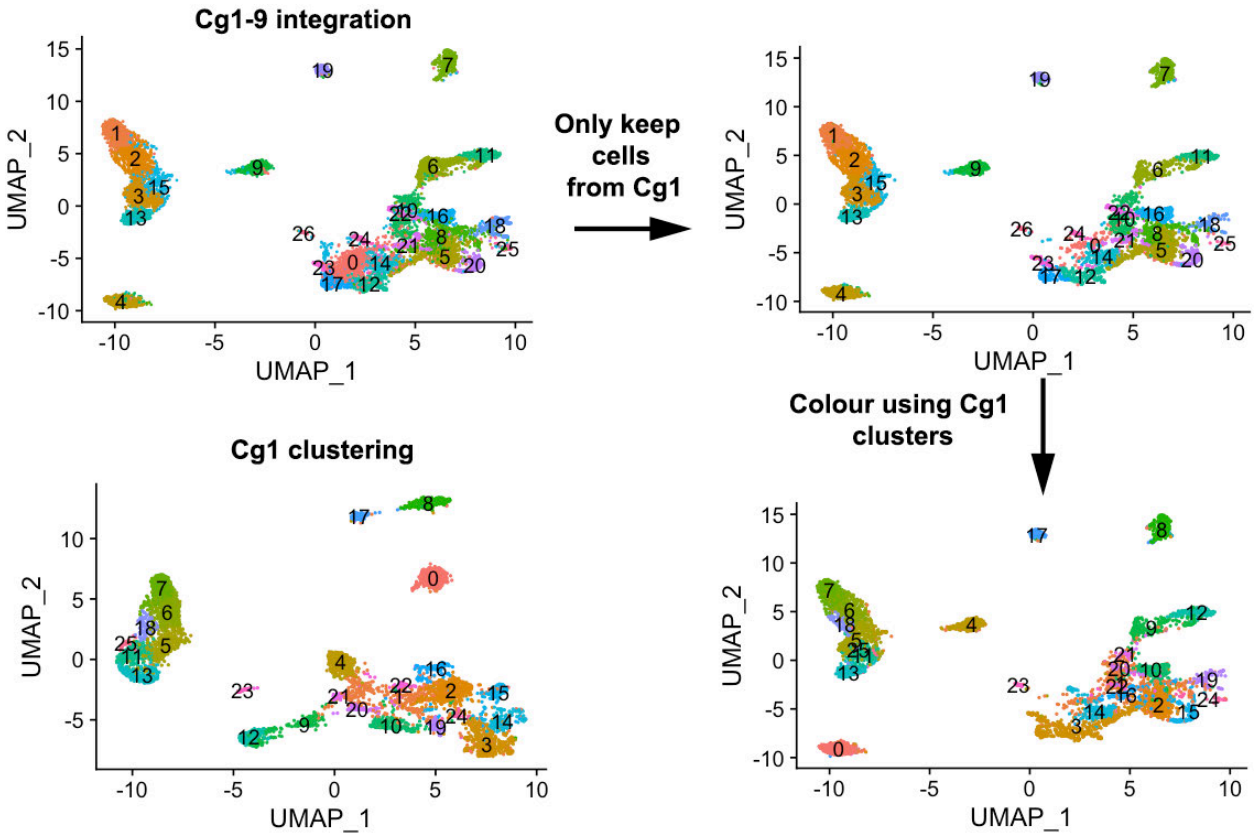


Figure 20. UMAP plot showing similar clustering between integrated samples Cg1-9 and Cg1 (the best quality sample).

Top left: UMAP plot of samples Cg1-9 integrated using Seurat v3 integration pipeline; Top right: same UMAP only showing cells from Cg1; Bottom left: UMAP plot of Cg1 cells; Bottom right: same UMAP as B but colour coded with clusters from Cg1 showing that clusters from integration of all batches resemble those obtained with the best quality batch only.

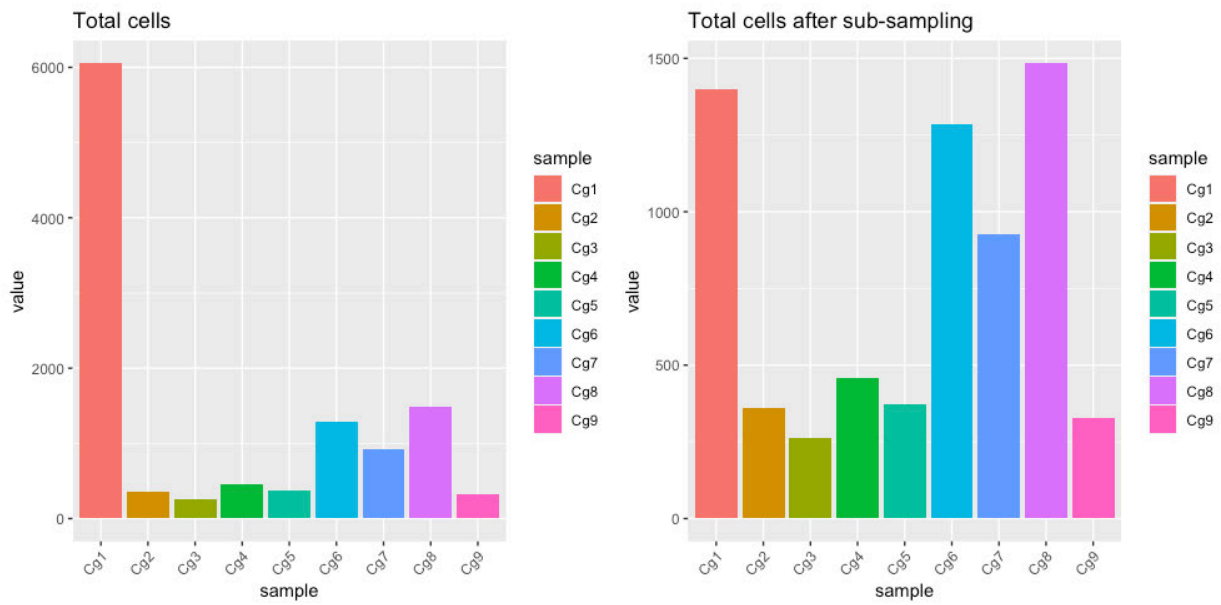


Figure 21. Cg1 has more cells than any other batch, could this drive the clustering?

Number of cells per sample used for integration and number of cells per sample used after down sampling of Cg1 to try and understand if the higher cell number of Cg1 was driving the clustering.

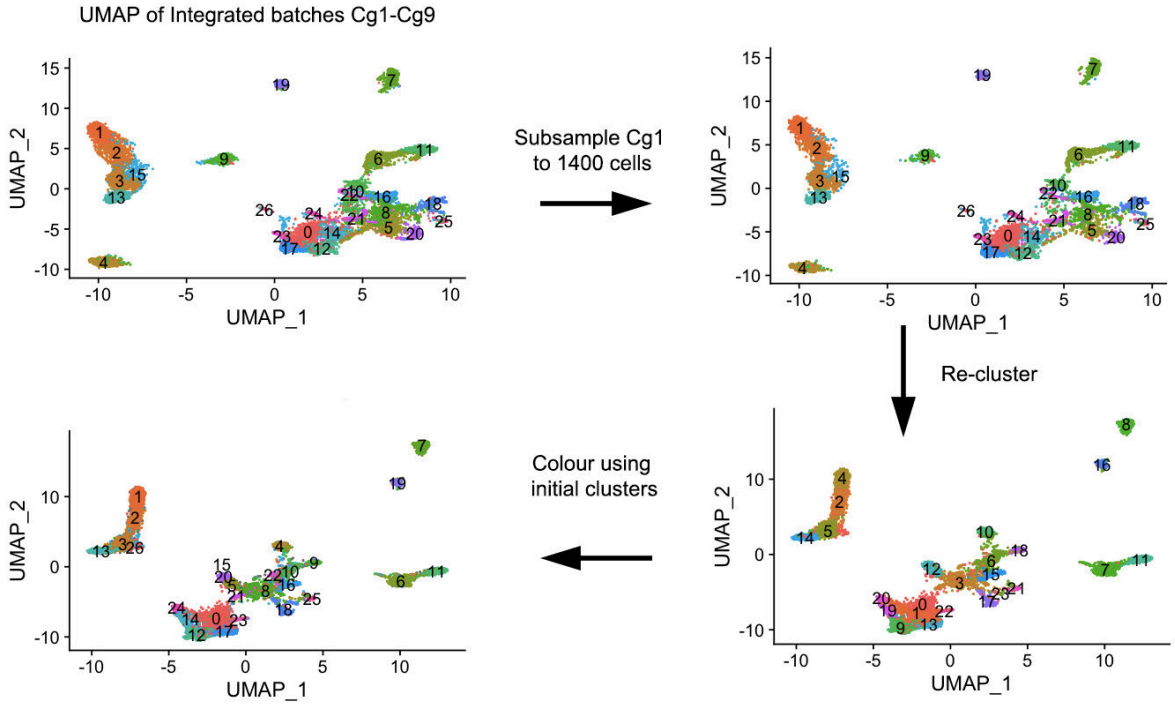


Figure 22. The higher number of cells in sample Cg1 is not driving the clustering.

In fact, if I subset Cg1 to contain as many cells as the other samples I still obtain very similar clustering. Top left: UMAP plot of samples Cg1-9 integrated using Seurat v3 integration pipeline, Top right: same UMAP as top left randomly sampling only 1400 cells from Cg1, bottom right: re-clustering of cells from top right, bottom left: same UMAP as bottom right

but coloured to match original clustering from top right and left. This process shows that cell clustering is not driven by the most numerous batch (Cg1).

3.3.4 Re-sequencing of Cg1

Since we have shown that a) samples collected show different qualities b) cells from different samples cluster together after integration and c) clustering of lower quality samples matches clustering obtained using only the best quality sample (Cg1) I will from here onward only use cells belonging to sample Cg1. This is because this sample has the highest number of cells, genes and UMI and will be the most informative and its clustering is backed up by our other lower quality repeats. I also decided to re-sequence Cg1 in order to obtain the best possible number of cells, genes and reads per cells. After resequencing, sample Cg1 contains ~8000 cells with approximately 1000 median genes per cells and 2000 median UMI per cells and in total we recovered 27,509 genes. With these new results I also obtained a higher number of clusters (32 instead of ~20) the identity of which will be discussed in the next chapter (figure 23).

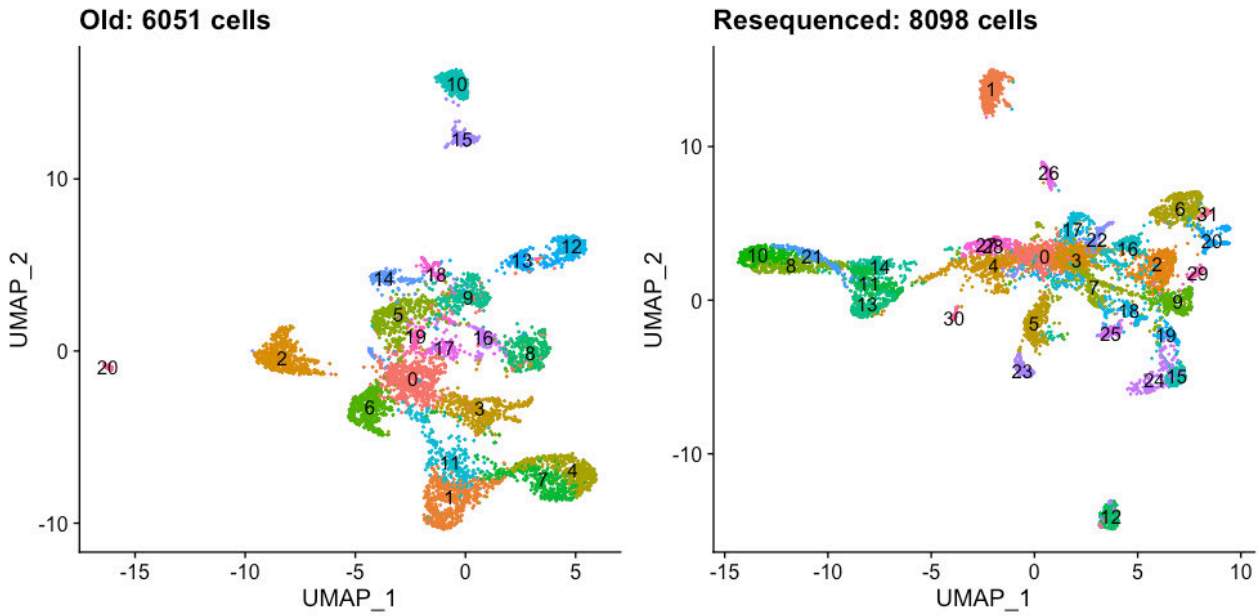


Figure 23. Resequencing of Cg1 leads to higher cell and cluster number.

UMAP of sample Cg1 before and after re-sequencing shows higher cell number as well as a higher cluster number.

3.5 Conclusions

3.5.1 Summary of results

In this chapter I presented the highly variable results I obtained in the 9 different scRNAseq runs performed on the oyster larva. Firstly, I showed how my first runs presented very low mapping percentages and how re-annotation of the genome using scRNA-seq reads helped to improve this percentages drastically as well as helping to recover more cells with more genes. Then, using different approaches (such as counting overall UMI numbers per sample and checking the % of mitochondrial reads in cells) I showed that different samples presented differences in quality and explained what could have caused them. I also demonstrated how this difference in quality did not substantially influence clustering when

samples were correctly integrated using Seurat. Moreover, I showed how the clustering of my highest quality sample (Cg1) was backed up by all other repeats. Since I could show that the result was reproducible I decided to re-sequence Cg1 in greater depth and to use this sample for downstream analysis as it would be a waste to try and compare low quality samples. In fact, a reduced depth in gene coverage per cell could easily cause artefacts during the comparison.

3.5.2 Reannotation of the oyster genome

In the first paragraphs of this chapter I showed how the mapping of the scRNA seq reads obtained for my first scs runs appeared to be fairly low. We also observed that many reads appear to map to intergenic or intronic regions which could indicate that the 3' UTR of genes in the genomes was not annotated. For this reason, we decided to try and re-annotate the genome using are newly produced 3' biased scRNAseq data and obtained a much higher mapping as well as an overall increase in cells, genes and UMIs per cells. This improvement was so substantial that it is in itself a notable result and proves the importance of a high-quality genome (and genome annotation) to perform successful scRNA seq experiments.

3.5.3 Technical challenges of scRNA seq experiments

As described in depth in this chapter many different factors play an important role in the success of a scRNA seq capture. Starting with the dissociation step it is important to try out different protocols, familiarise with the manual trituration (if needed) in order to obtain a

solution of single cells in a fast, gentle and effective way. In fact, a protocol that is too long or too harsh can cause an increase of dying cells for which many reads recovered would be mitochondrial genes which will have to be discarded. As shown in this chapter even the user can make a great difference in the final result obtained.

Then the cells need to be eluted in a buffer in which the cells can survive until the solution is loaded onto the multifluidic chip but that is also compatible with the downstream process. In the case of marine animals this process can be very problematic since both salt concentration and the presence of other additives (such as EDTA) can easily cause problems during reverse transcription. The many challenges I faced during this step convinced me to try different salt concentration during the preparation of the flatworm larva samples which will be discussed in chapter 4.

All of these aspects need to be considered when performing scRNA-seq experiments and have to be adjusted for each animal. In my case the several issues encountered during preparation of the different repeats was greatly educative albeit slightly depressing.

3.5.4 Computational challenges of quality control of scRNA seq samples

Once scRNA samples are obtained and sequenced one has to retrospectively ascertain the quality of the cells and decide how to:

1) correctly filter out empty droplets
2) filter out low quality/dying cells
3) get rid of multiplets

Empty droplets can sometimes be mistakenly confused with cells by CellRanger software when the solution that cells are in contains a lot of background RNA. This can be caused by leaky cells and in the case of marine invertebrates appear to be often a pretty substantial problem (personal observation) possibly due to adjustment of the salt concentration of the elution buffer (and subsequent cell osmoregulation). Usually filtering of other parameters such as minimum gene per cell content or mitochondrial percentage can help getting rid of these artefacts as well as of dying cells.

Multiplets can also be present when more than one cell gets captured in the same droplet. These “cells” should present a considerably higher gene and UMI content and can also be easily removed by filtering out higher outliers. In my case I didn’t detect any cluster which presented abnormal gene or UMI content and so I do not think multiplets caused particular problems (results not shown).

Finally, for some samples (such as in this case) integration might be necessary to get rid of batch effects. In general integration should be avoidable when possible, and for this reason we will only use Cg1 for the comparison between species, however in our case it proved to be a powerful tool to overcome differences in quality.

4 Identification of cell types in the Oyster larva

In the previous chapter I discussed quality control of the different single cell experiments performed on the larva of the oyster. I have shown how all repeats, regardless of their differences in quality, support the clustering obtained with my best sample (Cg1). These steps were important to prove that the same cell types could be recovered in different experiments but also showed clearly that sample Cg1 had the highest quality of all repeats. On this basis I decided to sequence this specific sample in greater depth and then to use Cg1 for downstream analysis including final clustering, cell type identity prediction and for comparison across species. There would be little point carrying out comparisons between species using low quality samples.

In this chapter I will discuss briefly how cell clustering was carried out and then show how I have tried to assign cell type identities to the different clusters by looking at cluster specific markers and by comparing those with already described markers of different cell types that I have found in the literature. I then proceed to validate the expression of cluster specific markers using chromogenic *in situ* hybridisation (ISH) or *in situ* Hybridisation Chain Reaction HCR. Following this approach, I was able to assign possible cell identities to most clusters, which is fundamental for carrying out meaningful comparisons between the two larvae.

4.1 Clustering of Cg1 sample

After assessing the overall quality of the extra sequencing date from sample Cg1, I proceeded with single cell sequencing analysis in Seurat following the same steps described in the previous chapter with the following cut offs:

1) Only cells with at least 200 genes expressed were kept.

2) Only cells with no more than 10% of mitochondrial gene content were kept.

I then used the top 2000 most variable genes to perform the PC analysis and then used the top 50 PCs for the clustering. When using a resolution of 2, I obtained 32 clusters (see fig 24), this seemed appropriate since with a resolution of 1 some small clusters that were detached would be grouped together to form larger clusters (i.e. cluster 29).

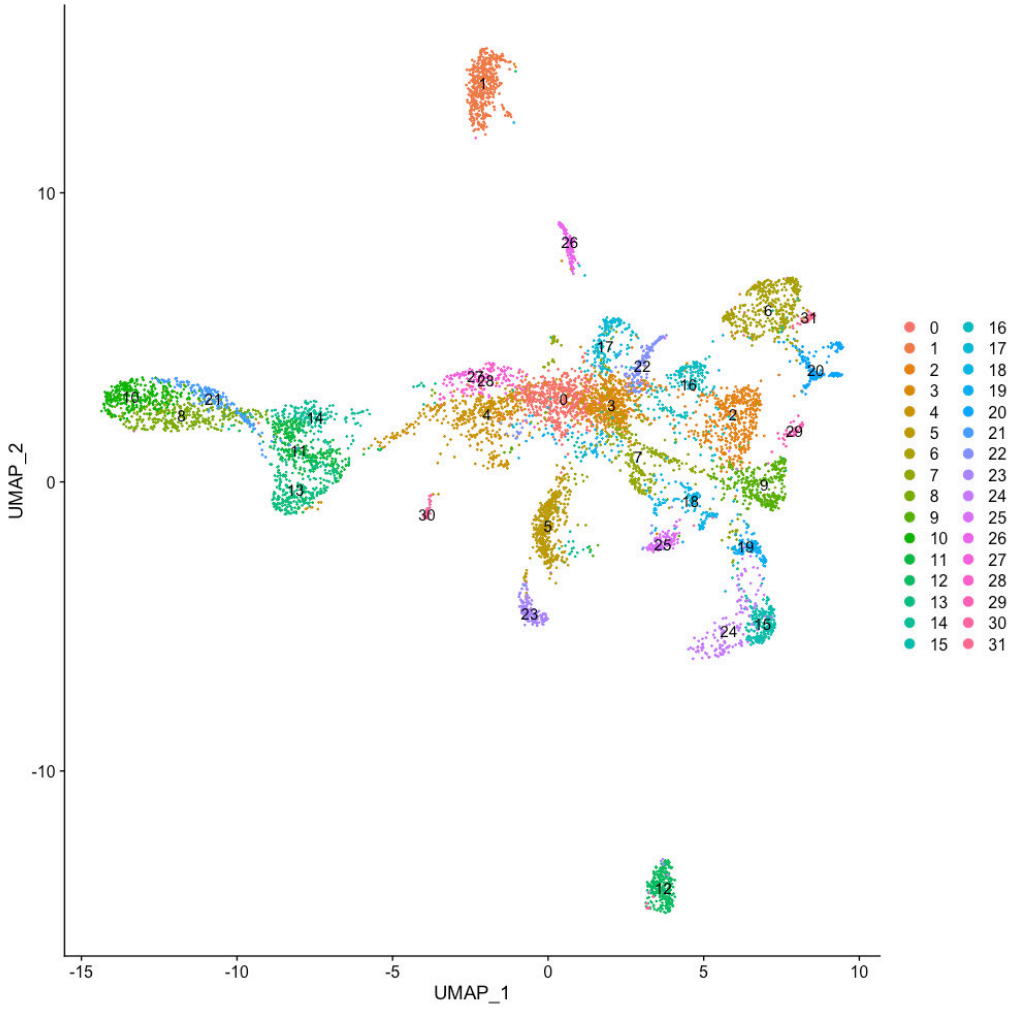


Figure 24. UMAP of Cg1 shows 32 different clusters.

After resequencing, the Cg1 sample contains ~8000 cells with approximately 1000 median genes per cells and 2000 median UMI per cells and in total we recovered 27,509 genes.

4.2 Analysis of different cell types

In the next sections I will describe the work done to assign cell type identities to the different clusters by characterising cluster specific markers and by searching the literature for already described markers of different cell types.

4.2.1 Muscle clusters

The first clusters of cells I am going to investigate are myocytes - the cells that make up muscles. Most animal muscle cells contain thick and thin filaments. Thick filaments are composed of myosin which is itself composed of 3 pairs of proteins: two heavy chains, two essential light chains and two regulatory light chains. The thin filaments are made of actin and can be associated with proteins such as tropomyosin and troponin. The specific combination of thick and thin filaments classically gives rise to two morphologically distinct muscle type: smooth and striated muscles. In molluscs, however, the distinction between the two is often difficult to discern. Moreover, molluscan smooth muscles (and possibly some striated muscles too) can exhibit a mechanical state called “catch” that allows them to maintain passive tension for long period of times with minimal energy requirements. This mechanism is used by bivalves to keep their shell closed for long periods of time and it is also been found in other invertebrates (such as insects, crayfish, nematodes, brachiopods and others) (Hooper and Thuma, 2005). Many proteins (including paramyosin, twitchin, catchin and calponin) have been suggested to play a role in the catch properties of muscles, however, no single gene marker has been yet found that A) seems to be present only in catch muscles and/or B) is necessary for catch properties.

For these reasons, to identify possible myocyte clusters in our single cell data I decided to look at a wide range of muscle related genes which highlighted several possible myocyte clusters: 12, 15, 19, 23 and 24 (see fig. 25). For these clusters I performed chromogenic *in situ* hybridisations on individual cluster markers to localize cells belonging to each of these clusters in the mollusc trochophore larva (see fig. 26). First, I looked at the expression of *Cg-troponin-T* which has a widespread expression in all the putative muscle clusters (see fig 25). *Cg-troponin-T* is expressed in two symmetrical triangular patches, one on either side of the animal. This expression is similar to what has been detected with immunohistochemistry against F-actin in other bivalve larvae (Wurzinger-Mayer et al, 2014; Kurita et al, 2016).
I then looked at more specific cluster markers. For cluster 12 I chose the gene *Cg-mab21-like-2*. This stains an area near the shell gland in the dorsal part of the trochophore (see fig. 26, staining is indicated in red in the schematic). Due to the position of the staining, I believe that cluster 12 may contain cells belonging to the anterior adductor muscle. This muscle is a bivalve innovation that controls the opening and closing of the shell plates (Kurita et al, 2016). The gene expression in this cluster is quite interesting as it is the only cluster to show expression of both paramyosin and calponin which is a combination that is typical of catch muscles (see fig. 25).

The marker I chose for cluster 23 is *Cg-FMRF-receptor* and it stains very few cells (possibly only two) in the apical organ near the apical tuft (see fig. 26, staining is indicated in blue in the schematic). I am not yet sure what these cells are as they seem to express a combination of muscle markers as well as a unique combination of transcription factors (these will be discussed in detail later).

Clusters 15, 19 and 24 share many markers, for this reason I chose a general marker for all three (*Cg-myosin-9-like*) which stains two almost symmetrical patches on the anterior part of the trochophore that extend towards the mouth (see fig. 26, staining is indicated in green in the schematic). I hypothesize that these clusters represent the velum retractors as the expression resembles that of previously published immunohistochemistry (Dyachuk et al, 2012).

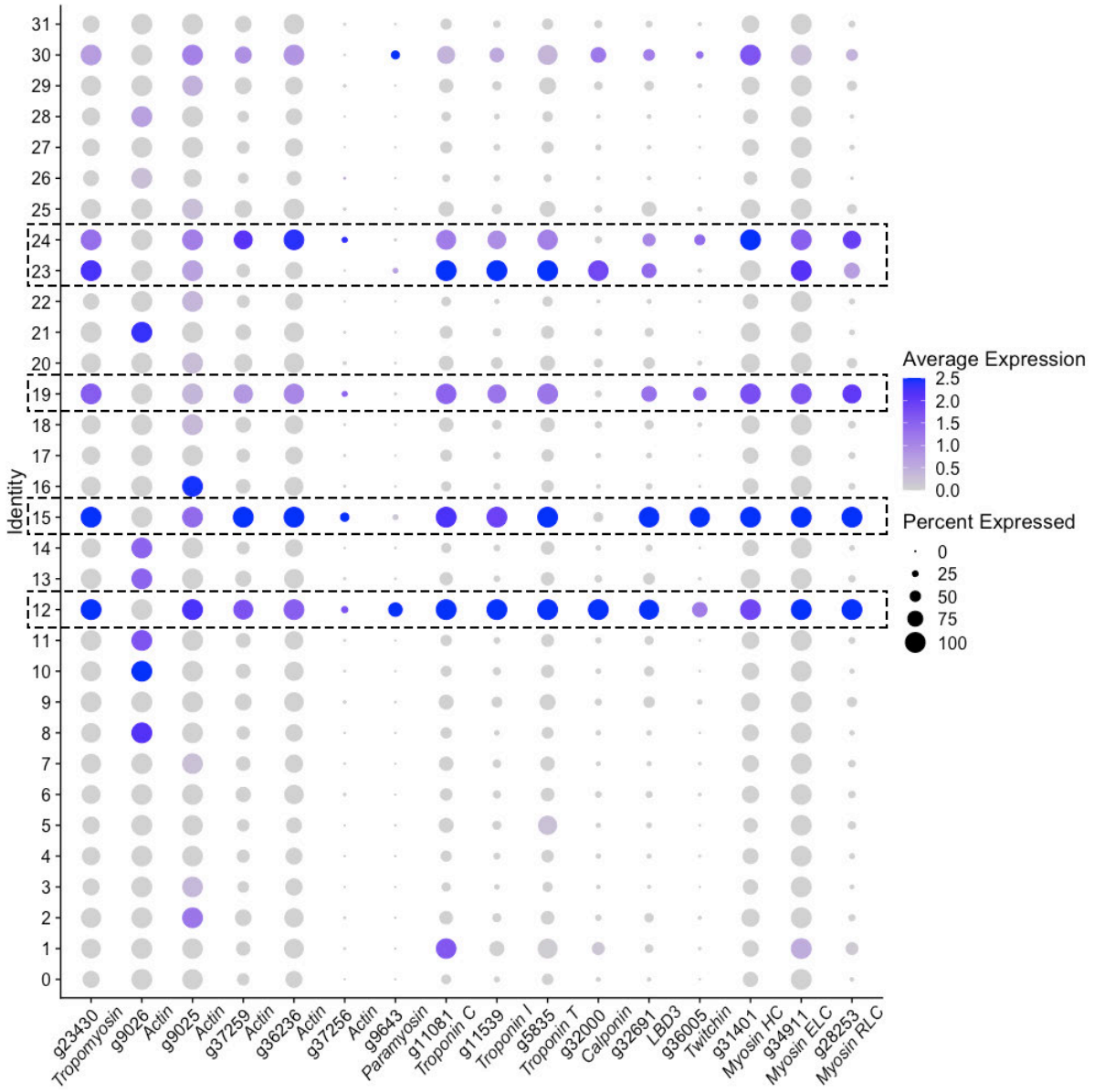


Figure 25. Clusters 12, 15, 19, 23 and 24 are likely myocyte clusters as they express several muscle markers.

Dot-plot showing gene expression of several myocyte markers (x axis) in different clusters (y axis) highlighting the presence of several myocyte clysters, specifically: clusters 12, 15, 19, 23 and 24. Size of the dot indicates the percentage of cells from the cluster expressing the gene (small means few cells are expressing the gene), colour indicates how expressed the gene is (blue highly expressed, grey lowly expressed).

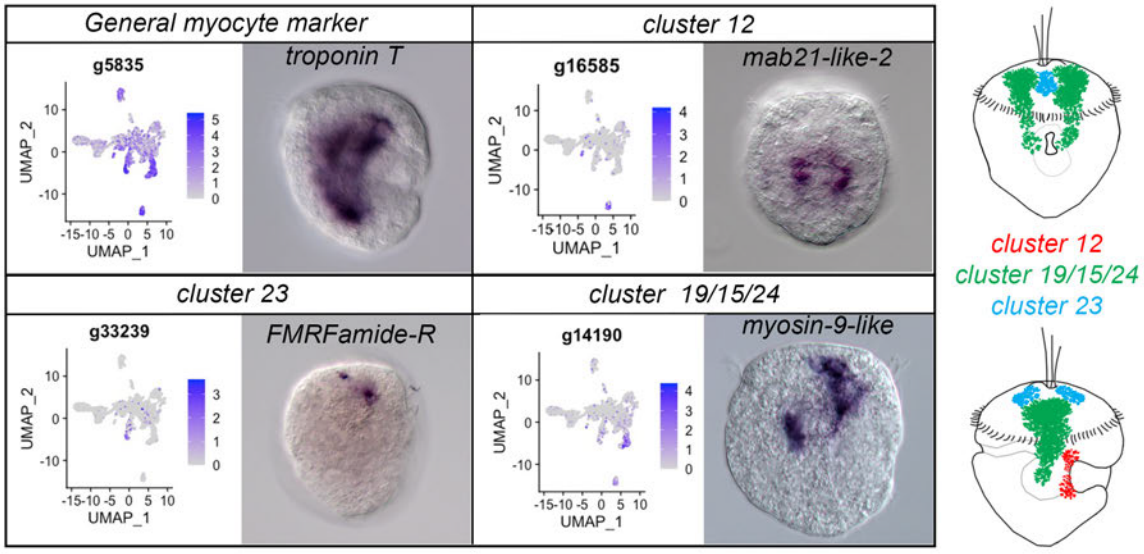


Figure 26. Muscle clusters form two triangular patches on either side of the larva.

Small UMAP shows in blue the expression of the gene, next to ISH of the same gene. Diagram on the right summarize the expression of the different markers for the clusters. On top frontal view of the larva, on the bottom lateral view (mouth is on the left, shell gland on the right).

4.2.2 Shell gland clusters

One of the most well-known features of molluscs is their shell and for this reason this the second group of cells I decided to look for are the shell gland cells. Adult shells vary drastically in composition across species: they can be made of different polymorphs of calcium carbonate - aragonite or calcite - or both; have one or several layers; and different layers can have different compositions. Larval shells, on the other hand, are all made of aragonite and share similar ultrastructure and hence seem to be more conserved than adult shells (Weiss et al, 2002). During bivalve mollusc development, the first shell, called prodissoconch I, starts to be secreted from the shell field at the trochophore stage and it will go on to form two D shaped shells. Right after this, the prodissoconch II starts to be secreted from mantle tissue of the veliger larva. Once the larva settles and metamorphoses,

the juvenile will start secreting the adult shells or dissochonch (Zhao et al, 2018). To identify the shell gland (or shell field) cluster, I looked for a selection of previously described markers, and found these genes are expressed in three clusters: cluster 6, cluster 20 and cluster 31 (see fig. 27) (Miyamoto, 2005; Kong et al, 2009; Huan et al, 2013). I then selected some clear cluster markers and performed *in situ* hybridisation to localize the three clusters in the larva.

For cluster 6, I picked three different markers, *Cgi-tyrosinase, Cgi-mucin and Cgi-soxE*. All three genes are expressed in the dorsal part of the larva, inside an area where the shell gland has been previously described to be. I believe that these cells could be responsible for the secretion of the prodissoconch I since their shape and location is similar to that of the first D shaped shells (see fig. 28).

For cluster 20, the genes I investigated are *Cgi-prisilkin, Cgi-dopamin-β-hydroxilase* and *Cgi-engrailed*. All three genes are expressed in a group of cells at the border of the shell gland. I speculate that these cells could be the predecessors of cells that will be involved in the secretion of the prodissoconch II at the veliger stage, as they are localised at the border of cluster 6 cells (see fig. 28).

Finally, for cluster 31 I chose two gene markers: Cgi-BMSP and Cgi-g7838, the latter didn't blast to any previously annotated gene but only to uncharacterised proteins. These two genes are expressed in a patch on either side of the shell gland, in an area that is near the anterior adductor muscles previously described (see fig. 28).

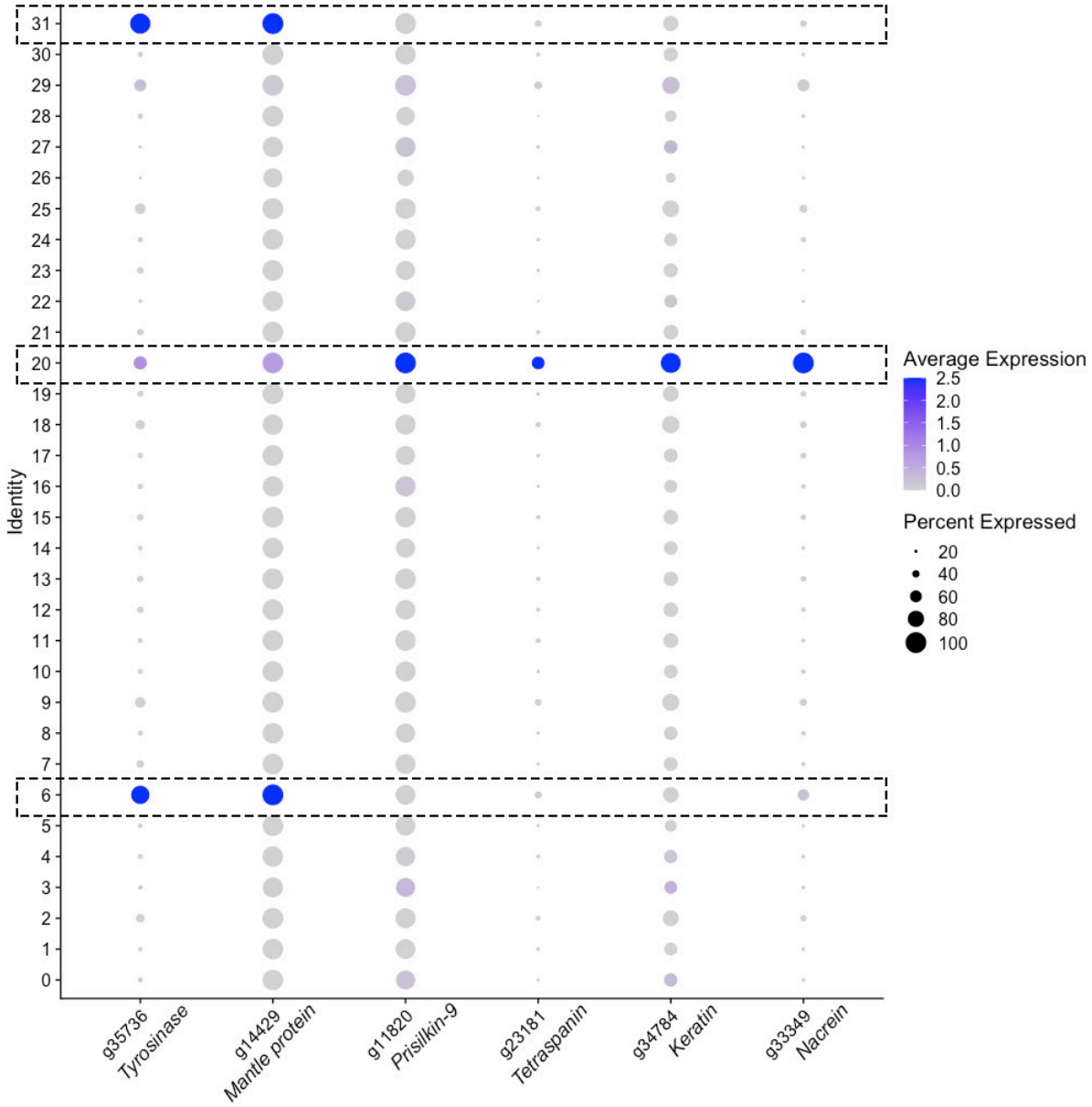


Figure 27. Expression of several previously described shell gland markers is concentrated in cluster 6, 20 and 31.

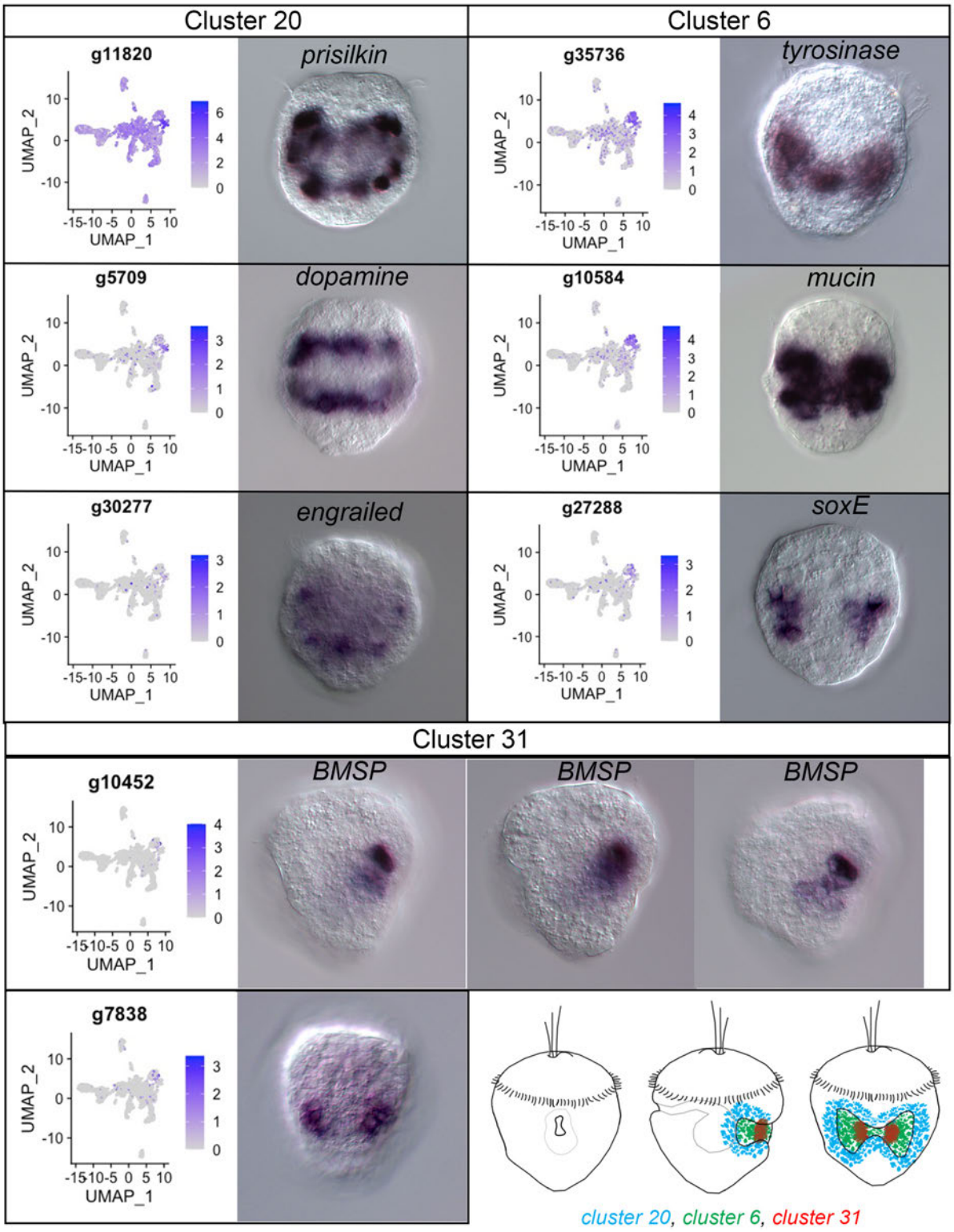


Figure 28. Expression of several previously described shell gland markers is concentrated in cluster 6, 20 and 31.

Small UMAP shows in blue the expression of the gene, next to ISH of the same gene, for BMP several focal planes are shown. Diagram on the bottom right summarize the expression of the different markers for the clusters (on the left frontal view of the larva, lateral view in the middle with mouth on the left, on the right dorsal view of the larva).

4.2.3 Ciliary band clusters

Ciliary bands are one of the main diagnostic features of spiralian larvae and, at least for trochophore larvae, authors agree that primary ciliary bands (the prototroch) are homologous. For this reason, I was keen to localise the ciliary band clusters and to use their expression signatures to compare with that of the Muller's larva ciliary bands (see chapter 6).

As I did for the other cell types, I firstly scanned the literature for ciliary band markers (Kakoi et al, 2008; Wu et al, 2020; Wang et al, 2020). As one can see in figure 29, most of the genes are expressed in three clusters: cluster 2, cluster 22 and cluster 25. ISH confirms that cells belonging to these clusters are localised to the prototroch as well as in a ring of few cells at the bottom of the embryo, which I believe may be a telotroch (see fig. 30). To try and discern which cluster is localised where I decided to use HCR on three marker clusters: one general ciliary band gene (g31376), one gene specific to cluster 25 (g11844) and one specific to cluster 2 (g3234). As one can see in figure 31 cluster 2 appears to contain cells of both the telotroch and prototroch whilst clusters 22 and 25 are made up only of prototrochal cells. Two of the ciliary band markers I used in the dotplot shown in figure 29, lophotrochin and trochin, were noted in a recent publication that looked at the expression of spiralian specific genes and found that these two genes were expressed in the ciliary bands of many spiralian larvae (Wu et al, 2020). In that work the authors first screened the genome of the oyster *C. gigas* in search of spiralian specific genes and then used those to look for orthologues in their species of interest. I decided to take the whole list of spiralian specific genes found in

oysters and to look for their expression in our SCS dataset. Figure 32 shows that 14 more spiralian specific genes (16 in total together with lophotrochin and trochin) out of the original 37 found in the study are expressed in the ciliary band of the oyster trochophore larva.

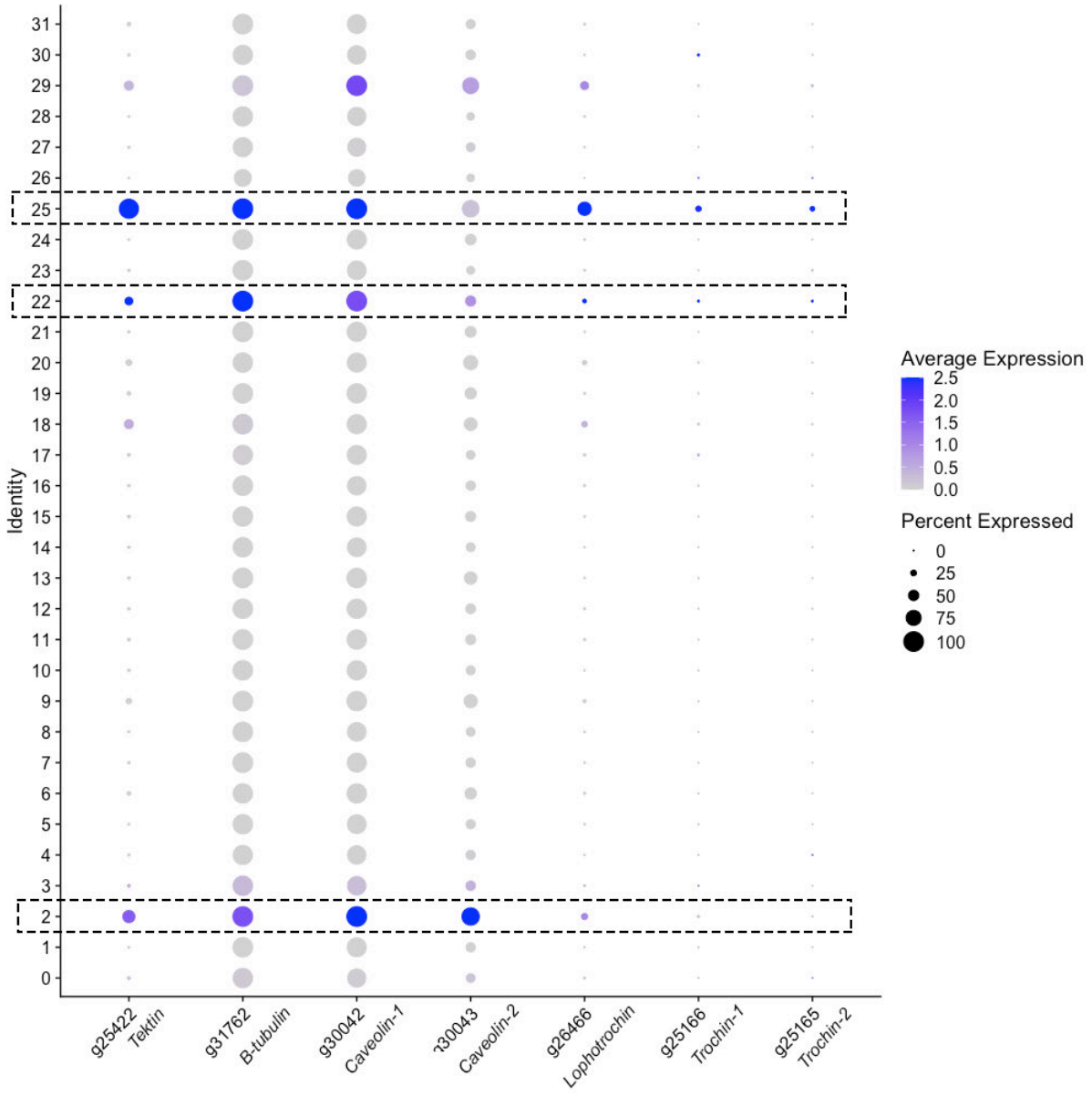


Figure 29. Shows the expression of several previously described ciliary band markers is concentrated in cluster 2, 22 and 25.

The last three genes, lophotrochin and trochin-1 and trochin-2 are from a recent publication that found these genes were expressed in the ciliary bands of many spiralian larvae (Wu et al, 2020)

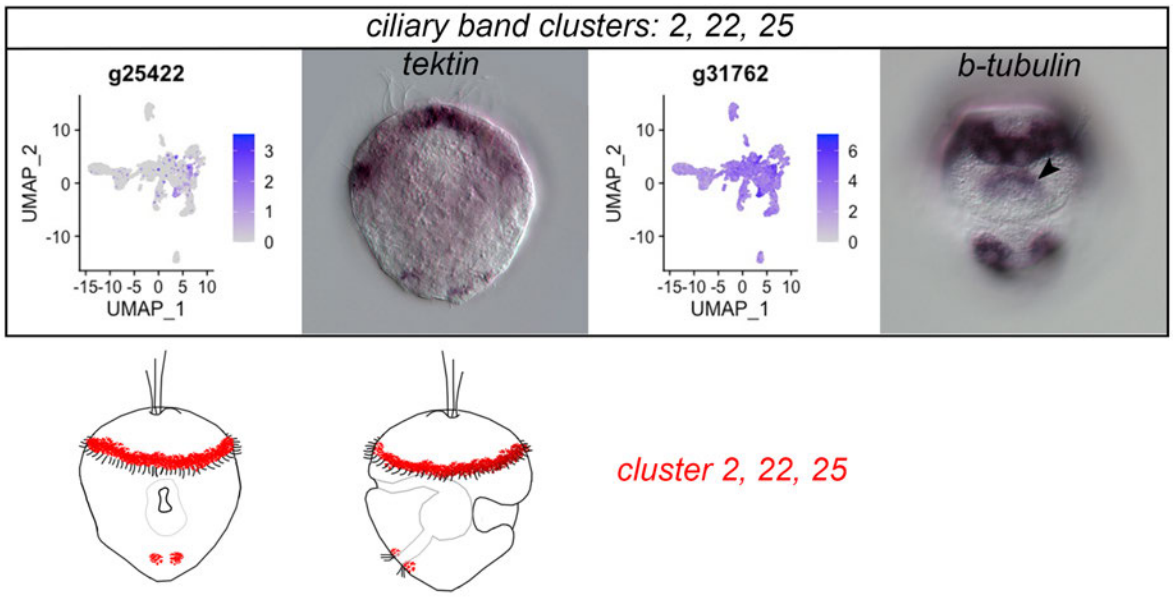


Figure 30. ISH of ciliary band markers cgi-tektin and cgi-β-tubulin show expression around the ciliary band and few cells near the anus.

Small UMAP shows in blue the expression of the gene, next to ISH of the same gene. Diagram on the bottom summarize the expression of the different markers for the clusters (on the left frontal view of the larva on the right lateral view of the larva, mouth on the left).

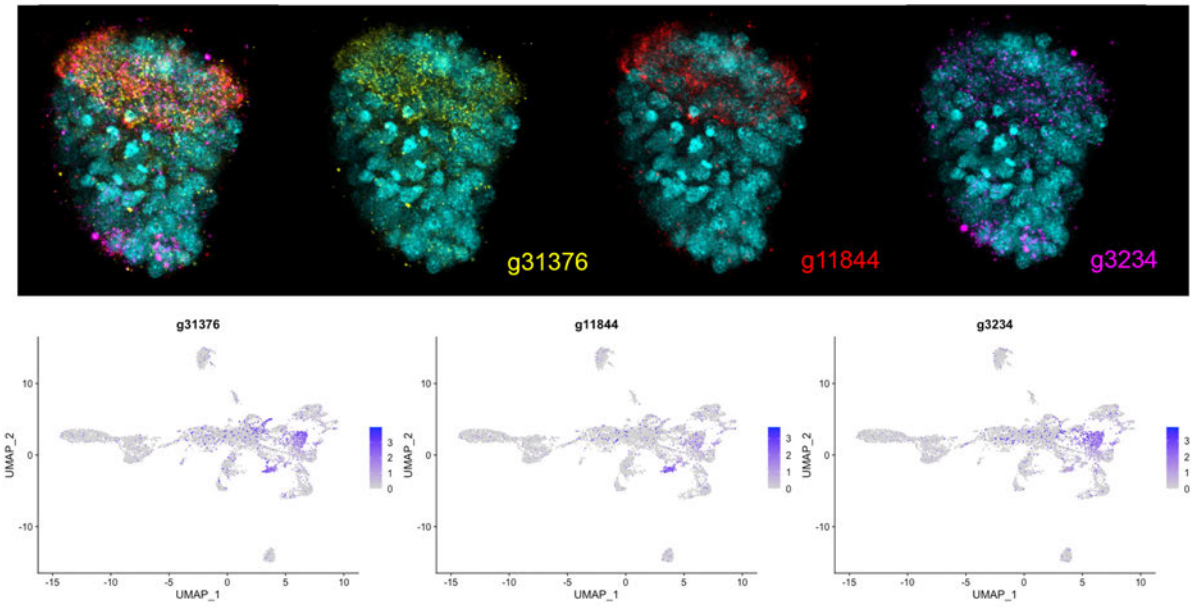


Figure 31. Cluster 2 contains mostly cells of the telotroch whilst cluster 22 and 25 are prototrochal cells.

HCR of a general ciliary band gene (g31376 in yellow), one gene specific to cluster 25 (g11844 in red) and one specific to cluster 2 (g3234 in pink). Small UMAP below shows in blue the expression of the three genes.

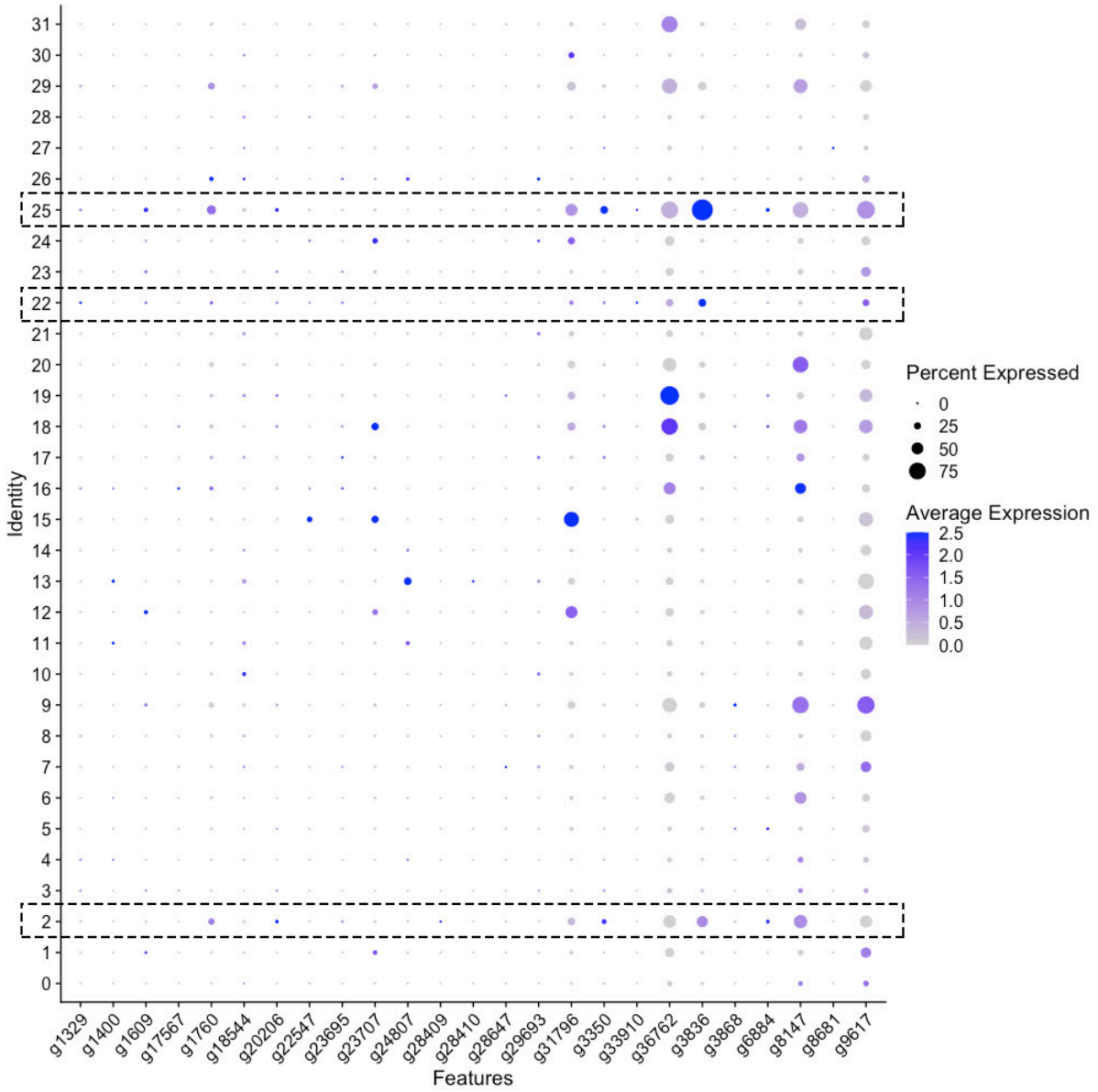


Figure 32. Many spiralian specific genes are expressed in the ciliary bands of the oyster larva.

Dotplot of spiralian specific genes from the study by Wu et al (2020) shows that 14 out of 37 such genes are expressed in ciliary band clusters.

4.2.4 Neuronal clusters

The larval nervous system of molluscs has been previously studied with the use of immunohistochemistry. These studies have shown that the first neurons appear right before the trochophore stage and that the larval nervous system typically initially consists of an

apical organ containing serotonergic neurons as well as a pair of dorsal and a pair of ventral peripheral neurons (Croll and Dickinson, 2004; Yurchenko et al, 2019).

Since neuropeptides are usually highly expressed in neurons, to search for neuronal clusters in our data I decided to look specifically at neuropeptide expression. In collaboration with Luis Yañez-Guerra (Jekely lab, Exeter) we firstly identified all the neuropeptides in our newly annotated oyster genome and I then searched for the expression of those neuropeptides in our SCS data. Fig. 33 shows the expression of the different neuropeptides in our single cell clusters and highlights two clusters: cluster 18 and 29.
ISH using suitable cluster markers show that cells belonging to cluster 18 are concentrated in the apical organ as well as in a patch of few dorsal cells above the shell gland (see fig. 34). These cells could be the anterior and dorsal neurons previously described by Yurchenko and colleagues (2018). Markers of cluster 29 are expressed instead in two cells on the ventral part of the larva and I think these could either be the ventral or the posterior neurons (Yurchenko et al, 2018).

To try and distinguish between different neuronal types at higher resolution I decided to use HCR with a combination of probes for three different neuropeptides, specifically GNQQNxp, FMRF and Myomodulin. Probing for a combination of three different neuropeptides allowed me to identify all neuronal types described using immunohistochemistry by Yurchenko (2018) and to link them to the different SCS clusters (see figure 33). As shown in figure 35, cells of the apical organ likely belong to cluster 18, posterior neurons belong to cluster 29 whilst ventral and dorsal neurons, marked with FMRFamide, are a sub-group of cells that is part of cluster 16. However, FMRFamide positive cells sit at the very edge of cluster 16 and

are slightly detached from other cells of cluster 16 so I believe they only get grouped together accidentally and probably would separate out at a higher resolution. In general, it appears that some neuropeptides are very specific to a few cells; another example for this is myomodulin which is expressed in only one cell of the apical organ. Although it would be tempting to increase the resolution further to try and divide out these different neuronal population it is worth noting that this tends to create further subdivision of larger clusters (such as haemocytes) without much apparent biological meaning. I tried subsetting the neurons separately but because of the very small number of cells the results were hard to interpret as I would, for example, find several sub-clusters with no specific markers.

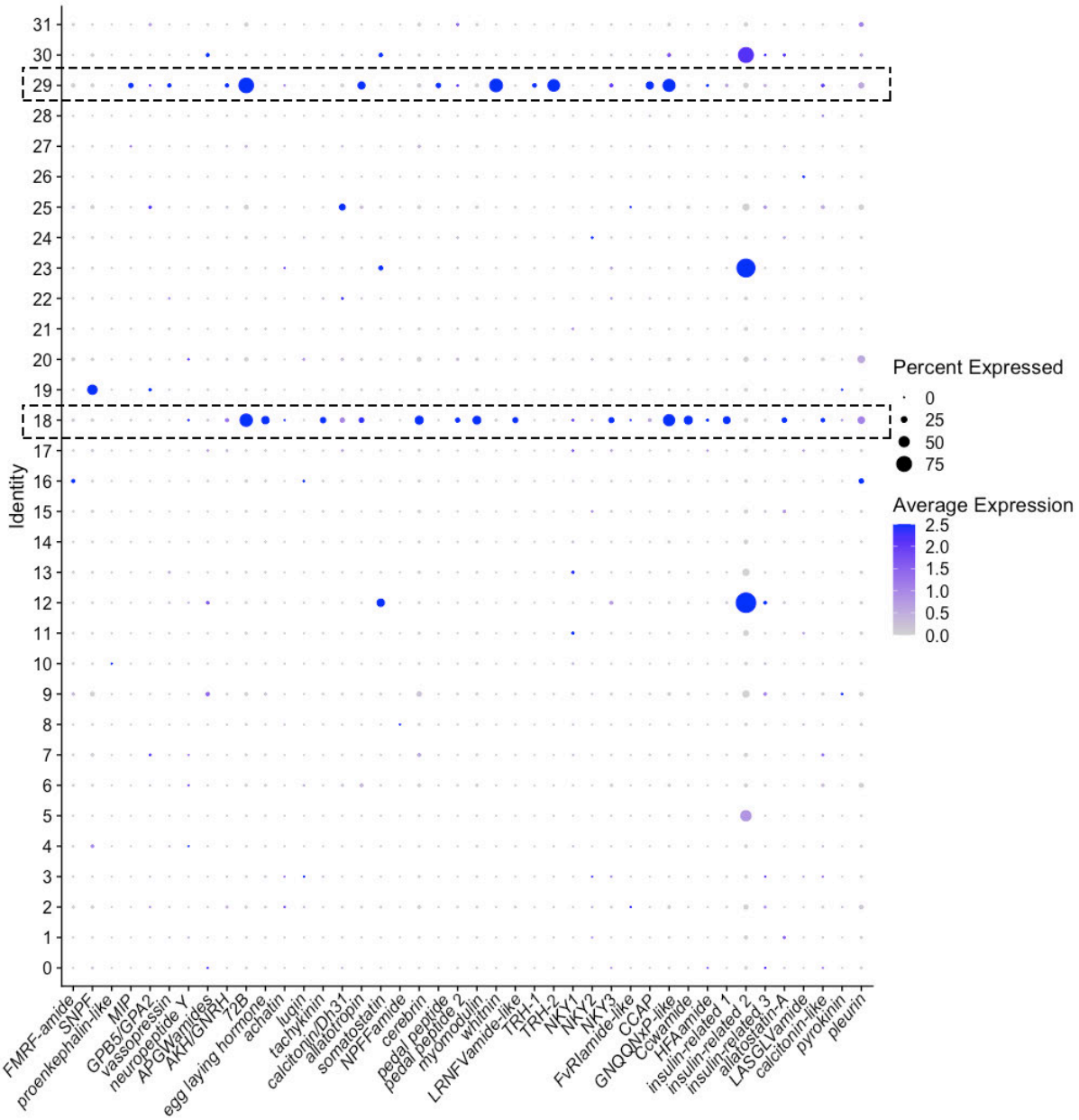


Figure 33. The majority of neuropeptides are expressed in cluster 18 and 29 which are likely neuronal cell types.

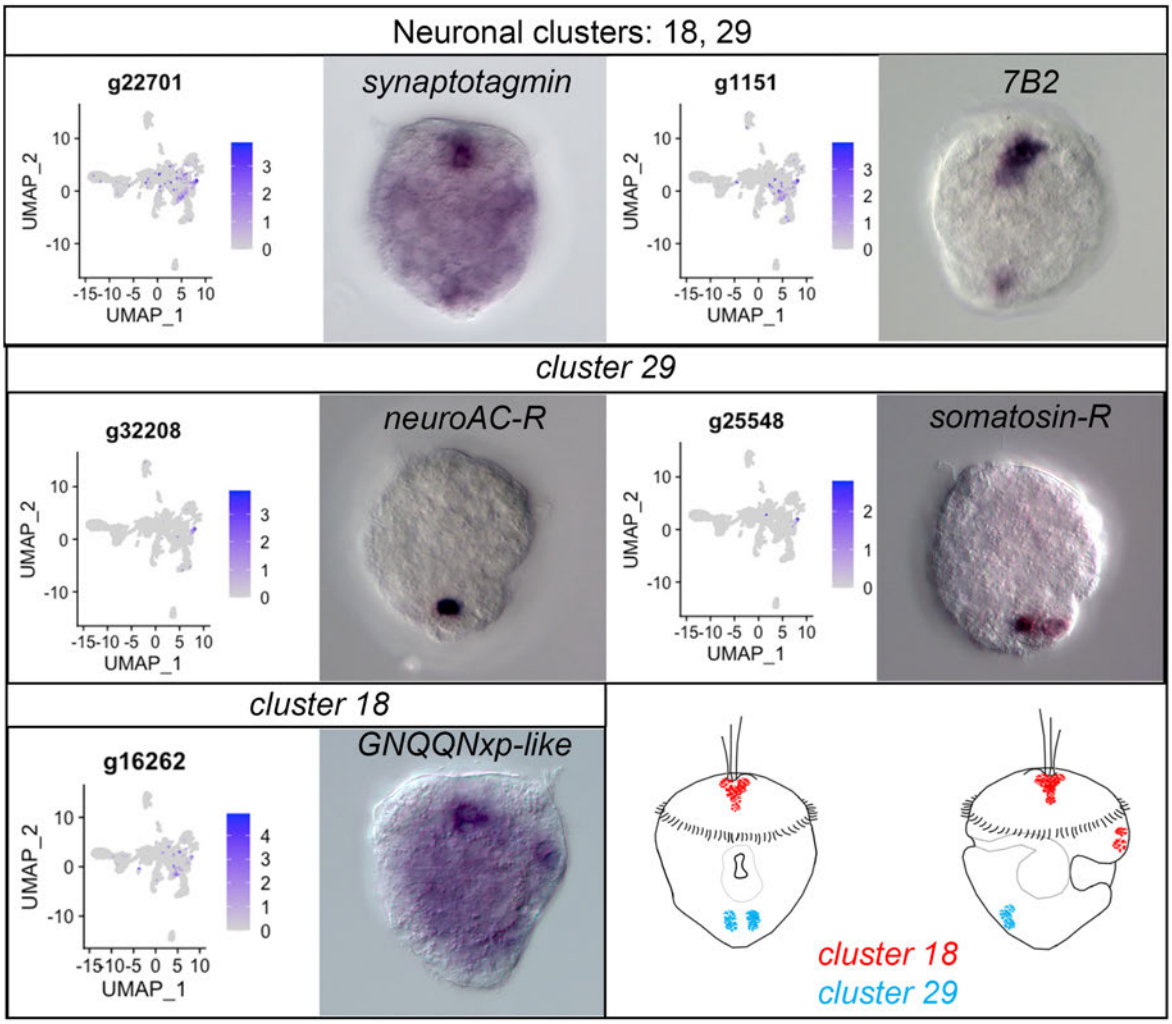


Figure 34. Neuronal clusters show expression in the apical organ (cluster 18) dorsally above the shell gland (cluster 18) and ventrally below the mouth (cluster 29).

Small UMAPs show in blue the expression of the gene, next to ISH of the same gene. Diagram on the bottom summarize the expression of the different markers for the clusters (on the left frontal view of the larva on the right lateral view of the larva, mouth on the left).

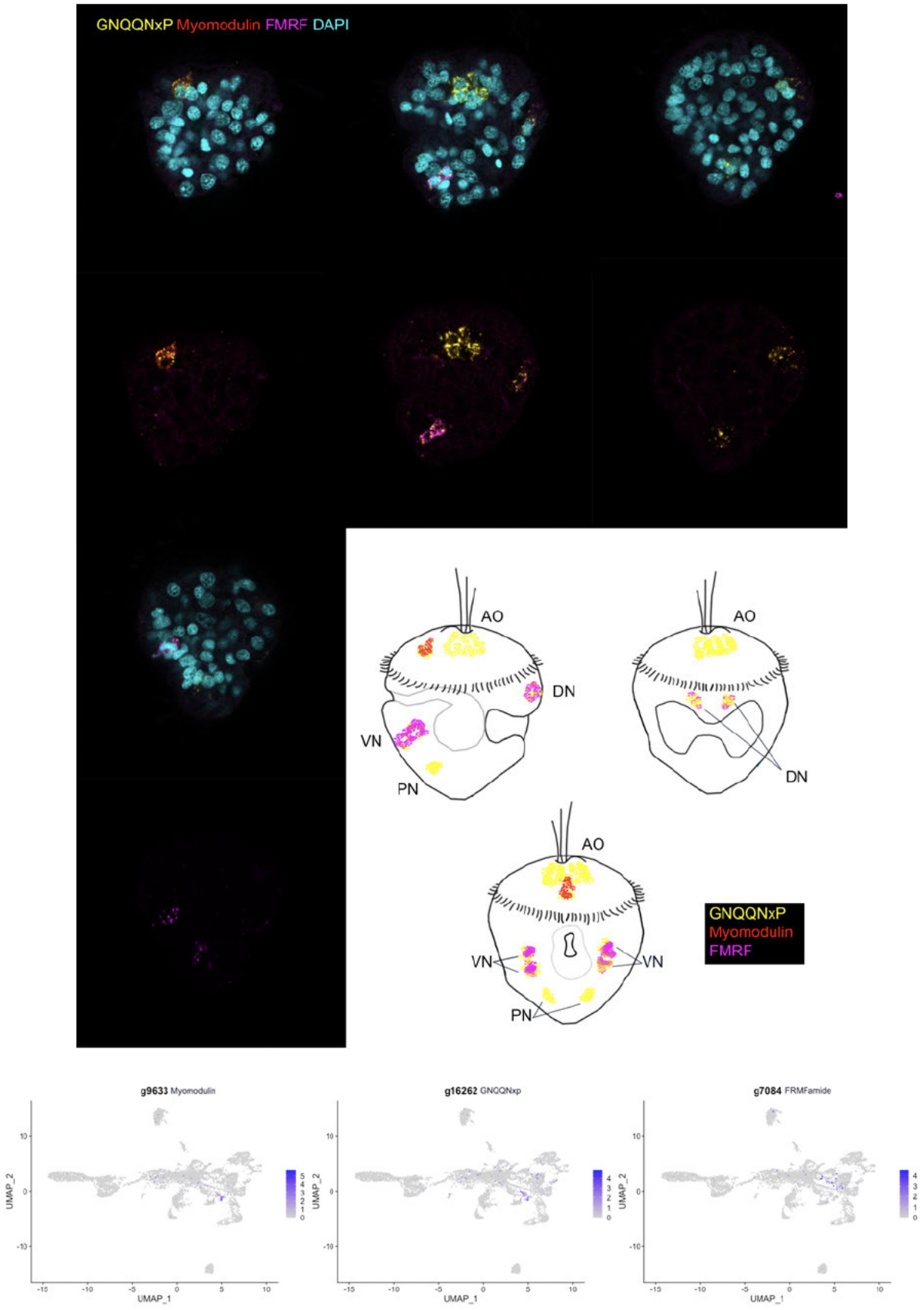


Figure 35. Expression of the neuropeptides GNQQNxp (in yellow), Myomodulin (in red) and FMRFamide (in magenta) using HCR highlights the different neurons of the larva.

Fluorescent images are a single stack moving laterally across the animal (mouth is left, shell is right). Diagram on the bottom summarize the expression of the different markers for the clusters (on the top left lateral view of the larva, like in the stacks, on the top right dorsal view, on the bottom frontal view of the larva). AO= apical organ, DN = dorsal neurons, VN= ventral neurons, PN= posterior neurons. Small UMAPs below show in blue the expression of the genes.

4.2.5 Hematopoietic clusters

The mollusc immune system relies mostly on blood cells (called haemocytes) and it has been extensively studied (probably due to the economic value of many molluscan species such as oysters). Haemocytes have been classified into granulocytes and hyalinocytes based on their cytoplasmic content, both can have phagocytic activity but the former are primarily involved in the cellular response to infection. Although it is well known that mollusc larvae are highly sensitive to viral, bacterial and fungal infections, their immune system is still poorly characterised. What is known so far is that in the oyster the immune system starts developing at the trochophore stage (Dyachuck, 2016).

To understand whether there were any haemocytes, or haemocytes precursors, in our single cell dataset I once again put together a list of haemocytes markers from the literature and searched for their expression in the single cell data (Tirapè et al, 2007; He et al, 2015; Song et al, 2016; Nuria et al, 2020). Together with some general haemocyte markers, I also used specific hyalinocyte and granulocyte markers from a recent proteomic study by Nuria and colleagues (2020) on a different oyster (*Ostrea edulis*). As shown in figure 36, most of the general haemocytes markers as well as specific hyalinocytes marker are concentrated in clusters 8, 10, 11, 13 and 14 whilst granulocytes markers appear scattered. It is possible that

the larval haemocyte molecular signature resembles that of adult hyalinocytes or that granulocytes differentiate at a later stage.

ISH of markers of cluster 8, 10, 11, 13 and 14 show expression in two patches on either side of the gut which appear to be connected anteriorly (see fig. 37).

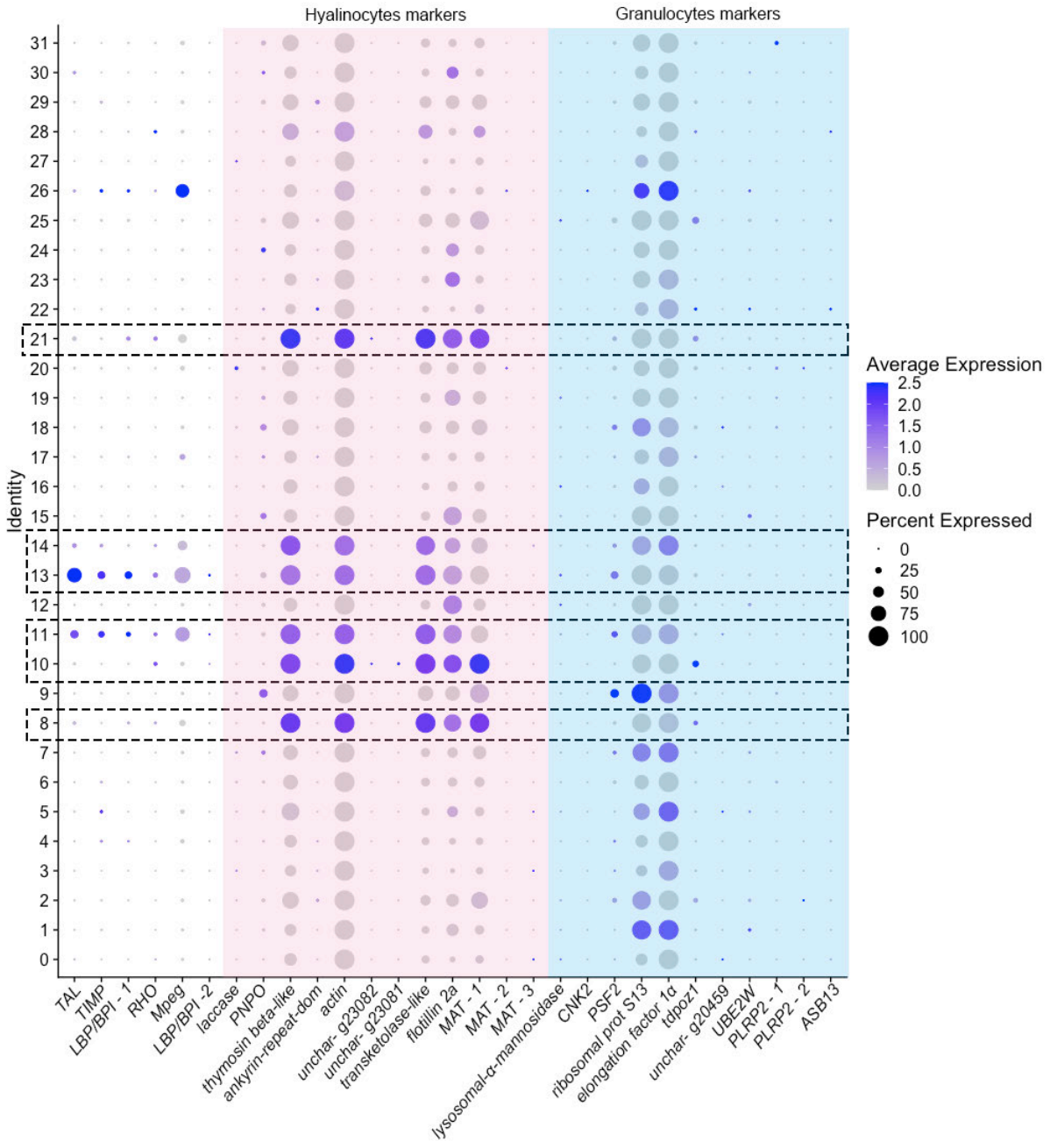


Figure 36. Expression of several general haemocyte markers as well as some hyalinocyte markers is concentrated in cluster 8, 10, 11, 13 and 14.

Most granulocyte markers are expressed in other clusters (such as cluster 26 which will be discussed later) that are likely not haemocytes. Hyalinocyte markers are shaded in pink, granulocyte markers are shaded in blue.

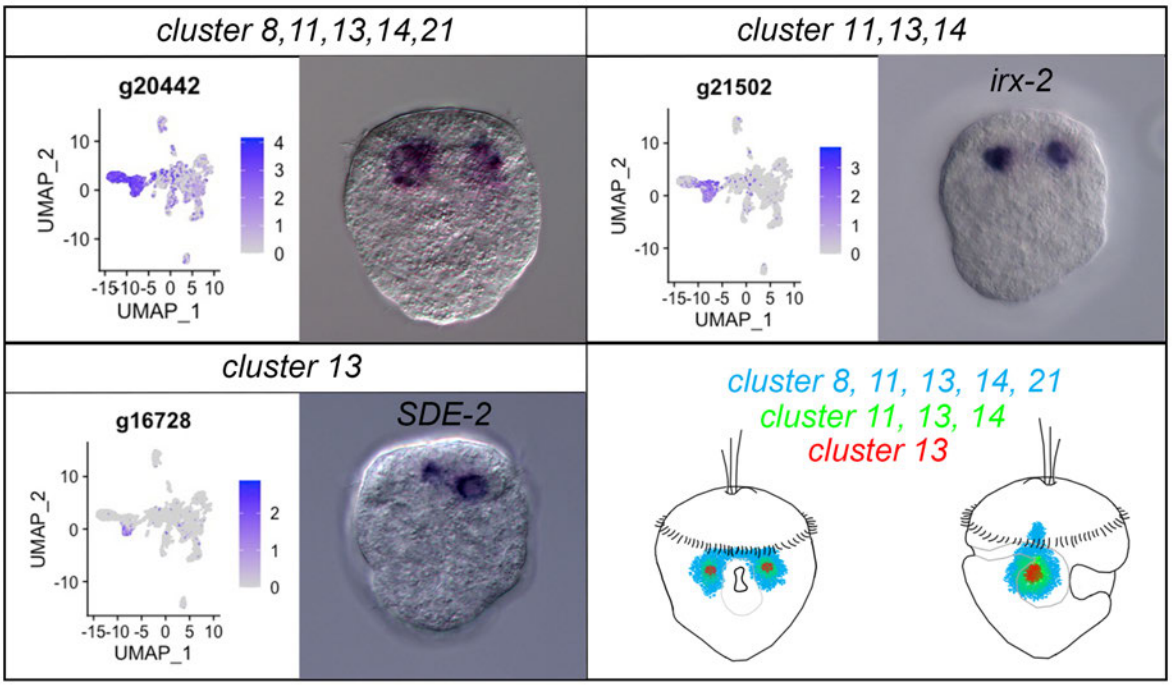


Figure 37. ISH of different haemocytes markers show expression is in two patches on either side of the gut.

Small UMAPs show in blue the expression of the gene, next to ISH of the same gene. Diagram on the bottom summarize the expression of the different markers for the clusters (on the left frontal view of the larva on the right lateral view of the larva, mouth on the left).

4.2.6 Other clusters

So far, I have tried to assign cell type identity by looking for known cell type marker expression in our SCS dataset and then localising the cells with ISH, however some clusters remain unassigned (see figure 38).

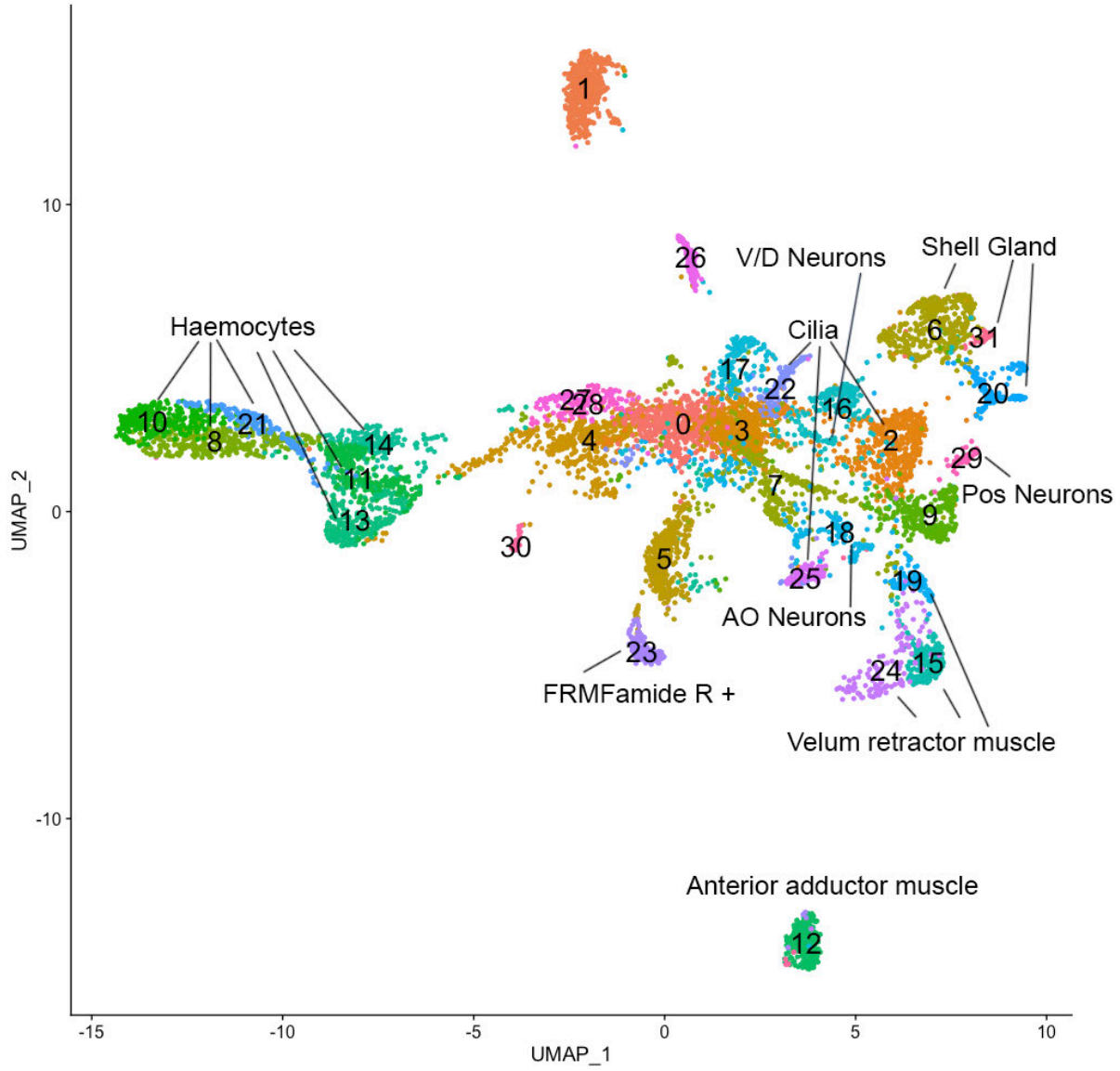


Figure 38. UMAP of the oyster larva SCS showing clusters identified so far with literature search of gene markers.

Pos= Posterior, AO= apical organ, V/D= ventral/dorsal.

4.2.7 Cluster 1

To try and establish the identity of cells belonging to cluster 1, I extracted the top 50 gene markers. Among these genes I found many myocyte markers such as: tnnc (troponin C), caldesmon like, tbb2 (tubulin-2-β-chain), myosin ELC (myosin essential light chain), actin and

muscle LIM protein. Moreover, I can see that many gene markers for cluster 1 are also expressed in other previously described myocyte clusters indicating that cluster 1 could be composed of myocytes (see fig. 39).

ISH of the cluster marker *cgi-pax6* shows that cells belonging to this cluster are scattered around the embryo in a very distinctive pattern (see fig. 40). Such characteristic expression does not, however, seem to resemble any immunohistochemistry of muscle markers previously described. Furthermore, by looking at the list of gene markers for this cluster one can observe many TFs such as twist, DOT1, alx, pax6, foxG and zinc finger homeobox 4 (a complete list of all TFs is found later in this chapter). This could indicate that cell from cluster 1 are still undergoing differentiation, however, among thses pax6 is a notable photoreceptor marker. For this reason, I decided to take a look at the expression distribution of a few other genes involved in bilaterian photoreceptors which have been recently investigated in a scaphopod larva (Wollesen et al, 2019). This include Pax6, as well as other TFs (Eya, Six1/2, Dachsund) different opsins (GO-opsin, peropsin, xenopsin, r-opsin) and a few other genes such as transient receptor potential cation channel (trpC) (involved in phototransduction), myosinV (myoV) which is implicated in intracellular r-opsin transport, and retinitis pigmentosa GTPase regulator (rpgr) used in ciliary opsin targeting. The result of these analysis can be seen in figure 41 and highlight how cluster 1 co-expressed Pax6, Eya and Six1/2 but no opsin. It is hence also possible that this cluster could give rise to photoreceptor cells at a later stage in development although it remains unclear why it would co-express so many myocyte related genes.

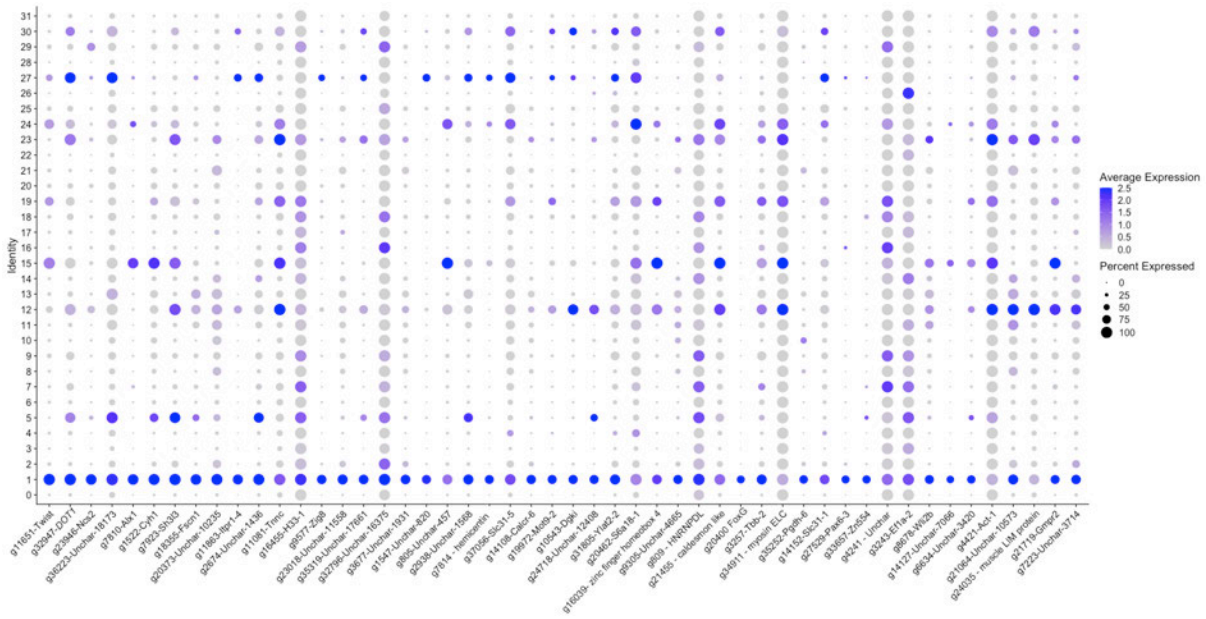


Figure 39. DotPlot of the expression of top 50 gene markers for cluster 1, many gene markers are myocyte related.

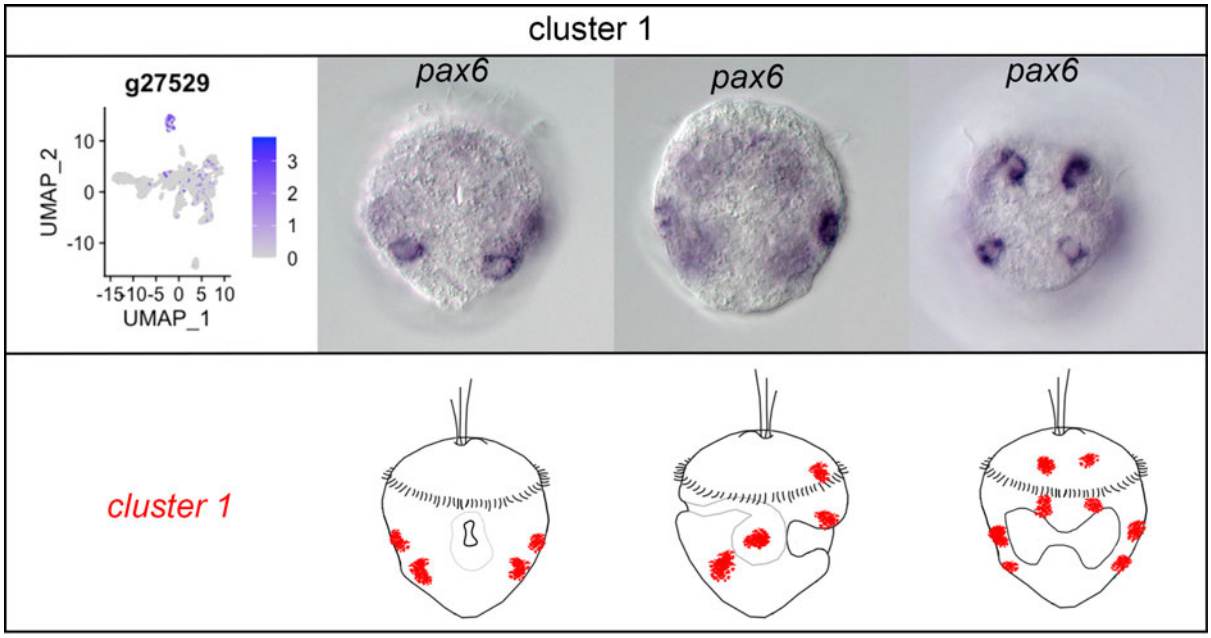


Figure 40. ISH of gene marker cgi-pax6 shows expression in cells scattered around the embryo in a distinctive pattern.

Small UMAPs show in blue the expression of the gene, next to ISH of the same gene (three focal planes moving frontally from the mouth towards the shell). Diagram on the bottom summarize the expression of the gene (on the left frontal view of the larva, same orientation as the ISH, in the middle lateral view of the larva, with mouth on the left, on the right dorsal view).

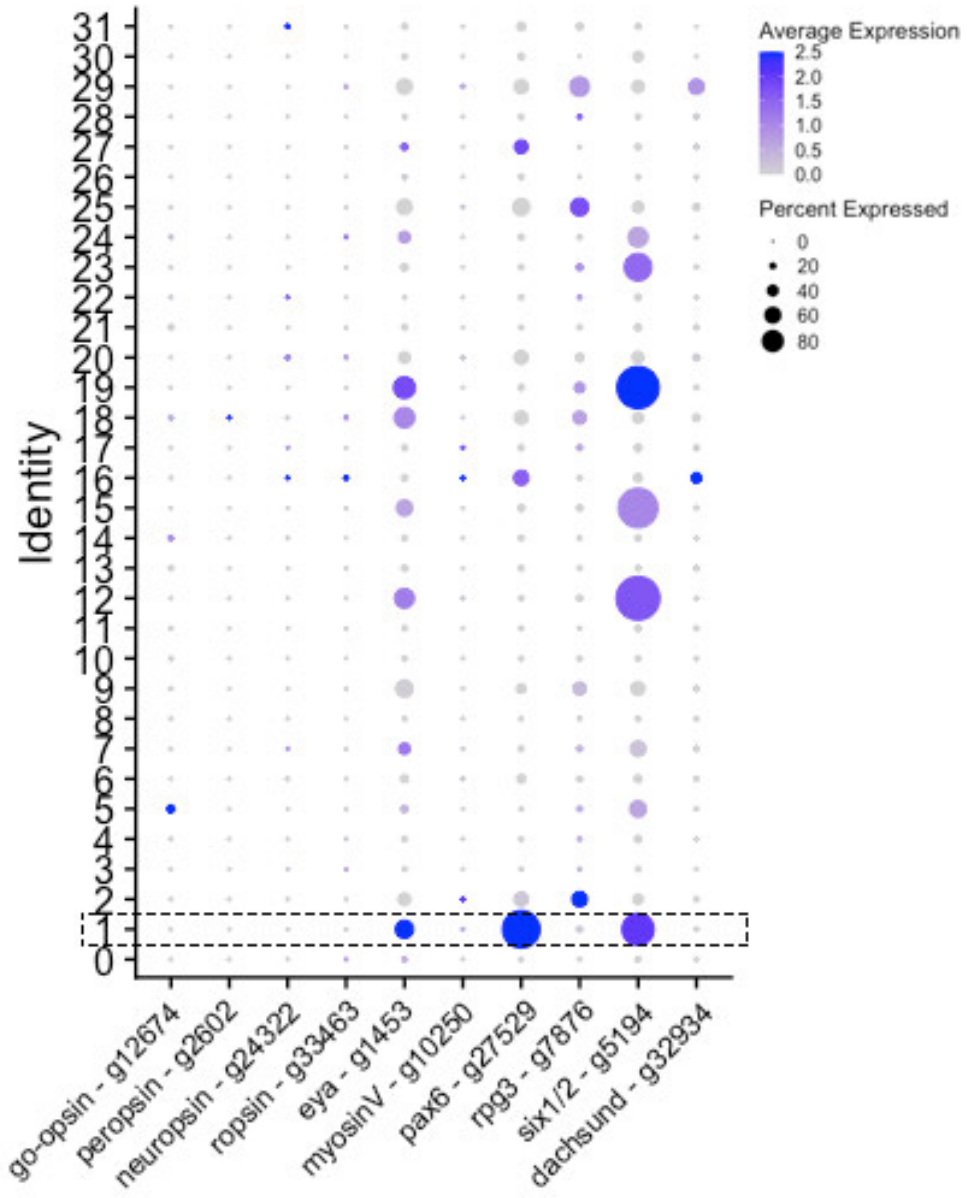


Figure 41. Expression of typical TFs of photoreceptors eya, pax6 and six1/2 1 are co-expressed in cluster 1.

Dotplot showing the expression of several opsins, and photoreceptor specific genes from Wollesen et al, 2019. Most TFs are expressed in cluster 1 although no opsins are expressed here.

4.2.8 Cluster 5

Similar to what I have already described for cluster 1, I extracted the top 50 gene markers

for cluster 5; as one can see in figure 42, many markers are shared with cluster 23 (a

myocyte cluster with FMRF-receptor positive cells). It is possible that cluster 5 could indeed represent another myocyte cluster since it also shares many markers with clusters 1, 12 and 15 (all previously identified as myocyte clusters). Moreover, amongst the top 50 gene markers I can find Tnnt (troponin-T), Act-1 (actin 1), unconventional myosin XVI and muscle M line assembly protein. As seen for cluster 1 cluster 5 also expresses many TFs such as Gata3, FoxF1, Six6, VBP, Elf1a, Erh, Rfox1, so once again it is likely that cells in cluster 5 are still undergoing differentiation.

Another possibility is that cluster 5 contains developing haemocytes, as some of the gene markers (such as Atcp, or actophorin, and BTG1) are associated in the literature with immune response. However, this is less likely considering that hardly any markers are shared between cluster 5 and our previously identified haemocyte markers.

As for their localisation, ISH of the marker *Hs3s5-1* shows that cells of cluster 5 are found medially in two patches, one, smaller, above the mouth and one below the shell gland (see figure 43).

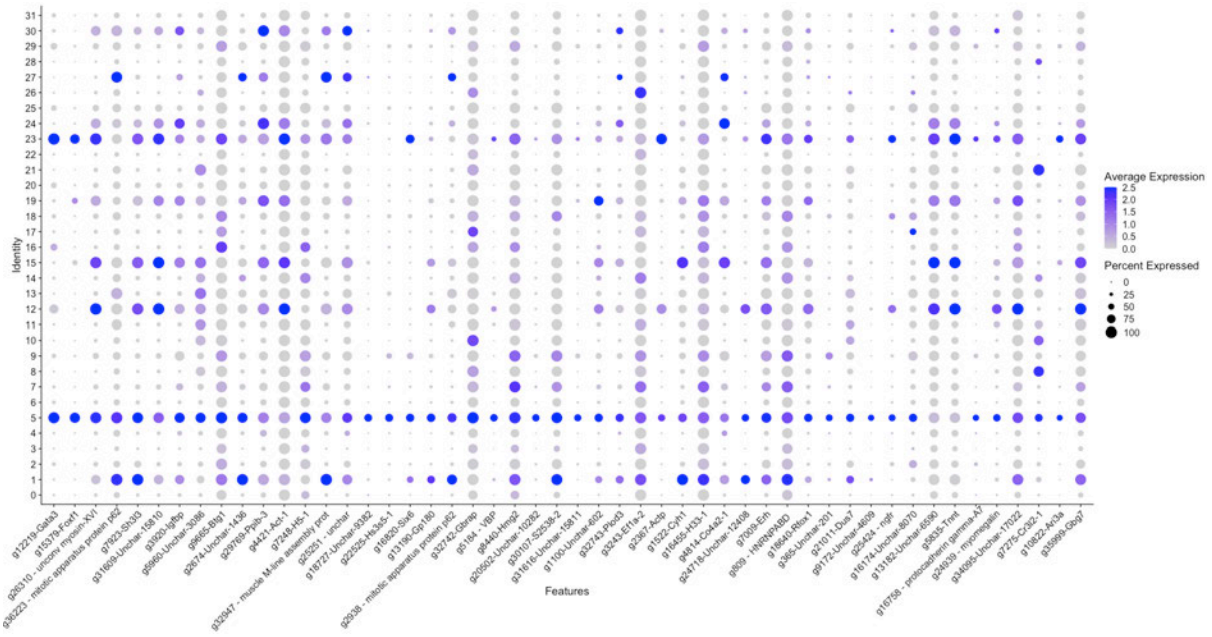


Figure 42. Markers of cluster 5 overlap with cluster 23, a muscle cluster.

Dotplot of top 50 markers for cluster 5 highlighting an overlap with markers of cluster 23.

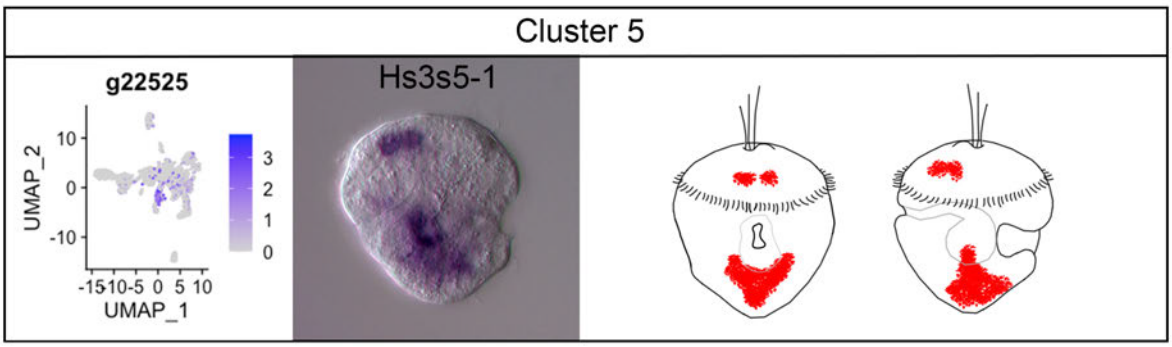


Figure 43. ISH of the cluster marker Hs3s5-1 showing staining in two patches.

Small UMAP shows in blue the expression of the gene, next to ISH of the same gene and diagram summarizing the expression of the gene (on the left frontal view of the larva, on the right lateral view of the larva, with mouth on the left, same orientation as the ISH).

4.2.9 Cluster 7 and cluster 9

Since cluster 7 and cluster 9 seem to share the majority of their markers they will be considered here together (see fig. 44). Interestingly, among their top markers I identified a few genes, such as Mgn2 (mago-nashi 2), Sumo3, Pcna, CBX1, that have been described in

the literature as playing a role in stem cell proliferation (Orii et al, 2005; Eisenhoffer et al, 2008; Thiruvalluvan et al, 2018; Kimball et al, 2020).

To check whether indeed cluster 7 and/or 9 transcriptional signature resembled that of proliferating cells, I decided to take a look at the expression of planarian neoblast markers in our oyster SCS dataset. To identify orthologous genes, I ran Orthofinder (Emms & Kelly, 2019) on our proteome alongside the complete set of genes from the genome of *Schmidtea mediterranea* and I then collected all the genes of the oyster that were orthologous to neoblast markers as described in the Plass et al. paper (2018). Figure 45 clearly shows that most of the neoblast markers of planarians that were expressed in our oyster SCS are indeed expressed in cluster 7 and 9 reinforcing our hypothesis that these could represent proliferating cells. It is unclear whether these actually represent neoblasts/stem-like cells or whether they're simply proliferating larval cells. In fact, neoblast (and stem cells in general) are usually characterised by three features: a specific molecular signature (that these cells have), active proliferation and totipotency (which I did not investigate).

ISH of marker genes for cluster 7 and 9, respectively APOBEC-1 and g24584 (which did not blast to any previously annotated gene) appears to show expression in the gut (see figure 46). It is possible that proliferating cells are present here since at the trochophore stage larvae starts to develop a gut to initiate feeding (Dyachuk et al, 2012).

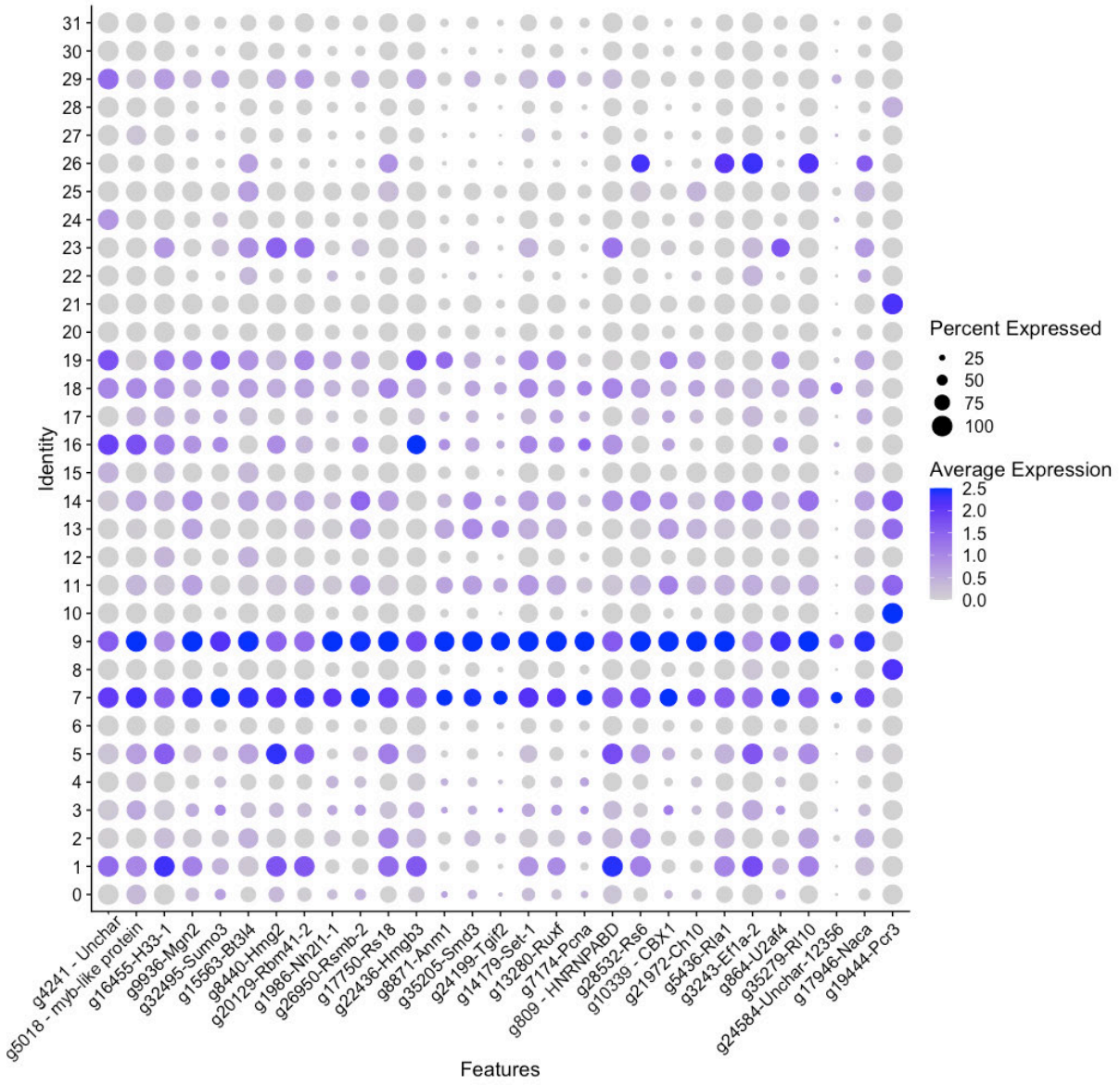


Figure 44. DotPlot showing the expression of the top 29 markers of cluster 7 which largely overlap with marker for cluster 9.

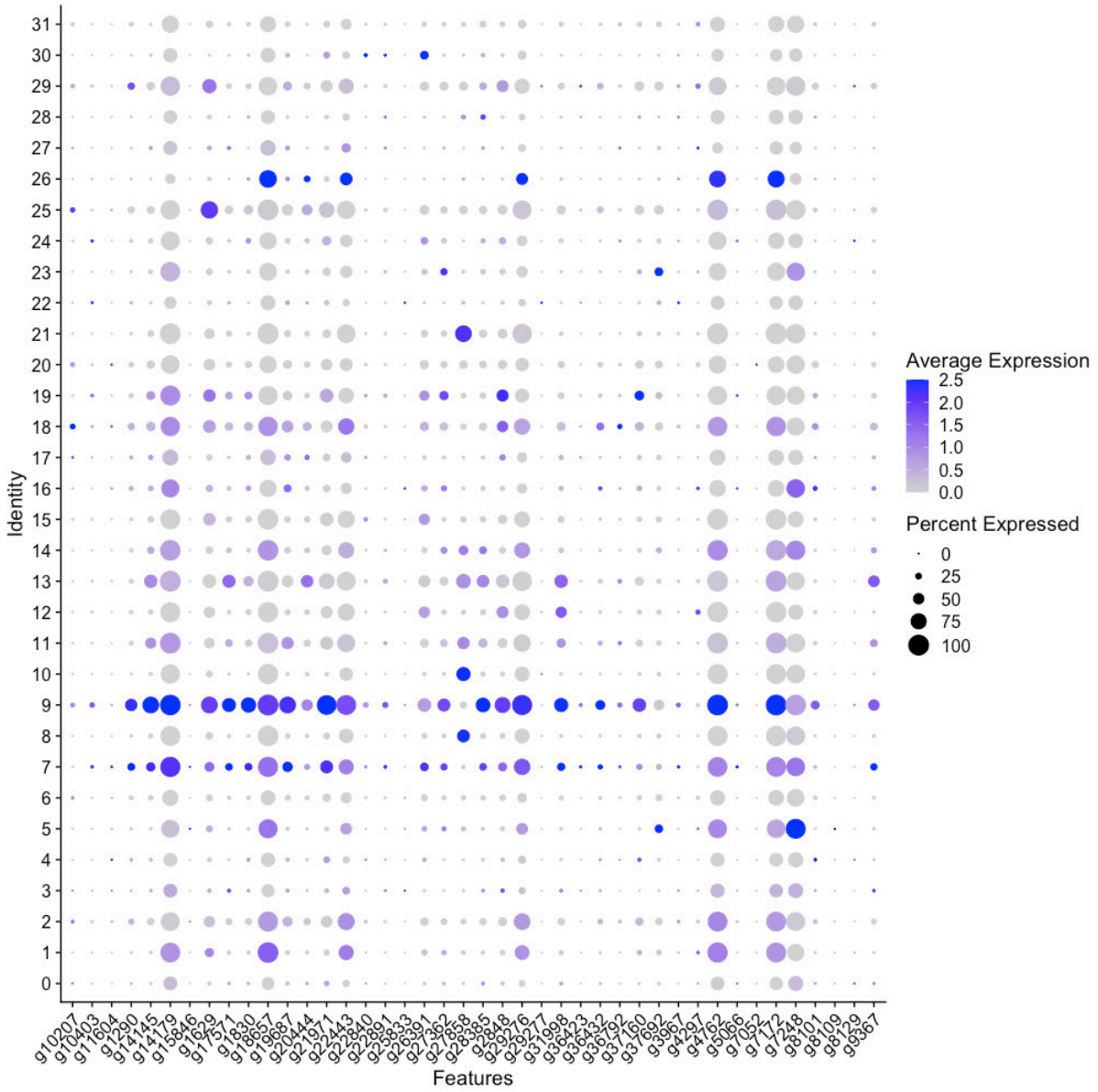


Figure 45. Expression of planarian neoblast markers is concentrated in cluster 7 and 9.

Marker genes of planarian neoblast taken from Plass et al. (2018), orthologs were identified using Orthofinder.

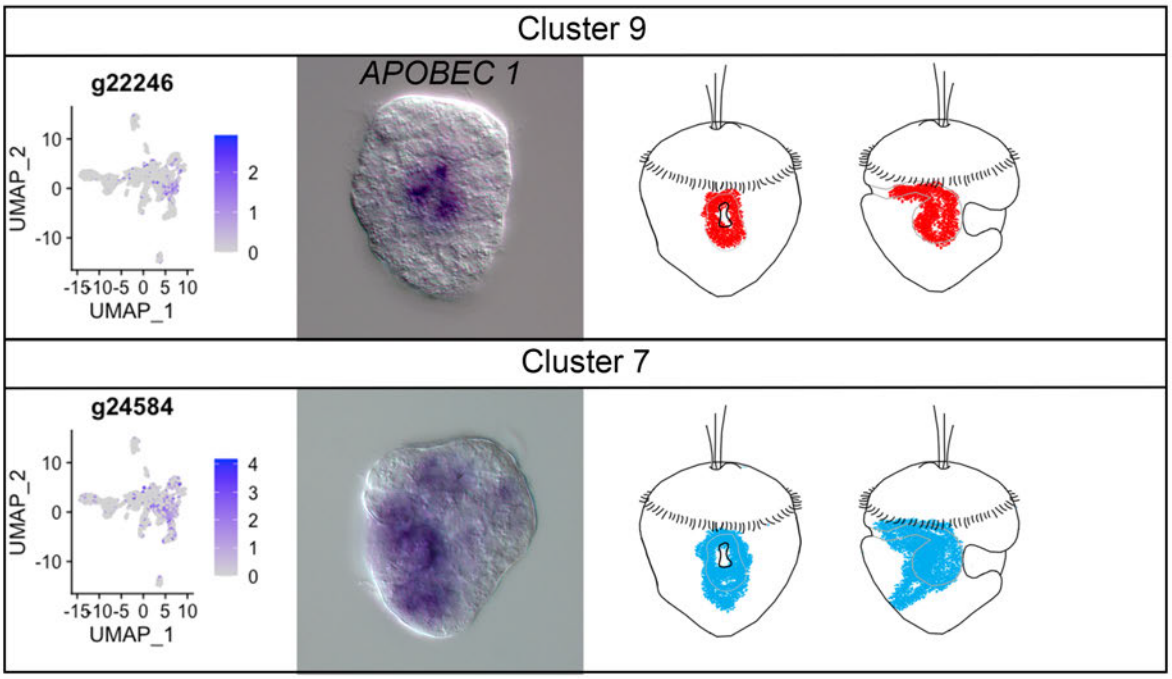


Figure 46. ISH of marker genes of cluster 9 and 7 show expression in the region of the developing gut.

Small UMAPs shows in blue the expression of the gene, next to ISH of the same gene and the diagram summarizing the expression of the gene (on the left frontal view of the larva, on the right lateral view of the larva, with mouth on the left). APOBEC 1 ISH is view frontal, ISH of g24584 is a lateral view, mouth on the left.

4.2.10 Cluster 16

Many markers of cluster 16, as one can see in figure 47, could not be annotated as a BLAST returned no hits against previously annotated genes. Different literature searches for the other annotated genes of these cluster did not lead to many results. In fact, the only published study that mentions many of the gene markers for cluster 16, such as calreticulin, annexin b9 and calmodulin is a transcriptomic study of the pallial gland of the date mussel, *Lithophaga lithophaga* (Sivka et al, 2018). In the same study the authors mention that the pallial gland secretions, which allow the mussel to bore into rock, contain glycoproteins, and cells of cluster 16 have, as their top markers, 2 mucin genes. However, to the best of my

knowledge, oysters do not seem to have a pallial gland. It is possible, since many markers from cluster 16 are shared with clusters 31 and 2 (both shell gland clusters) that cells in cluster 16 are part of a shell structure, possibly with a secretory role. This idea is backed up by my ISH which shows expression in cells arranged in two circles laterally to the animal outlining the border of the shell (see figure 48).

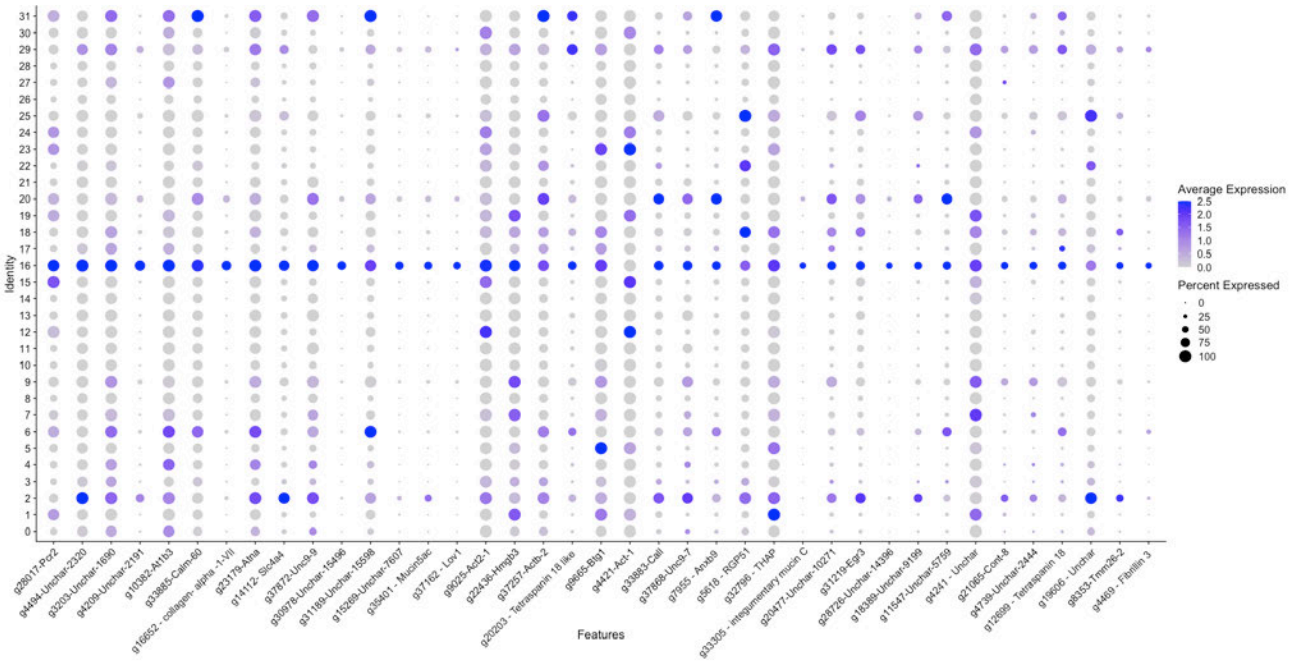


Figure 47. Dotplot showing the expression of the top markers for cluster 16 showing some overlaps with cluster 2, 6 and 31 which are shell gland clusters.

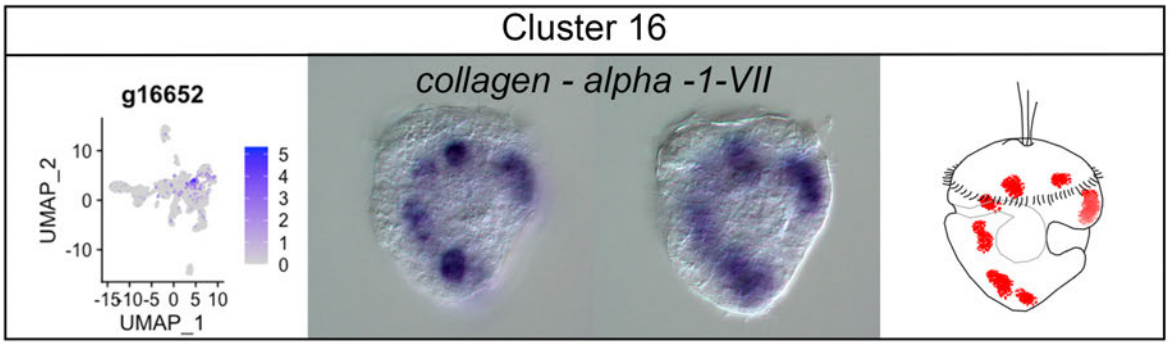


Figure 48. ISH of the gene marker collagen-alpha-1-VII shows that cells from cluster 16 are located on either side of the animal in two circles that follow the outline of the shell.

Small UMAP shows in blue the expression of the gene, next to ISH of the same gene (two focal planes moving laterally) and diagram summarizing the expression of the gene (lateral view of the larva, with mouth on the left, same orientation as the ISH).

4.2.11 Cluster 26

Cluster 26 is a small cluster that appears to have a very specific transcriptional signature with most of its top markers not shared across any other clusters (see figure 49). Moreover, when searching the literature for any of these top markers I could not find any substantial mention in previous studies with the exception of cathepsin L-1 which is indicated in a study as a phagocyte marker (Lv et al, 2018). Chromogenic *in situs* for a selected marker (g33097) did not show any staining, for this reason I decided to try HCR on another gene marker (g24588-FMRF-r-2 which is a different FMRFamide receptor from the one used for cluster 23). HCR of this gene show it is expressed in very few cells (possibly only one) located near the apical organ of the larva (marked here with the general neuronal marker GNQQNx) that co-express the neuropeptide GNQQNx (see figure 50). This result was not highlighted by our SCS data where these two genes do not appear to be co-expressed in any cells. However, as we can see in figure 28, cluster 26 appears to be among the clusters with fewest UMIs and fewest genes detected so it is possible that we missed some genes for this cluster. In general, since there is only one cell in the larva that belongs to this cluster it is likely that we would have a lower coverage for this cluster. In any case, considering that cells in this cluster are in close proximity of the apical organ and they co-express the general neuropeptide marker GNQQNx together with FMRF receptor and neuromedin-u receptor it is possible that cluster 26 is another small cluster of neurons although it doesn't appear to share many markers with other neuronal clusters.

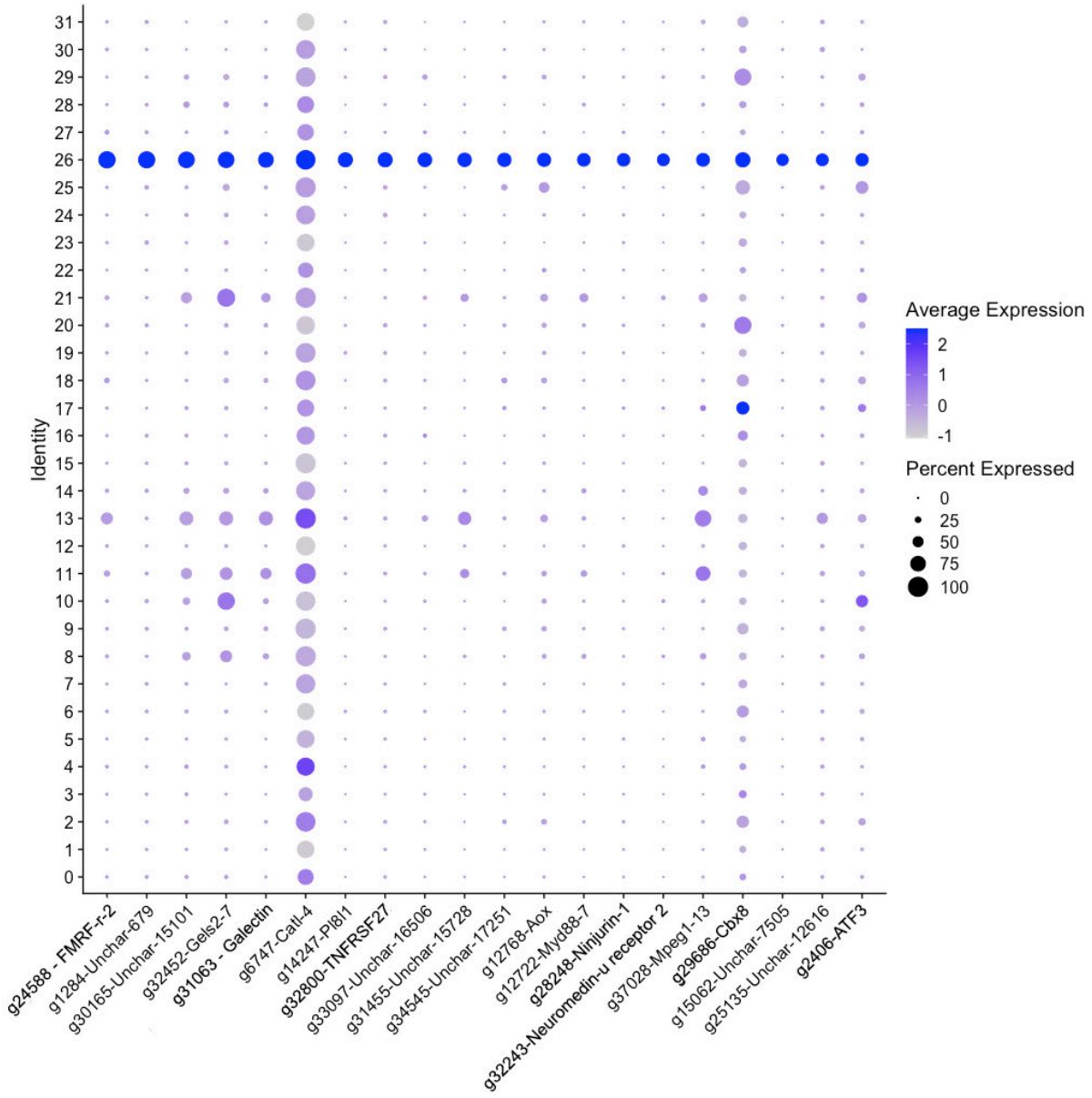


Figure 49. Dotplot of the top 20 marker genes for cluster 26 showing one phagocyte marker (cathepsin-L) and a few neuronal markers (FMRF-r and Neuromedin-u receptor).

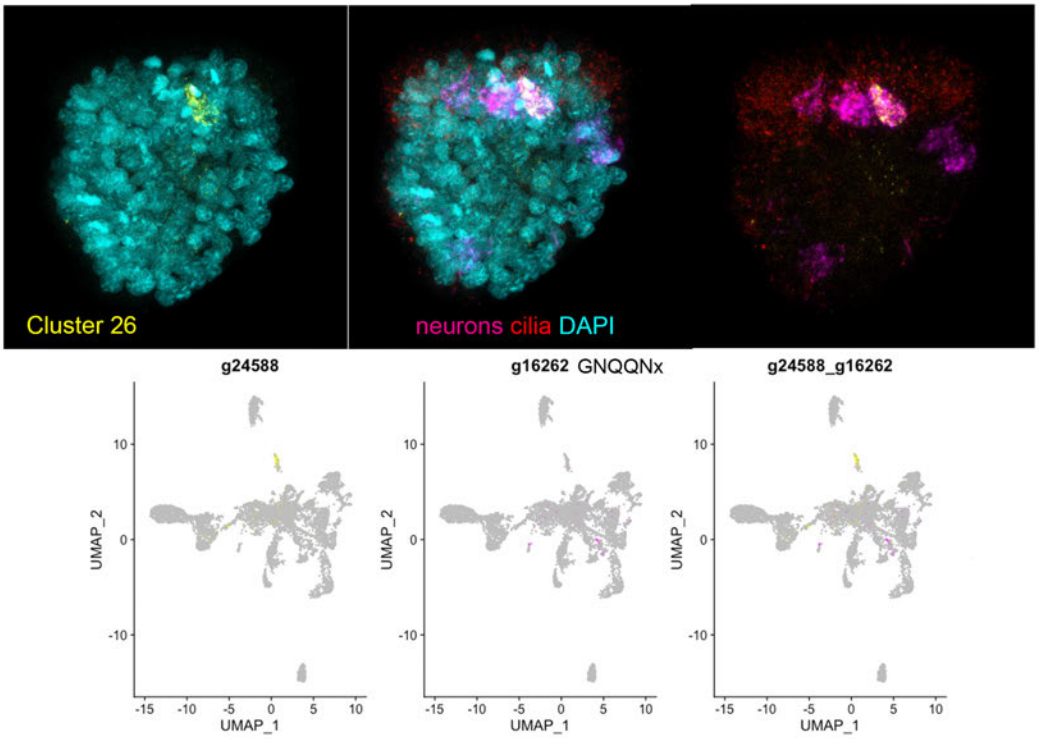


Figure 50. Cluster 26 is likely made up of only one cell that co-express the neuronal marker GNQQNx.

HCR of gene markers of cluster 26 (g24588 – yellow) together with a neuronal marker (g16262 – GNQQNx - magenta) and a ciliary marker (g31376 - red). DAPI in light blue stains nuclei. As shown in the small UMAP expression plots below, this result doesn't match our SCS data where no cells seem to co-express the two genes.

4.2.12 Remaining clusters

After several literature searches, analysis of marker genes and ISH I still have some clusters that remained unidentified. I believe that clusters 0, 3, 4, 27, 28 and 30 likely contain either undifferentiated cells in the larva or lower quality cells since they are among the clusters with lower UMI and gene content and higher mitochondrial gene content (see figure 51). These central clusters are indeed quite common in most SCS published papers and usually remain unidentified.

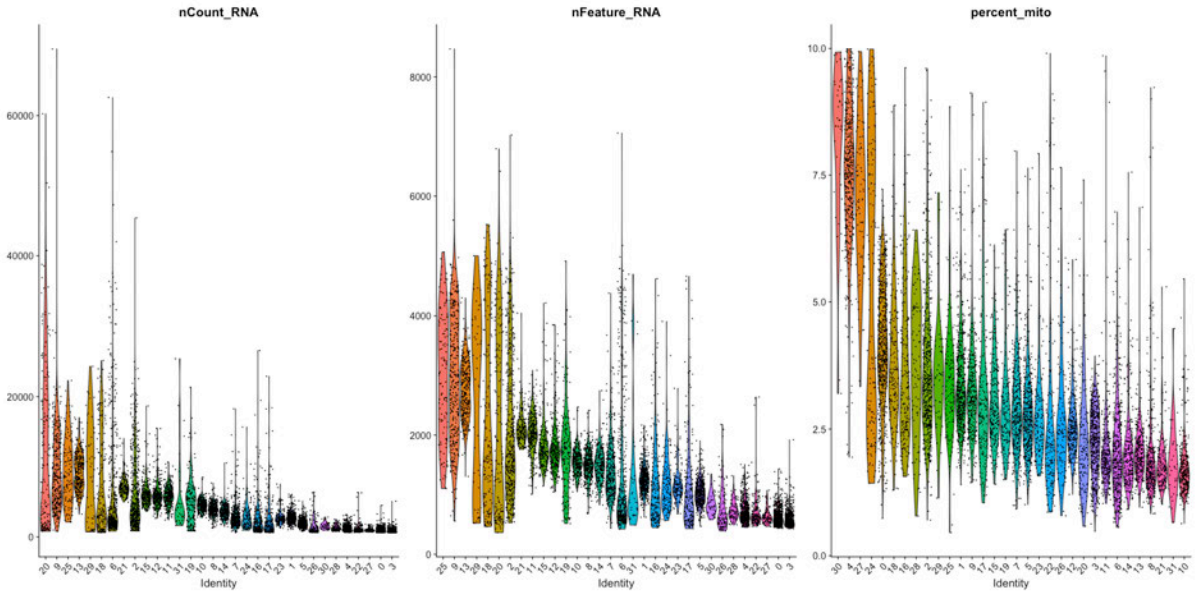


Figure 51. Clusters 0, 3, 4, 27, 28 and 30 show poorer quality compared to the others.

Violin plots of content of UMI (nCount_RNA), genes (nFeature_RNA) and percentage of mitochondrial genes (percent_mito) in cells of different clusters (x axis) shows that clusters 0,3,4,27,28 and 30 are among clusters with lower UMI and gene content and higher mitochondrial gene content.

4.3 Relationships between cell types in the oyster trochophore larva

In the previous paragraphs I attempted to assign cell type identities to the different cell clusters generated with our single cell sequencing data using a combination of known markers and *in situ* hybridisation. Using this approach, I realised that for most cell types (i.e. myocytes, haemocytes, ciliary band cells, shell gland cells, neurons) more than one cluster was identified. I was interested in finding out what the relationship between these clusters (which often seem to share many markers) may be, since UMAPs often do not depict relationship between nearby clusters well. Moreover, I though this information might be useful for the final aim of this thesis which is comparing cell types across Lophotrochozoan larvae.

To establish relationships between clusters, I used an R script devised by our collaborator Jacob Musser (Arendt lab, EMBL Heidelberg) that takes the log transformed average expression matrix per cluster and uses it to calculate the Euclidean distance between each cluster. Bootstrap support for the tree of relationships between clusters is computed as the percentage of replicates that found a particular clade shown in the tree (out of 10000 repeats). The result of this approach is the tree shown in figure 52. As one can see, oyster clusters are divided into three major 'clades': haemocytes, muscles and a large clade containing shell gland cells, neurons, ciliary bands and neoblasts. Support for haemocytes, muscle, neurons, shell gland, ciliary bands and neoblast clades are quite high. Support for relationship between different cell types are fairly low (~50%), except for the clade containing ciliary bands and neurons (97%).

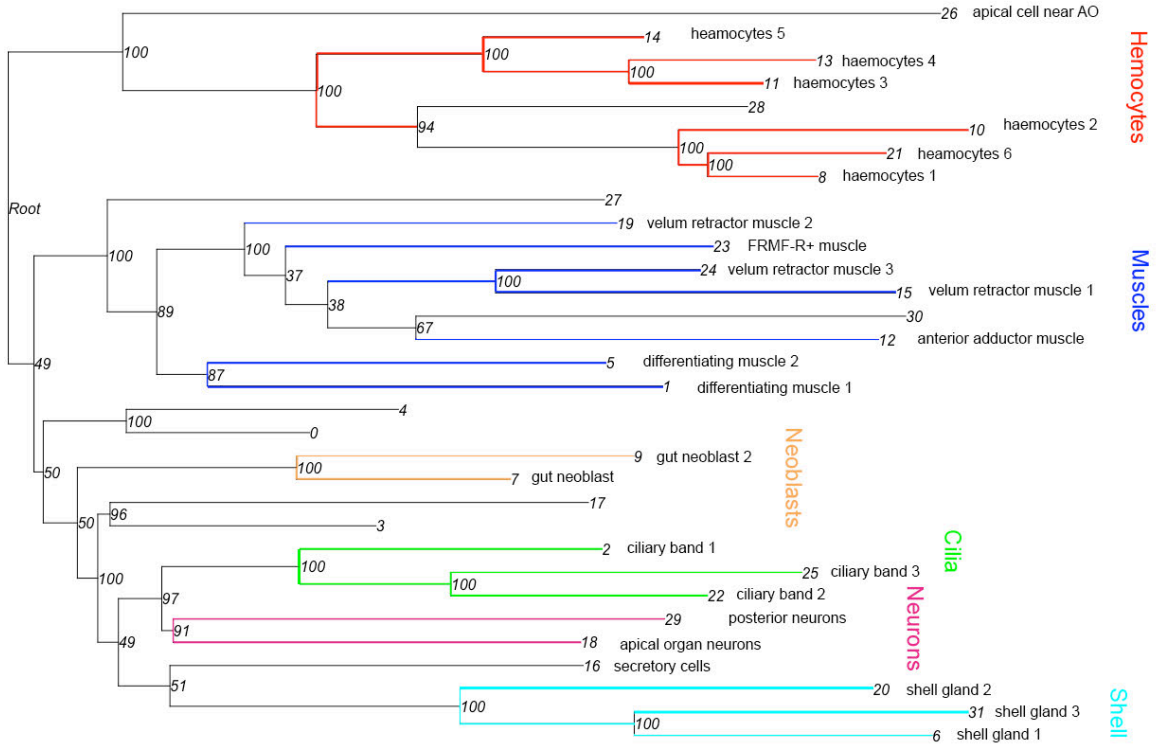


Figure 52. Transcriptional profile of ciliary cells and neurons appear similar.

Cell type trees showing relationship between clusters calculated as Euclidean distance; most similar cell types (i.e. muscles, neurons, shell glands cells, haemocytes, ciliary bands) group together. Bootstrap values represent % of time that clade was recovered (10000 repeats). Tree was rooted via midpoint rooting (roots at the midpoint between the two tips most distant from each other).

4.4 Transcription factor signatures in different clusters

As previously mentioned, my SCS clustering showed the presence of many different clusters of similar cell types such as muscles, haemocytes, neurons, cilia and shell gland cells. In the previous section I have shown how each of those cell types are indeed more similar to others of the same type but fairly different to other groups (with possibly the exception of ciliary cells and neurons), for this reason I thought it would be interesting to try and work

out if they showed a different transcription factor signature. In order to identify transcription factors amongst the oyster genes I ran Orthofinder with a list of TFs from human and *Drosophila* and then retained all genes of the oyster that belonged to an orthogroup that contained at least one TF from either fly or human. The expression of transcription factors in different clusters, which is shown in figure 53, is clearly very cluster specific with different blocks of TFs activated in pretty much all clusters except for clusters 0, 3, 4, 27, 28 and 30 which, as discussed earlier, likely contain lower quality cells. Moreover, even clusters with the "same" cell identity (i.e. muscles, haemocytes, neurons, cilia and shell gland cells) show very distinct TFs signatures.

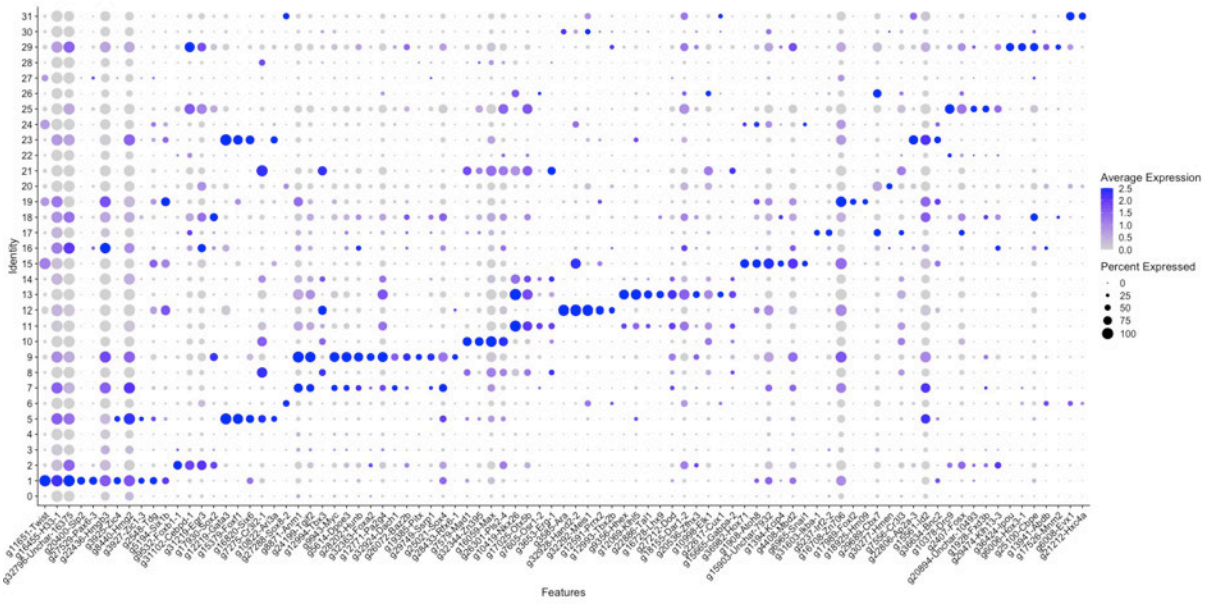


Figure 53. Different clusters have substantially different TFs signatures.

TFs were identified as orthologs to human and/or drosophila TFs using Orthofinder. Only TFs that are also cluster markers are shown here.

4.5 Transcriptome age index (TAI) for single cell clusters in the oyster

After having identified most cell clusters and having worked out the relationships between different clusters, I was interested in taking a look at the transcriptome age index of these cell types.

I wanted to use the ages of the genes expressed in each cell cluster to ask whether some cell clusters have younger genes than others. This might indicate either that that the cell cluster has originated more recently or that it has started expressing new genes more recently (for example due to selection) than other clusters with an older TAI.

I thought this could be interesting because, as already mentioned in the introduction, three recent studies have shown that the trochophore larvae seem to express "younger" genes than other developmental stages (Xu et al, 2016; Wu et al, 2019; Wang et al, 2020). This result has been interpreted by some authors as a proof for a more recent origin of trochophore larvae. Since these results were obtained using bulk RNA seq data I wanted to explore the differences in TAI between different cell types.

To compute the TAI, I firstly needed to assign all genes from our focal species (in this case the oyster) to a different phylostratum, to do so I used the R based framework phylostratr (Arendsee et al, 2019). A phylostratum represents a major taxonomic level in the evolution of the species of interest so for instance, for the oyster, some of the phylostrata will be Bivalvia, Mollusca, Lophotrochozoa, Metazoa and so on. To assign genes to each phylostratum, phylostratr blasts each gene from the focal species (the oyster) against a database of other living organisms and assesses how far down in the tree there is a significant hit. For instance, if a gene A has hits only in bivalves it will be assigned to the

Bivalvia phylostratum, if a gene B also has a hit in *Drosophila* (but not in more distant taxa) it will be assigned to Protostomia and so on.

At this point, I decided first to look at the distribution of genes across different phylostrata. As one can see in figure 54 there's a big difference in number of genes belonging to different phylostrata, with the majority of genes being shared across cellular organisms. Importantly, very few genes appear to be restricted to either the Protostomia, Lophotrochozoa or Spiralia clades. I do not know for sure what factors may drive these differences but they will surely affect downstream analysis since the probability of a gene to belong to a particular PS or another are not equal.

I then moved on to calculate the TAI for each cluster using the R package myTAI (Drost et al, 2018). Briefly, I first calculated the log transformed average of gene expression per cluster – in the classical TAI approach genes with higher expression are given a higher weight, for this reason log transforming the expression reduces the bias towards highly expressed genes. Then, for each cluster the TAI was calculated as the weighted mean of the evolutionary age (phylostratum) of each gene weighted by the (log-transformed) average expression level of that gene in that cluster. Using this approach small PS values correspond to older PS and so theoretically the smaller the TAI the "older" the transcriptome of that cluster. However, there's a few things to keep in mind when using this approach: firstly, fast evolving genes could be incorrectly assigned to younger PS as it is harder to find a homolog in a distantly related species (Natsidis et al, 2021), secondly, since gene age is weighted by their expression level some highly expressed genes could skew the results. For this reason, as previously explained, I decided to log transform the expression average of each gene per

each cluster but one could also consider simply using genes that are expressed above a minimum threshold in a specific cluster.

As seen in fig 55, shell gland clusters appear to have a considerably higher (younger) TAI compared to the rest of the trochophore. This is a striking result since it may indicate that the peak in TAI observed in previous studies could, at least in molluscs, be driven by shell gland gene expression starting at the trochophore stage and not by the fact that the larva is young. This younger TAI could be due to the later addition of a shell in an older larval body or to the fact that the shell underwent faster evolution than the rest of the animal. Regardless, this shows the importance of looking at gene expression at the cell level rather than in bulk and is definitely an interesting analysis that could be carried out on other larvae.

TAI numbers only make sense in a comparison, they tell us if a developmental stage or cell type expresses more “young” or “old” genes compared to another, however we can also investigate the contribution of genes belonging to different PS to the total TAI. This analysis, shown in figure 56 clearly shows that the higher TAI observed for the shell gland clusters is caused by genes belonging to mollusc or younger phylostrata. This backs up the idea that the peak in TAI is indeed due to the co-option of the shell gland in the mollusc larva and could potentially indicate that the rest of the larva is older.

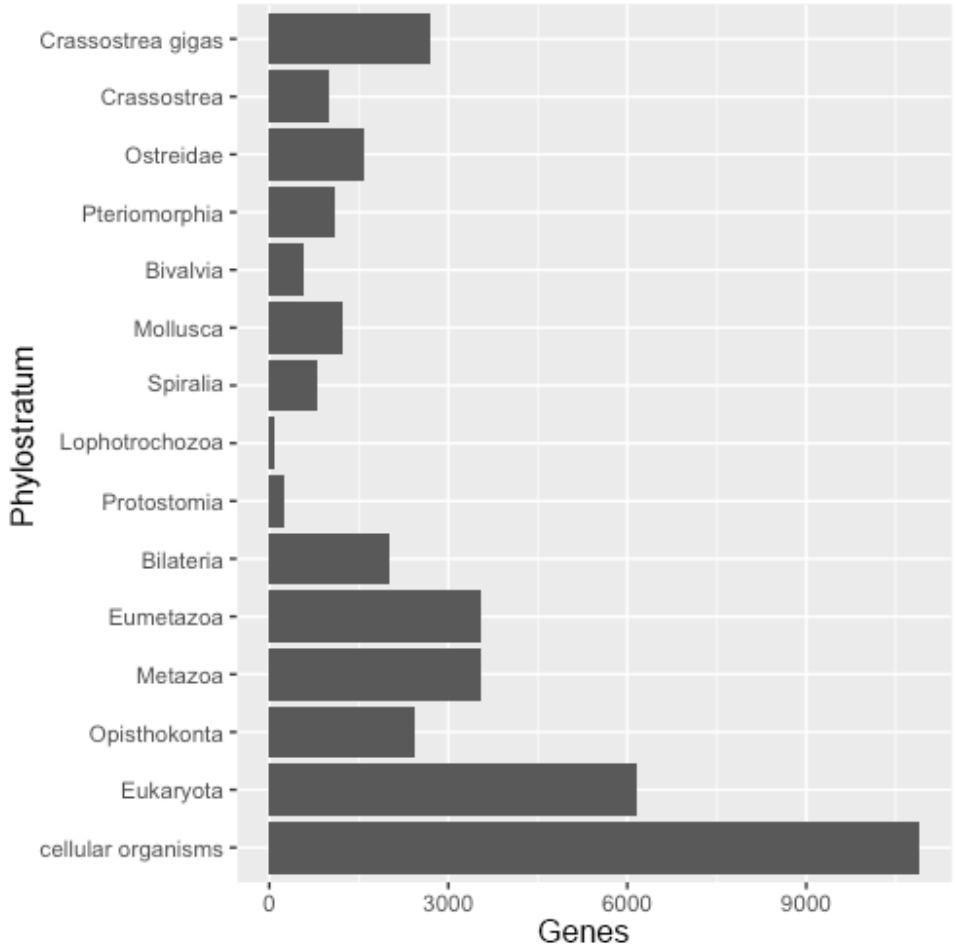


Figure 54. Distribution of oyster genes across different PS is very variable with the majority of genes having a hit across cellular organisms.

Very few genes belong to the Protostomia, Lophotrochozoa and Spiralia phylostrata.

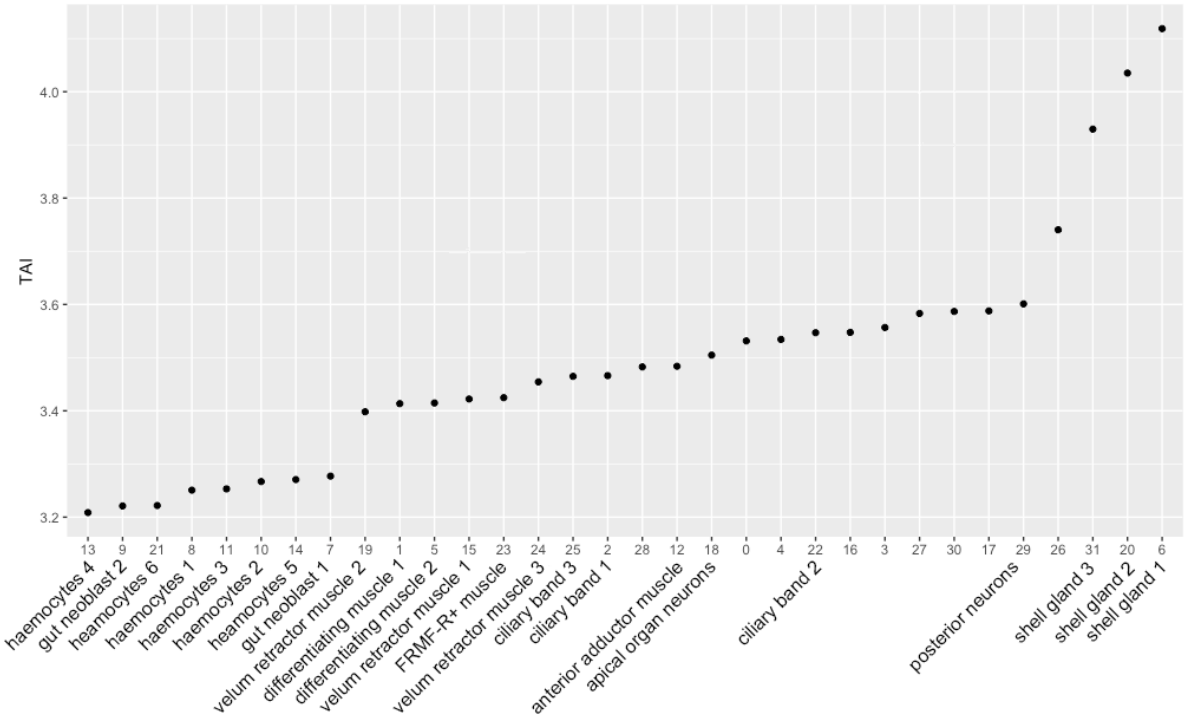


Figure 55. Shell gland clusters express "younger" genes than any other cluster.

TAI values of different clusters in the oyster, shell gland clusters present a considerably higher TAI then the rest.

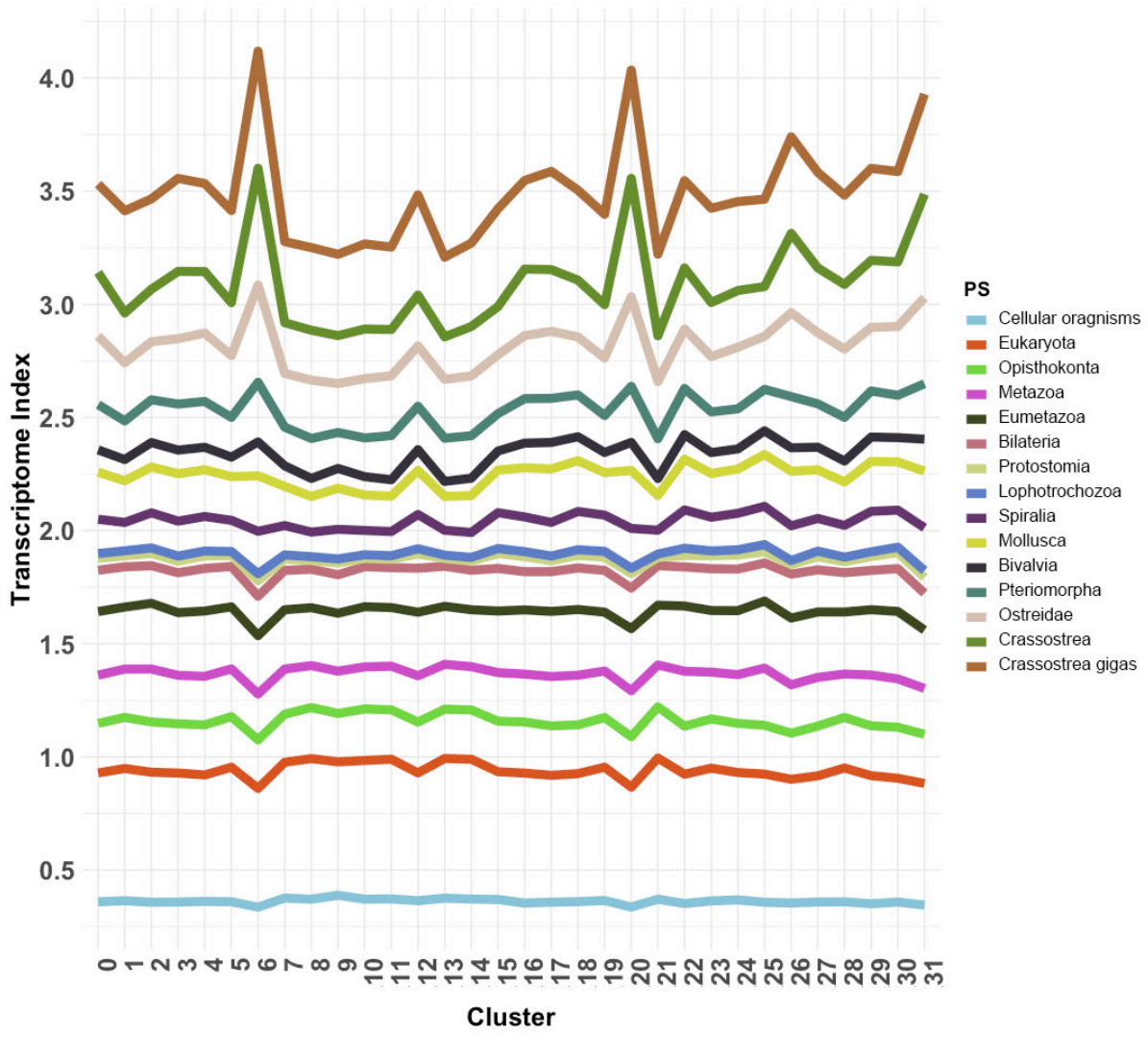


Figure 56. "younger" gene contribution for the shell gland clusters (6, 20 and 31) starts at the Mollusca level.

Contribution of different phylostrata to the total TAI of the oyster divided by cluster.

4.6 Conclusions

4.6.1 Summary of results

In this chapter I tried to identify the main cell types present in the trochophore larva of the oyster. Using a combination of literature searches and *in situ* hybridization I managed to characterise almost all clusters obtained. Moreover, I used the average transcriptional profile of each cluster to establish relationship amongst them showing that similar cell types (such as muscles, haemocytes, ciliary bands and neurons) are indeed more similar to each other than any other groups. Finally, I have looked at the transcriptional age index (TAI) of each cluster and demonstrated that the higher TAI recently observed by some authors in mollusc trochophore larvae is likely due to the rapid evolution of the shell gland. All this information will be used in chapter 6 to compare the trochophore larva of mollusc with the Muller's larva of polyclad flatworms.

4.6.2 Classical features of trochophore larvae found in the oyster larva

We picked the oyster larva for this study because to the best of our knowledge it was a canonical trochophore larva. In fact, it presented a visible prototroch and previous studies had shown it had an apical organ (Yurchenko et al, 2019). Moreover, when I started investigating the identity of the different cell types I found several neuronal clusters (or subclusters) with cluster 29 containing cells of the apical organ and three distinct ciliary cell cluster with clusters 22 and 25 mostly containing cells of the prototroch and cluster 2 containing cells of the telotroch. I also discovered several muscle and hematopoietic

clusters. Among the cells I identified in the oyster trochophore larva, I believe that neuronal and ciliary cells would be the most interesting to compare across species since they are more likely to present a specific larval signature, that is because the organs they are part of (the ciliary bands and apical organ) are larval specific. Moreover, the ciliary bands of the oyster express a total of 16 lophotrochozoan specific genes that were taken from a recent study (Wu et al, 2020), these will be further dealt with in chapter 6 (see figure 32).
What would be particularly interesting would be to find out whether A) we can find matches across larvae of specific subtypes of cells (for example a specific neuronal subset) and B) to find shared regulatory elements as well as differentiation genes in these cells. In fact, one striking result I obtained from the oyster SCS datasets is a very distinct TFs signature for each cluster which should be easily compared with that of the flatworm larva. The presence of similar regulatory elements in larval specific cell types would likely indicate a common larval evolutionary origin rather than a multiple co-option from the adult.

4.6.3 Peculiarities of the oyster trochophore larva

As discussed in the previous section, the oyster larva presents some classical trochophore features such as ciliary bands and an apical organ, however, it seems to lack (at least at the stage considered) a fully developed gut and protonephridia. Moreover, it is overall considerably smaller than other larvae: oyster larvae are ~50um in diameter compared to the usual 150um of flatworm or sea urchin larvae. The small larval size definitely helped us in achieving a good coverage of all cells, in fact we even recovered some clusters made up of only one or two cells in the larva, and this is because with a dataset of almost 8000 cells (and larvae that only have about 200 cells) we would expect up to 40 repeats of each cell.

Furthermore, the small cell number allowed us to recover many reads and genes per cells, and in fact most clusters not only show a very specific gene but also a distinct transcription factor signature (see figure 53) which is impressive when considering that usually TFs are expressed at a lower level and hence are harder to detect in scRNA-seq datasets.
This smaller size does not necessarily translate in a simpler body system as oyster trochophore appear to have up to 6 haemocytes clusters, 7 distinct muscle clusters and 3 shell gland clusters, many of which have a clearly distinct transcriptional signature; if anything, this study shows how underrated the diversity of cell types in mollusc larvae may be. Specifically, the shell gland clusters, which are a peculiar feature of mollusc larvae, present a high proportion of novel (or fast evolving) genes expressed which may as well prove they originated more recently than the rest of the larva. Indeed, if we believe that phylostratigraphy can accurately infer evolutionary events, the fact that the younger cells of the larva (the shell gland cells) express mostly genes that are mollusc specific (or younger) could mean that the rest of the larva is older. However, as explained before, phylostratigraphy presents some major flaws such as the tendency to misidentify homologs of fast evolving genes and the rather arbitrary decision to use gene expression to weight different genes in calculating the TAI.

Overall, although the oyster larva proved incredibly interesting and complex, in spite of its small size, some of the features we identified appear to be very specific of oysters/bivalve/mollusc larvae and it would have possibly been more interesting to pick a larva with protonephridia or a fully developed gut. In fact, these two organs are often lost during metamorphosis and could present some larval specific cells worth investigating.

5 Single cell sequencing in the polyclad flatworm *Prosthecereaus crozieri*

In the previous chapter I discussed the mapping, data quality, clustering and cell type identification for the trochophore larva of the oyster. In this chapter I will present similar results for the Muller's larva of a polyclad flatworm. This chapter will set the basis for the comparison between the two larvae which is the final aim of my thesis that will be presented in the next chapter.

5.1 Mapping of flatworm SCS data

For the Muller's larva of the polyclad flatworm *Prosthceraeus crozieri* I performed four single cell captures which, similarly to what done with the oyster, I have given the species initials Pc and a chronological number (Pc1-Pc4).

Initially I mapped the single cell data onto a transcriptome that was available in our lab. I obtained between 3000 and 5000 cells per sample with quite low median genes per cells (between 40 (Pc 2) and 400 (Pc 4) compared to an overall average of 1000 for the oyster scs) (see fig. 57). When looking at the mapping percentages I found that a strikingly high 11-15% of reads mapped antisense to the transcriptome. This is a common problem when using non-stranded RNAseq data to build a transcriptome and it is caused by transcripts getting wrongly assembled antisense. Several other members of the EVOCELL network I am part of also encountered this issue. As explained in the previous chapter for intergenic regions,

antisense mappers are automatically discarded, which could partially explain the low number of genes per cells recovered.

For this reason and because overall mapping seemed to be quite low (<72%) I decided to sequence the genome. DNA preparation as well as assembly and annotation were carried out by Daniel Leite (details in methods).

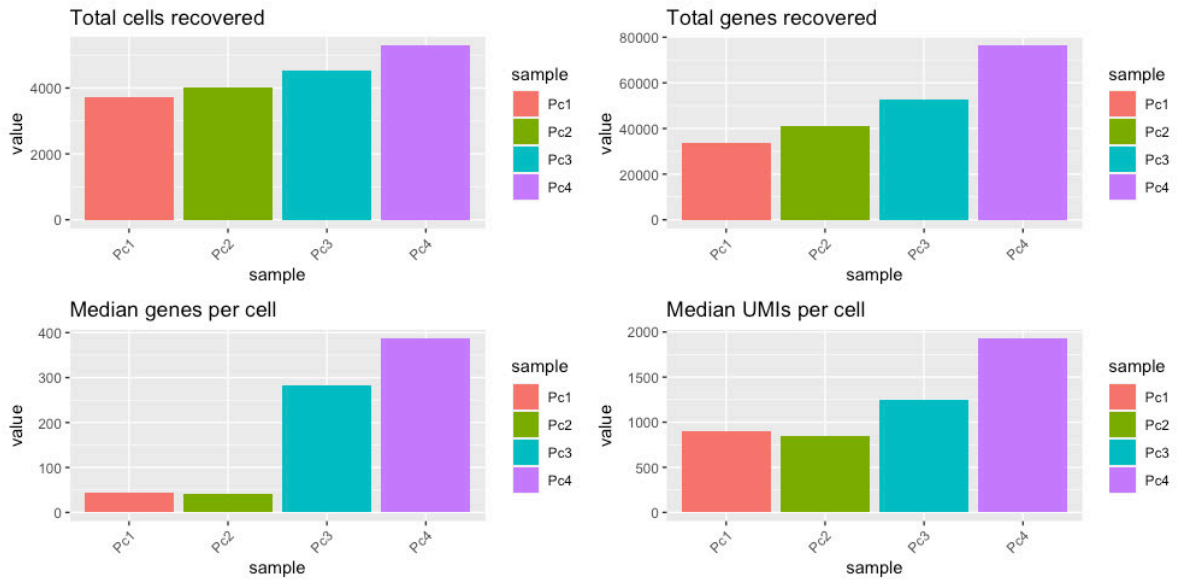
The final genome assembly generated was 2.07 Gb in size with an N50 of ~30 kb and a total of 43,325 predicted genes. 886/978 (90.6%) BUSCO genes of *P. crozieri* were identified, and of these 767 (78.4%) were present as a single copy.

After remapping the data onto our genome, I found several positive shifts; the overall percentage of reads mapping to the genome increased (>87%), the percentage of antisense mappers dropped (<3%) and the number of genes, reads and UMIs per cell all increased. The overall number of genes detected decreased, however, and this may be explained by the considerably lower number of genes predicted in the genome (43,325) compared to the transcriptome (216,151). This higher number of genes in the transcriptome, however, was almost certainly due to many redundant transcripts since I actually see an overall increase in mapping (see figure 58).

I also observed a slight decrease in cell numbers (for Pc1, for example, from ~4000 to 3000) which could be due to the loss of some borderline cells - droplets that were previously considered cells and are now considered empty - which could be due to the overall increase of UMIs per cell (see figure 59). The software I used for mapping and assigning barcodes to

cells (Cell Ranger) uses the UMI content per barcode (i.e. per droplet) to decide what is a cell and what is not. It specifically looks for a significant drop in the plot of numbers of UMIs indicating the stark difference expected between full and empty droplets. It is possible that this drop would be more obvious after recovering more reads (and more UMIs per cell) and hence that some empty droplets at the border would get discarded (see figure 60).

Finally, the optimized mapping clearly shows that, once again, I have some quality differences between my samples. Samples Pc3 and Pc4, in particular, recover more cells, have more genes as well as higher median numbers of genes and UMIs per cell. In the next paragraphs I will discuss the differences between the samples and how I performed quality controls on them (see figure 58 and 59).



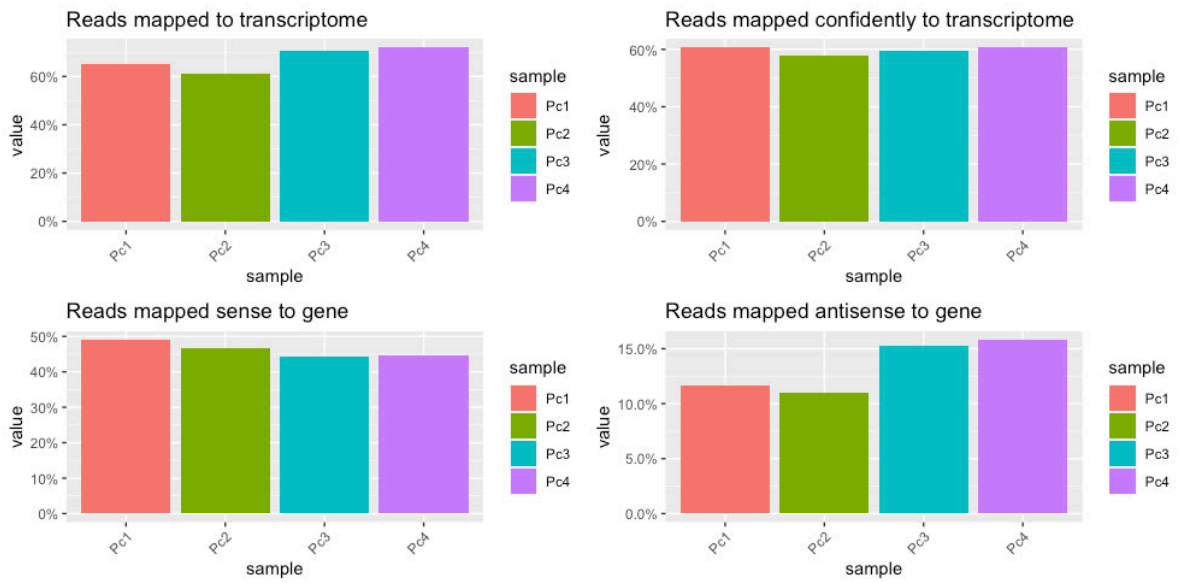


Figure 57. Very few reads from scRNA seq map to the flatworm transcriptome.

Mapping single cell sequence read data from the Müller's larva of the polyclad flatworm to a transcriptome gives between 3000-5000 cells per sample with a low gene content (50-400). Overall mapping is fairly low (around 60%) with up to 15% of reads mapping antisense to transcripts, a common problem when building transcriptomes from non-stranded RNA-seq data.

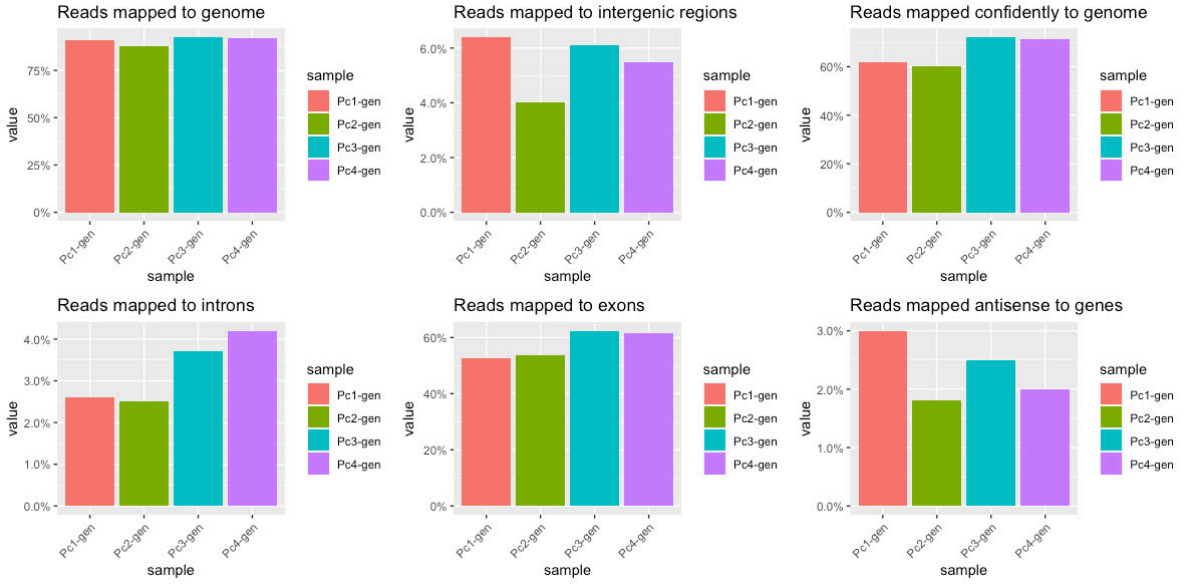


Figure 58. Mapping single cell data from the Müller's larva of the polyclad flatworm to our newly sequenced genome improves overall mapping (>87%) and drastically reduces antisense mappers (<3%) giving us more useful data for cell type analysis.

Pc-gen indicates mapping to the genome.

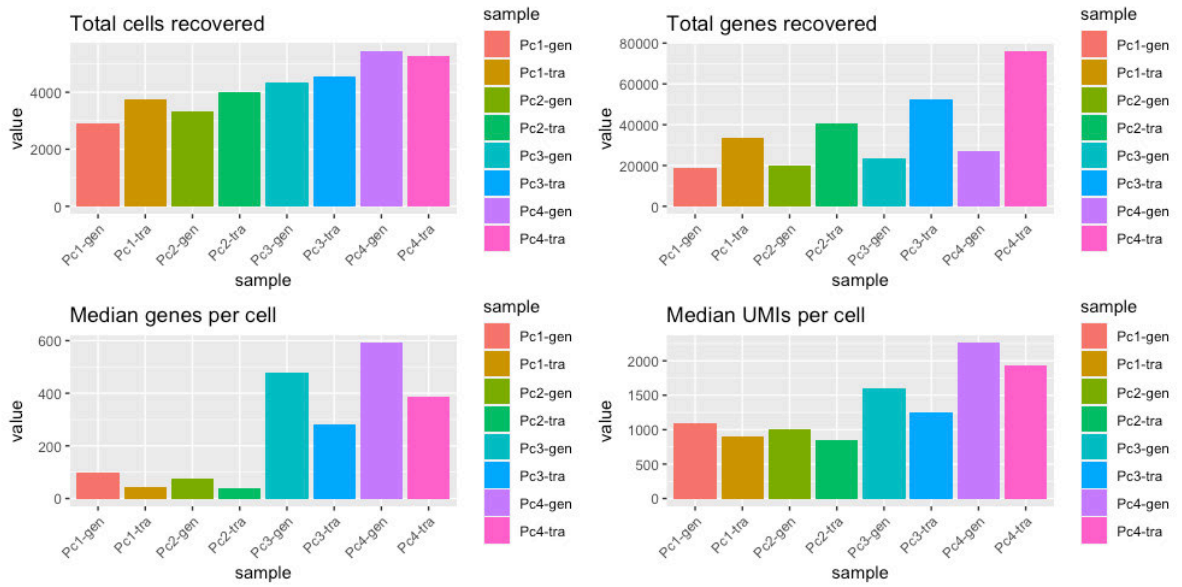


Figure 59. Mapping single cell data from the Muller's larva of the polyclad flatworm to our newly sequenced genome improves median numbers of genes and UMIs per cells, however we also see a reduction in cell numbers as well as total genes.

Pc-gen indicates mapping to the new genome, Pc-tra indicates mapping to the old transcriptome.

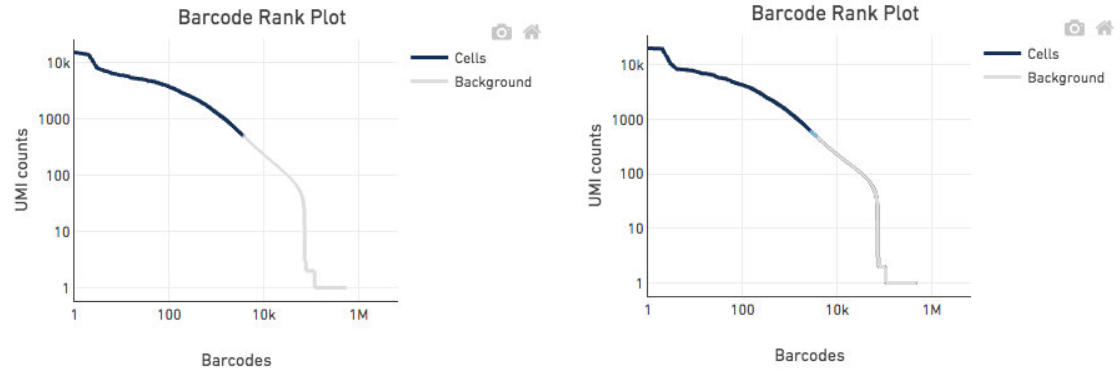


Figure 60. Higher recovery of UMI per cells helps correct identification of cell containing droplet vs background.

UMI rank plots for Pc1 when data is mapped to transcriptome (left) or genome (right) showing a more visible dip in the curve (commonly called knee) when mapping genes to the genome. This likely caused a drop in cell number since empty droplets could be more correctly detected.

5.2 Quality control of different samples

As explained in detail in the Methods section, for the flatworm larva I performed two cell dissociations experiments and from each I loaded two samples into separate wells of the microfluidics chip with different salt concentration for the final elutions. Immediately before capture, Pc1 and Pc3 were eluted in CMF-SW whilst Pc2 and Pc4 were eluted in a mix of NoCaNoMg-EDTA-free-ASW and nuclease free water (so that the final salt concentration was below 250 mM which is the highest NaCl limit tested by 10x). I did this because I was worried that the higher salt concentration of cells eluted in pure NoCaNoMg-EDTA-free-ASW (higher than the manufactures' tested limit) could impede the activity of the retro-transcriptase enzyme and/or that a lower salt concentration (for cells eluted in NoCaNoMg-EDTA-free-ASW and nuclease free water) could reduce cell viability. As shown in figure 59, there is a clear difference in quality between the two dissociation experiments (Pc1-Pc2 and Pc3-Pc4), with the second repeat yielding:

1) higher overall cell numbers
2) higher genes per cell
3) higher UMI per cell

However, in comparison, the differences between the two conditions, NoCaNoMg-EDTA-free-ASW (Pc1, Pc3) vs NoCaNoMg-EDTA-free-ASW+NFH2O (Pc2, Pc4) are not as striking, although samples eluted in NoCaNoMg-EDTA-free-ASW+NFH2O appear to have higher mitochondrial gene content which could have been caused by osmotic stress (see figure 61). There doesn't seem to be, however, a clear effect on genes or UMI recovered although with only two replicates this is not a strong conclusion.

Regardless of batch specific differences, all samples show a relatively higher mitochondrial gene content than I observed with the oyster, with most cells having mitochondrial gene content higher than 10%. For mitochondrial gene content, samples Pc1 and Pc2 show two density peaks, one at around 20-30% and one at around 50-60%. It is very likely that the higher peak corresponds to dying cells, in fact, as one can see in figure 62, most cells with high UMI content have less than 30% mitochondrial gene content. It is unclear why the flatworm larval cells I captured have an overall higher mitochondrial gene content than those of the oyster larva, it could be due to problems with cell viability. However, as I will show more in detail in the next paragraph, these cells seem still to contain a lot of information, and this could indicate that higher numbers of mitochondrial transcripts could represent the natural state of these cells and be a species-specific trait.

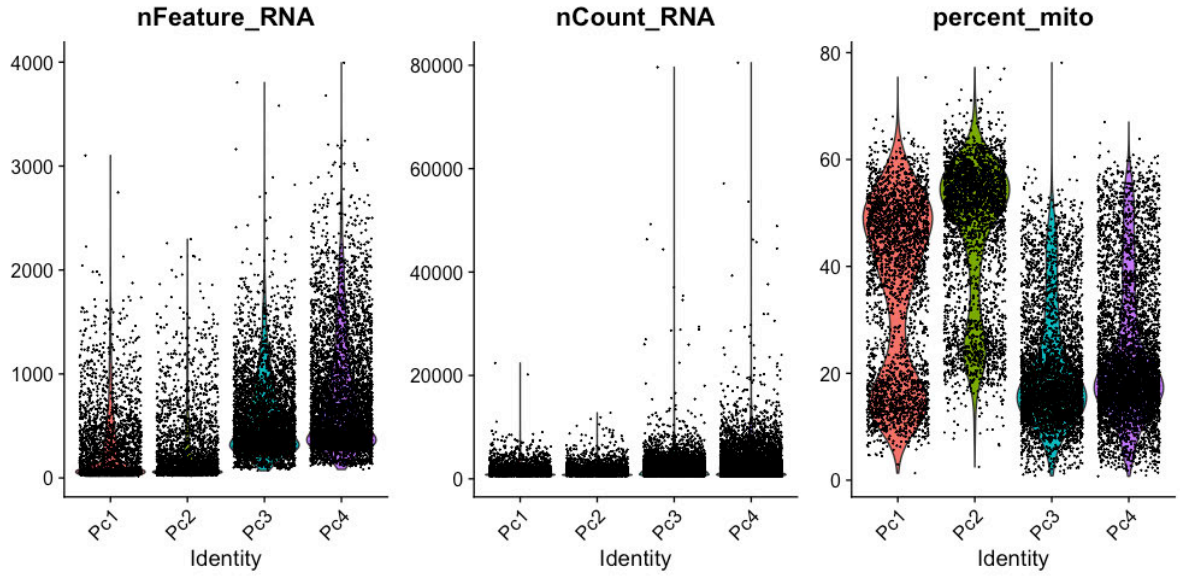


Figure 61. Samples Pc3 and Pc4 show overall better quality than samples Pc1 and Pc2.

Violin plots of gene numbers per cell, UMIs per cell and percentage of reads mapping to mitochondrial transcripts per cell of different samples. This shows that samples Pc3 and Pc4 have higher overall quality (gene numbers per cell, UMIs per cell) then Pc1 and Pc2. Moreover, samples eluted in NoCaNoMg-EDTA-free-ASW (Pc1, Pc3) have lower mitochondrial gene content (a sign of better cell viability) than those eluted in NoCaNoMg-EDTA-free-ASW+NFH2O (Pc2, Pc4) possibly due to osmotic stress. Elution in water doesn't seem to affect the other parameters much.

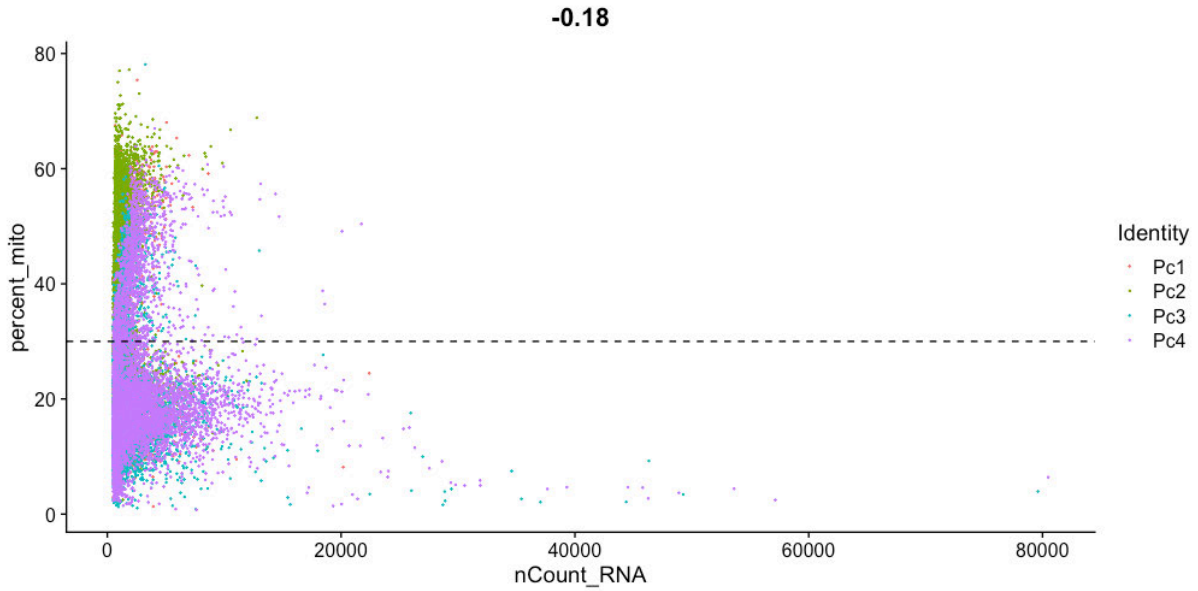


Figure 62. Cells with low mitochondrial gene content have higher quality.

Scatter plot showing that cells with low mitochondrial gene percentage have a higher UMI number.

5.2.1 Filtering out low quality cells

As a first quality control measure, I decided (as I did for the oyster) only to keep cells expressing at least 200 genes; this is a common procedure to get rid of low-quality cells. As one can see in fig. 63, simply filtering out cells with fewer than 200 genes helped remove most cells with high mitochondrial gene content.

Nonetheless, since the base mitochondrial gene content in flatworm cells is higher than in my oyster samples I decided to try and select subsets of cells according to their mitochondrial gene content to see if and how this impacted cell clustering. My idea was that, above a certain threshold of mitochondrial gene content, cells will start to lose their

transcriptional signature which should result in a loss of clustering structure. Once I can identify this threshold, I can use it to filter out cells for downstream analysis. Figure 64 shows the differences in gene number, UMI number and mitochondrial gene content for the different subsets selected. Each subset contains the same number of cells, 1142, except for the highest mitochondrial content subset which contains 1146. As one can observe in figure 64, it appears that gene numbers per cell and UMI numbers per cells start to reduce from group 8 onwards, which corresponds to cells with more than 30% mitochondrial gene content.

I then clustered these different cell subsets to see how the overall structure of the data would change. Clustering was done with the same steps as described for the oyster in chapter 4. Figure 65 shows that pretty much all subsets appear to contain meaningful clustering (i.e. cells divide out nicely into approximately 10 clusters), this result reassured me of the quality of my data. Since there didn't seem to be a clear-cut difference between the various subsets, I decided to set the maximum cut-off for mitochondrial gene content threshold to 40% for all downstream analysis. I should mention that playing around with the threshold (lowering to 30% or increasing it to 50%) didn't cause any big difference in the final clustering (results not shown).

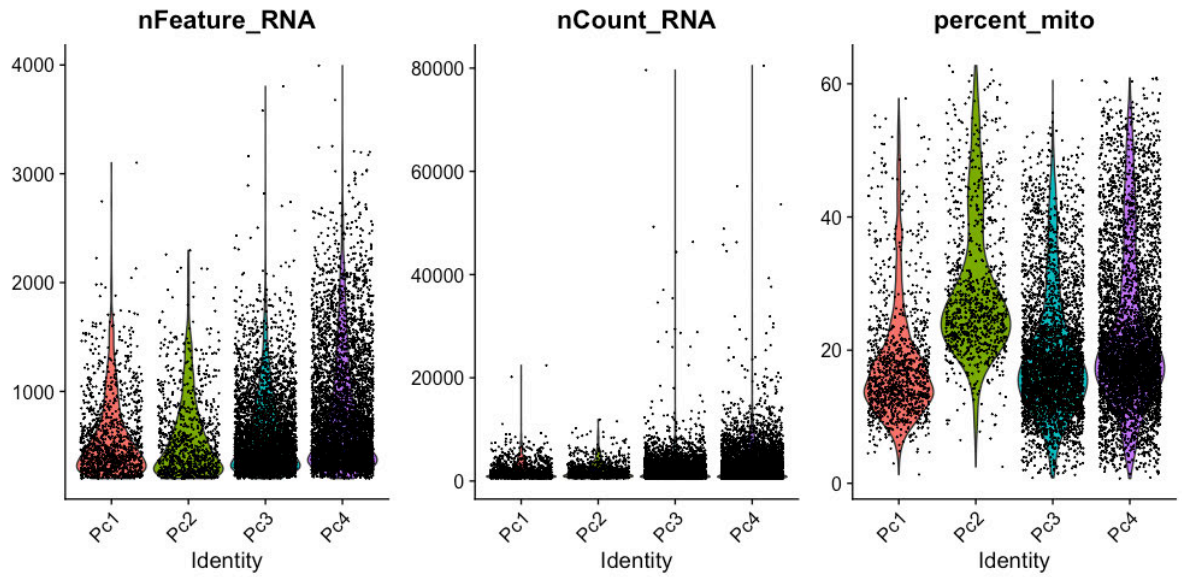


Figure 63. Filtering out cells with fewer than 200 genes removes most cells with high mitochondrial gene content.

Violin plot showing the number of genes per cell, UMI per cell and mitochondrial gene percentage per cell after removing all cells expressing less than 200 genes.

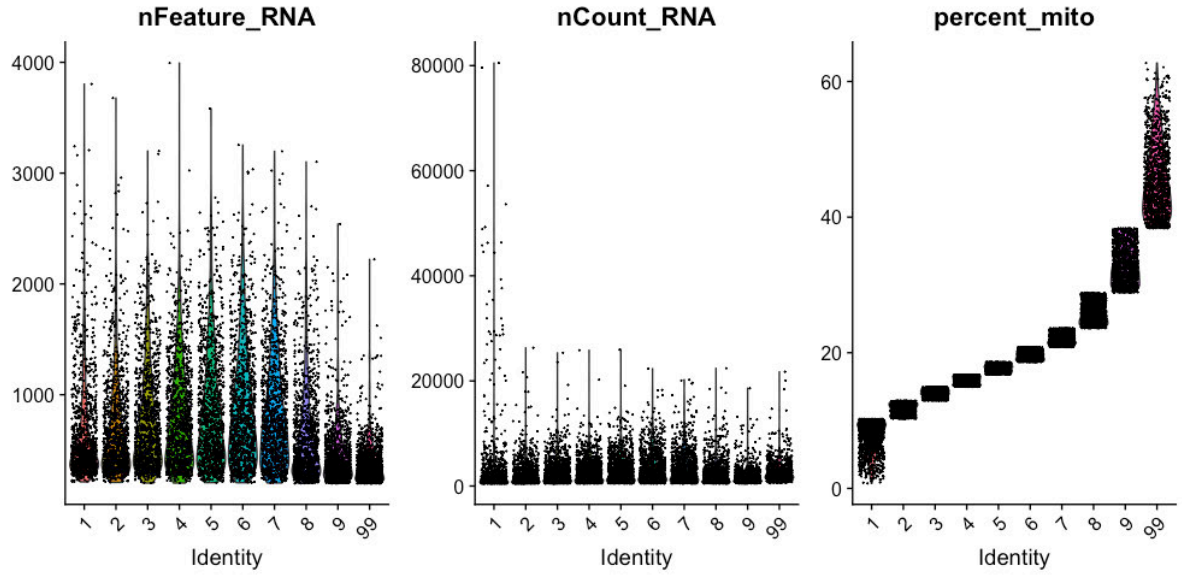


Figure 64. Cells that have up to 30% of mitochondrial gene content have similar quality.

Violin plots showing the number of genes per cell, UMI per cell and mitochondrial gene percentage per cell cells subsets in groups of 1142 cells ordered by their mitochondrial gene content.

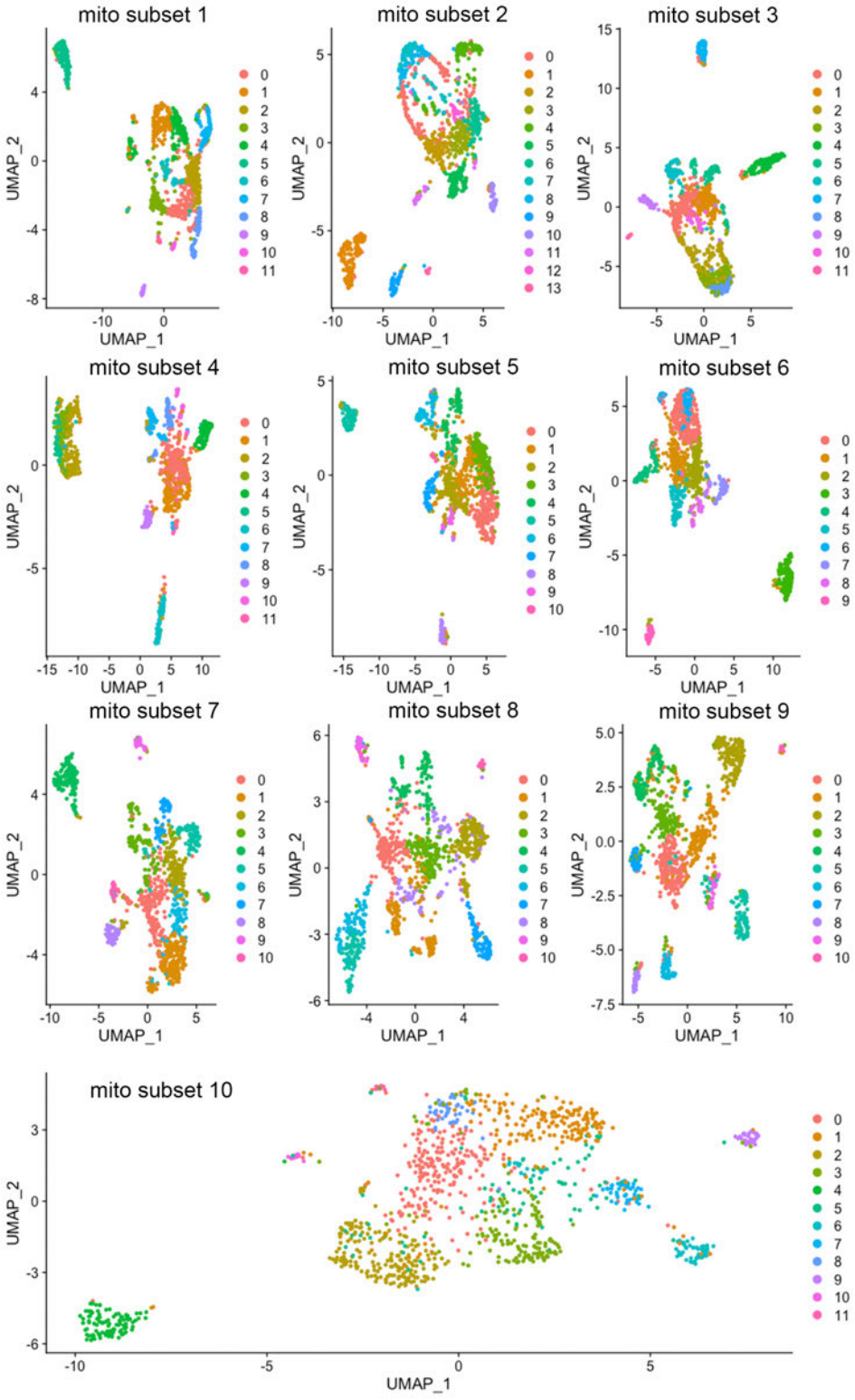


Figure 65. UMAP of cells subset by mitochondrial gene content showing most subsets retain a lot of information.

The mitochondrial subsets are the same as shown in figure 7, Mito subset 10 image is larger just to fill the space.

5.2.2 Sample aggregation

Differently from what I had observed with the oyster, where different samples would not cluster together unless coerced by integration, flatworm samples appear to cluster together using cell ranger aggregation (without normalization) (see fig. 66). This is probably because, although there is a slight difference in quality among samples, this is not as substantial as it was for the oyster. To ensure that only higher-quality cells were used for clustering, as discussed above, I removed all cells with more than 40% mitochondrial gene content, I set the minimum number of genes per cell to 200 genes and regressed out UMI number as well as mitochondrial gene content to make sure they did not influence the clustering. As one can see in figure 67, these steps allowed me to remove any bias in clustering that could be due to difference in quality among cells.

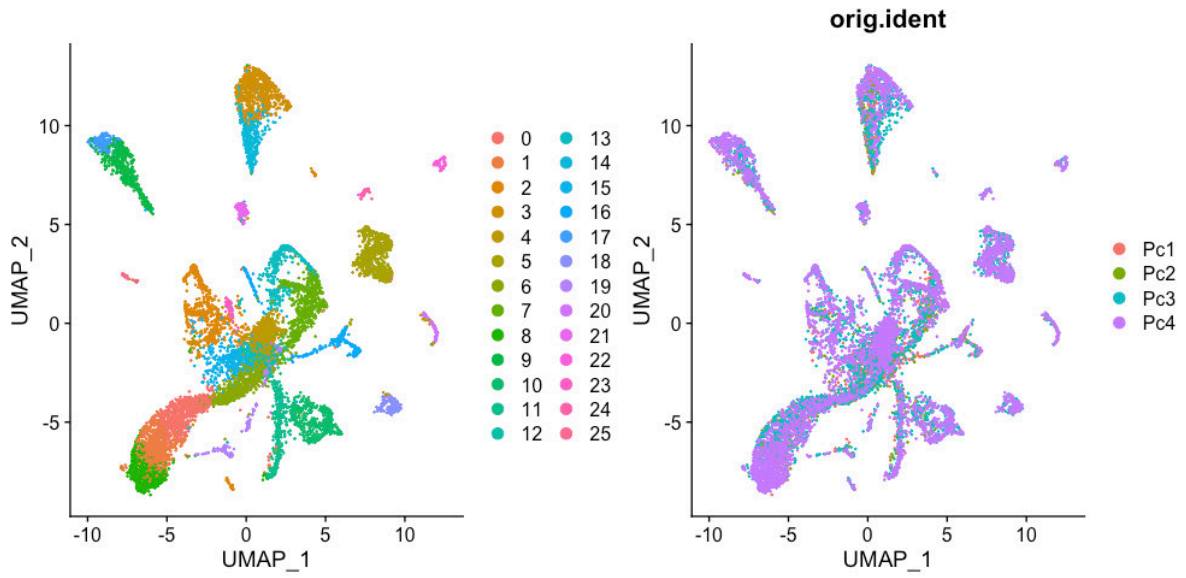


Figure 66. Cells from different samples cluster together without the need to integrate.

UMAP coloured by cluster (left) and by sample identity (right) showing that Cell Ranger sample aggregation is enough for cells from different samples to cluster together.

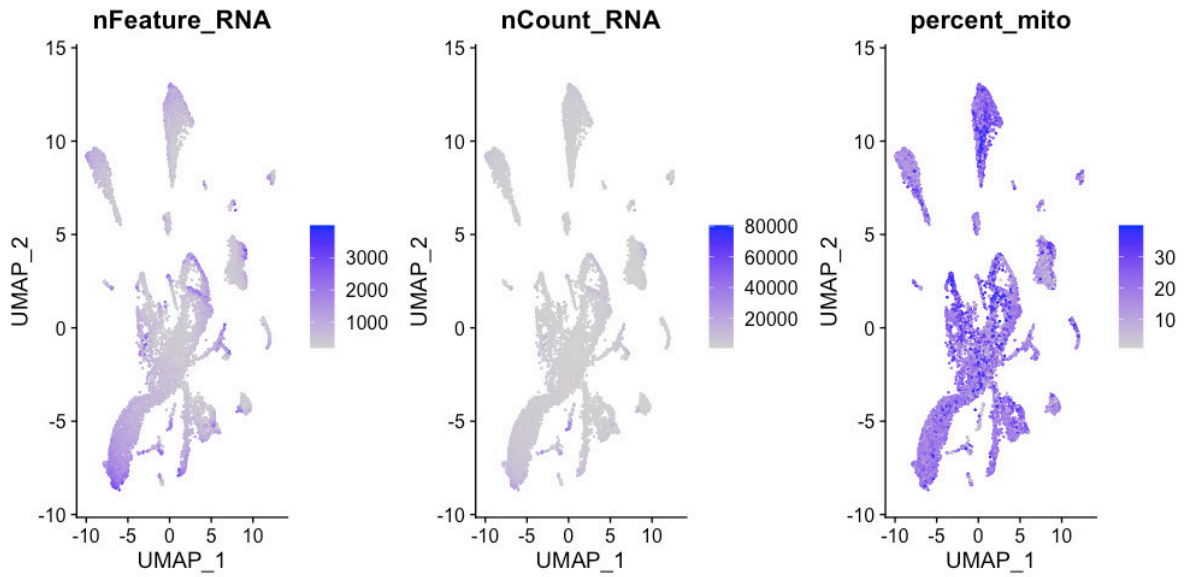


Figure 67. UMAP coloured by gene content (left), UMI content (centre) and mitochondrial percentage (right) showing that clustering is not driven by differences in these parameters.

5.3 Re-sequencing of samples Pc3 and Pc4

To improve the quality of our flatworm larval data I decided to re-sequence our two best samples (Pc3 and Pc4) in greater depth, and to use these for downstream clustering analysis as well as for comparison with the oyster. As mentioned in chapter 3 I believe it is best to only use high quality data for cross species comparison to be sure that similarity is not driven by artefacts (such as low read coverage per cell, high mitochondrial gene content or markers of cell death). After re-sequencing and quality filtering (minimum 200 genes per cells and mitochondrial gene content < 30%) Pc3 and Pc4 contain a total of 13,457 cells with median of 418 genes per cell, median of 1377 UMIs per cells and a total of 35,730 genes recovered. The threshold for mitochondrial gene content was lowered to 30% since on average Pc3 and Pc4 had fewer cells with high mitochondrial gene content than Pc1 and Pc2. Clustering following the same procedure outlined for the oyster in chapter 3 (and using

a resolution parameter of 2) gives the UMAP shown in figure 68. I obtained 34 clusters, all of which contain cells from both repeats. In the rest of this chapter I will try, as previously done for the oyster larva, to identify the main cell types and to confirm their localisation using *in situ* hybridisation techniques.

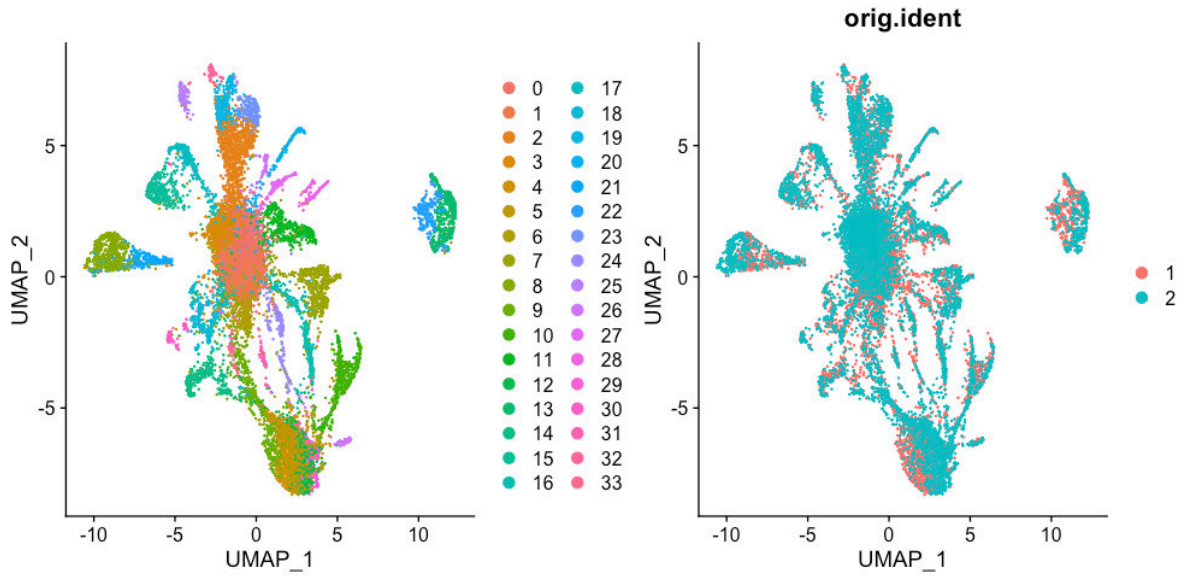


Figure 68. UMAP of re-sequenced Pc3 and Pc4 shows 34 clusters (left) all of which contain cells from both repeats (right).

In the UMAP on the right 1=Pc3 and 2=Pc4.

5.4 Developing a new and successful *in situ* hybridisation protocol for the flatworm Mueller's larva

One of the main challenges of my PhD thesis has been trying to develop a functioning *in situ* hybridisation method for the flatworm larva to use to identify the expression of my single cell sequencing data. This has been attempted for the past 10 years in various labs including ours without much success. We were initially hoping that scRNAseq data would give us a good idea of which genes were expressed at the larval stage as well as their per cell expression level and that that information could help us with the troubleshooting. However, with classical *in situ* hybridisation protocols, only a small handful of genes have ever

worked. What was most puzzling was that A) a gene that was working would work regardless of the protocol used and B) there seemed to be no consistent difference between working and non-working genes. Genes for which *in situ*s worked were not more highly expressed and they did not have a different AT/CG content from those that did not work (see table 2 for details). I initially thought that the size of the probe might play a role, since the three genes for which *in situ*s had worked were shorter (although not by much) than the non-working ones so I shortened some probes but that did not help. One of the genes that always worked well is r-opsin (Rawlinson et al, 2019), we hypothesized that this could be due to the superficial position of r-opsin positive cells on the outside of the larva. However, genes that were fished out of our scRNA-seq data as expressed in the same cells as r-opsin did not work. Moreover, another gene, troponin T, which is clearly expressed in cells that are deeper in the larva appeared to work (see fig. 69).

I eventually decided to try *in situ* DNA-hybridization chain reaction (HCR). I found this new protocol gave consistently better results with more than half of the genes so far trialled showing results consistent with our single cell data (Choi et al, 2018). In this chapter and the following ones, I will only show HCR results for the flatworm larva. As a proof of concept, I showed that HCR and chromogenic *in situ*s for the two genes r-opsin and troponin T gave the same results (see fig. 69). The fact that HCR gives consistent results in the polyclad flatworm larva is a great tool that came out of this work, and we hope it will open the door to more exciting research on the Mueller's larva.

Table 2. There are no clear differences between working (green gene names) and non-working (red gene name) genes used for classical chromogenic *in situ* hybridisation.

Mean counts per cell: mean counts per cell calculated using our scRNA-seq data. Cell expressing: indicate the number of cells that express the gene in our scRNA-seq data.

Gene ID	Gene Name	% GC	Size (bp)	Mean counts per cell	cells expressing	T7/ Sp6
g25197	Pc_pero_opsin	50.9	712	2.5	83	sp6
g13280	Pc_5HT	42.4	1008	4.862069	29	t7
g20284	Pc_mucin	44.1	1461	26.22689	119	sp6
g21369	Pc_jumonjii_C	43.2	1035	6.407407	27	sp6
g491	Pc_FRMFamide_receptor	38.1	1039	1.908163	98	t7
g3137	Pc_metalloproteinaseNAS13like	42.4	1042	5.528662	157	t7
g3163	Pc_Vasa	45.2	1016	1.888845	2519	t7
g24247	Pc_GuanineNucleotideBinding	47.5	809	1.487805	41	sp6
g18600	Pc_ropsin_Kate	52.3	523	4.242424	33	sp6
g14959	Pc_troponin_T	48.7	649	4.927361	413	t7
g896	Pc_Foxq2	54.1	554	1.520408	98	sp6

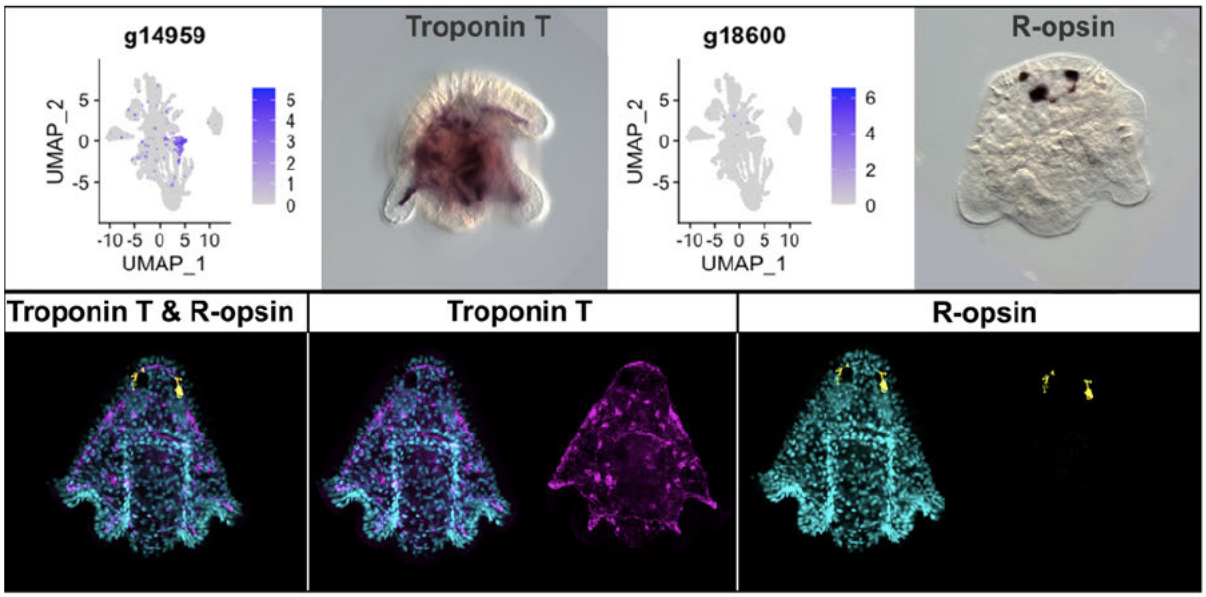


Figure 69. Chromogenic *in situs* (top) and fluorescent HCR (bottom) have the same expression patterns for R-opsin (yellow) and Troponin T genes (pink).

Top row: UMAP showing the expression of the genes (troponin T and R-opsin) in scRNAseq data next to chromogenic *in situ*s. Bottom row shows HCR of Troponin T (pink) and R-opin (yellow) with and without DAPI (in light blue) which stains the nuclei.

5.5 Cluster identification in the flatworm Mueller's larva

Since *in situ* hybridisation in the flatworm larva never reliably worked before I didn't have any species-specific results from previous work that I could use to identify cell types in our SCS data. However, luckily, I could use gene markers from other SCS datasets from adult flatworms similar to what I did for the neoblasts of the oyster. As described in the previous chapter, I first identified orthologous genes between our polyclad flatworm and the flatworm *Schmidtea mediterranea* using Orthofinder (Emms & Kelly, 2019) and then collected all genes of the flatworm that are homologs to cluster markers of different *S. mediterranea* cell types (taken from Plass et al, 2018).

5.5.1 Muscle cluster

Similar to what I have done for the oyster, I started off by looking for myocytes. I extracted the *S. mediterranea* myocyte marker genes and found their homologs in the flatworm. Their expression across cell clusters can be seen in fig. 70 which shows that most of *P. crozieri* homologs of *Schmidtea* myocyte markers are expressed in cluster 7. Moreover, one can appreciate how most genes are indeed myocyte related (several myosins, tropomyosins, paramyosins and collagens). Interestingly, it appears that the flatworm larva only has one myocyte cluster, which is quite different from what I had observed for the oyster larva. As expected, troponin T, one of the few genes for which I had working chromogenic *in situs*, was indeed a myocyte specific gene. Troponin HCR expression matches that of troponin-

chromogenic *in situ* and also resembles previous immunohistochemical staining (Rawlinson, 2010) (see fig. 69 and 71).

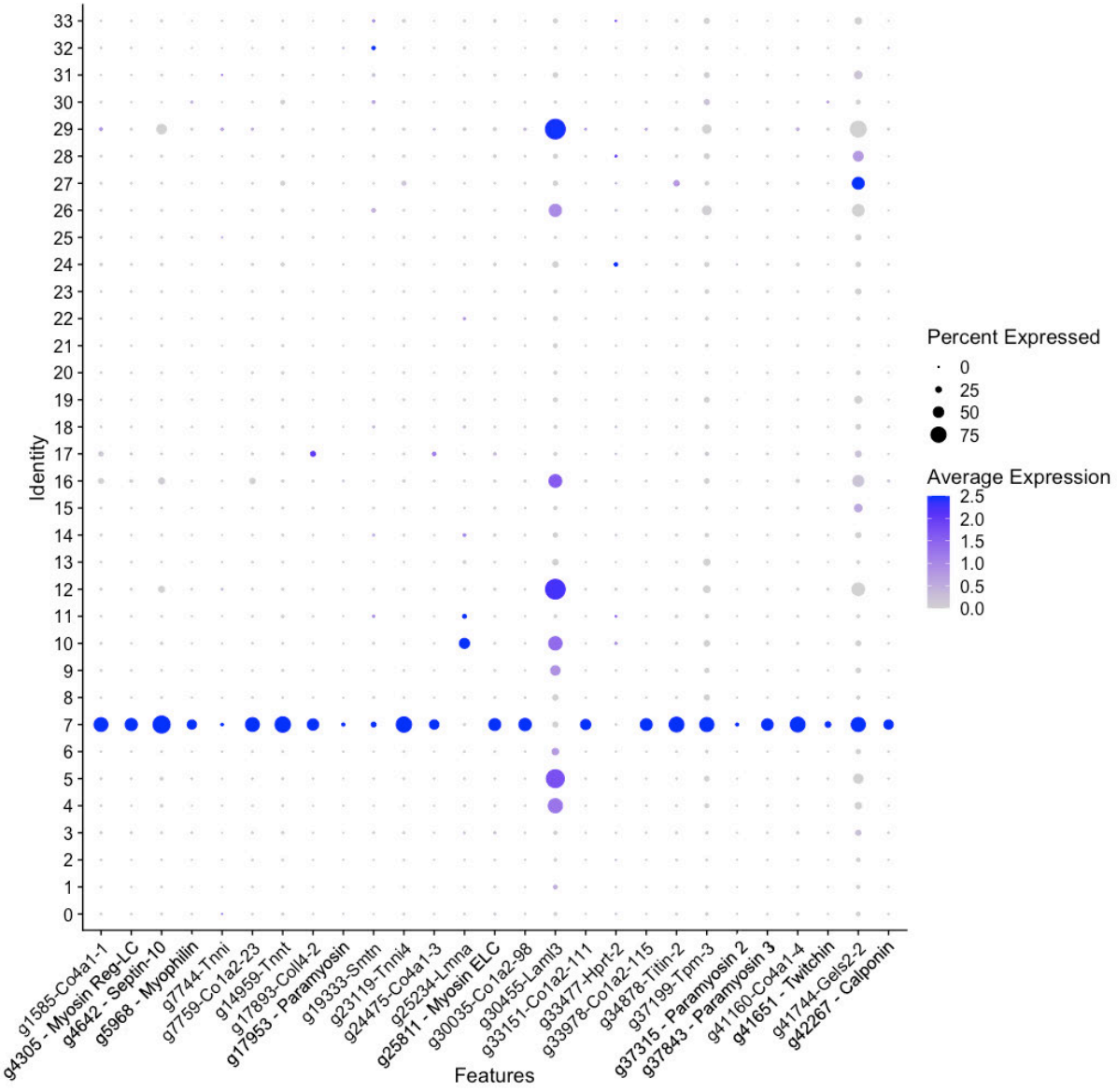


Figure 70. *P. crozieri* orthologs of *S. mediterranea* myocyte markers show expression in cluster 7.

Dotplot of orthologs of myocyte marker genes from the Plass et al paper (2018).

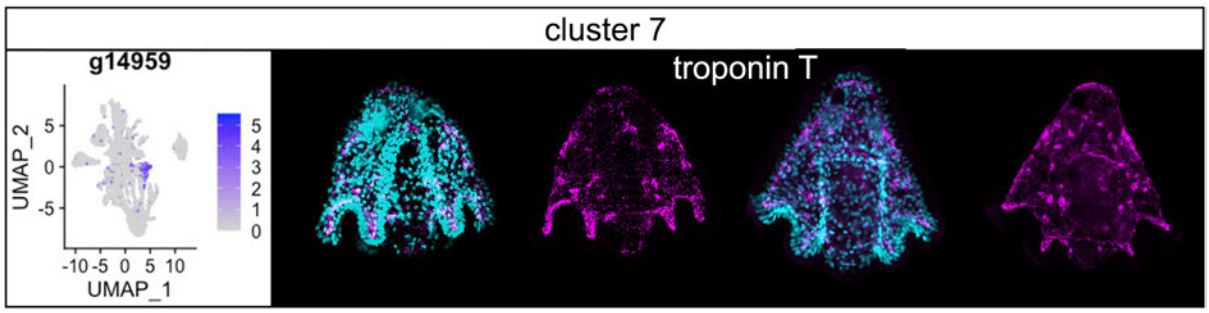


Figure 71. *P. crozieri* muller's larvae HCR of the cluster 7 myocyte marker Troponin (here in magenta) highlights the myocyte in the larva.

Expression resembles that shown with immunohistochemistry by Rawlinson (2010). Small UMAP on the left shows expression of the gene in scRNA-seq data next to HCR stainings in two animals (fist one, dorsal view, second one frontal view). DAPI in light blue stains the nuclei.

5.5.2 Neoblast clusters

Neoblasts are a population of stem cells found in flatworms that allow adult worms to regenerate almost all body parts. Previous studies from our lab using EdU pulse experiments on *P. crozieri* Muller's larvae had shown the presence of proliferating cells in the mesenchymal layer, similarly to what has been observed in adults of the macrostomid flatworm *M. lignano*. Since neoblasts are very well characterised in planarians, I decided once again to look at the expression of orthologs of *S. mediterranea* neoblast markers. Figure 72 shows that clusters 4, 5, 12 and 29 are likely to represent neoblasts or undifferentiated proliferating cells. I selected a common marker for these clusters and found that its expression is found in the mesenchymal layer of the larvae, similarly to that already observed in our lab for proliferating cells marked with EdU (see figure 73).

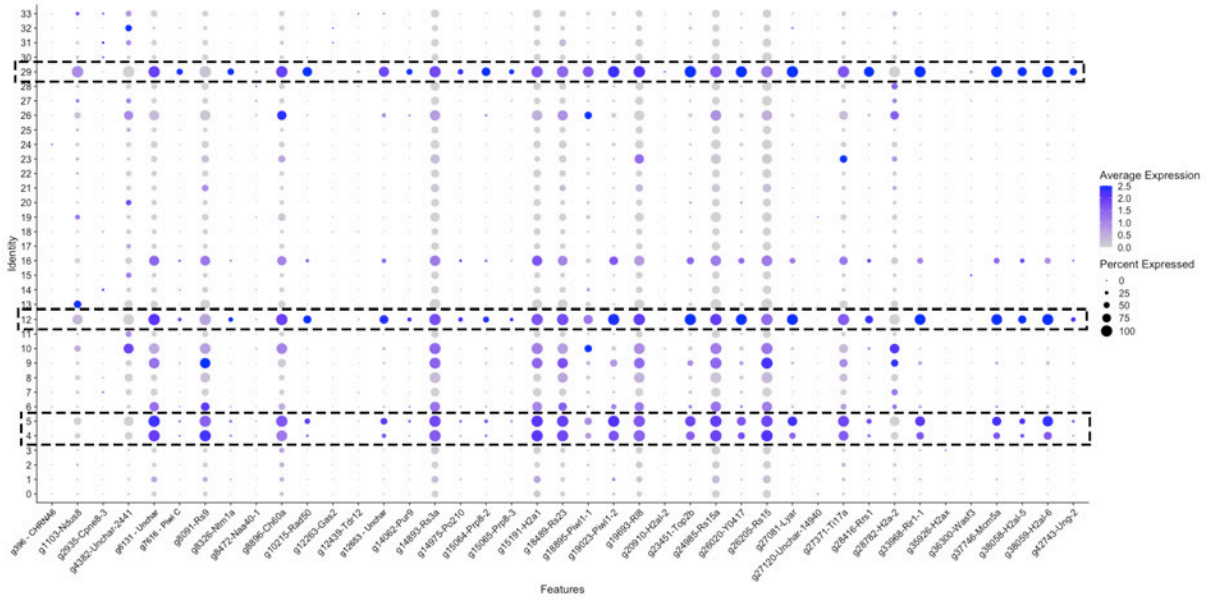


Figure 72. *P. crozieri* orthologs of *S. mediterranea* neoblast markers show expression in cluster 4, 5, 12 and 29

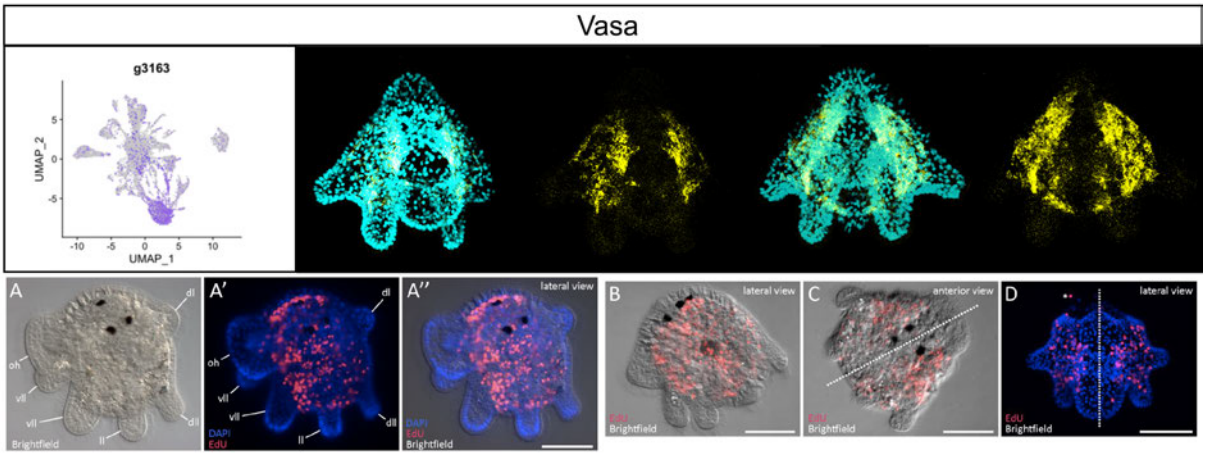


Figure 73. HCR of the cluster marker for clusters 4, 5, 12 and 29 Vasa (here in yellow) shows expression in the mesenchymal layer of the larva, similarly to what previously observed for proliferating cells labelled with EdU in our lab.

Top box shows on the left the UMAP of expression of vasa in the scRNA seq data and on the right two animals stained with HCR for the Vasa gene (first one frontal view, second one dorsal view). DAPI in light blue stains the nuclei. Bottom box shows Edu experiments carried out by Johannes Girstmeir. Larvae were labeled for S-phase cells with a 1-hour pulse of 10 µM EdU. (A) Brightfield image showing the lateral view of a Müller's larvae. (A') Same larvae visualised for DAPI staining and EdU positive cells (A'') merge of all three channels. (B) Further examples of an EdU stained Müller's larva showing the same result. (C-D) apical view of a Müller's larvae showing the bilateral distribution of EdU positive cells. (oh) oral hood, (dl) dorsal lobe, (vll) ventro-lateral lobe, (ll) lateral lobe, (dll) dorso-lateral lobe. Scale bar = 100 µm.

5.5.3 Neuronal clusters

As for the oyster I decided to use neuropeptides to identify the neuronal clusters of the flatworm. I asked our collaborator Luis Yañez Guerra (Jékely lab) to predict the neuropeptides in the genome of the flatworm and then I looked at their expression in our single cell dataset. As one can see in figure 74, most neuropeptides show expression in clusters 18 and 30. HCR *in situs* of markers for cluster 30 show it is concentrated in two groups of cells in the apical region of the larva, whilst a more general marker for both cluster 18 and 30 shows a few neurons scattered around the animal and a larger apical structure (likely the apical organ) (see figure 75). Since cluster 30 appears in two patches in the apical region of the larva, I wondered if these cells had any connection to the paired cerebral eyes of the larva. (The flatworm larva after hatching has three eyes: two cerebral eyes (expressing r-opsin, shown in figure 69) and one epidermal eye (Rawlinson et al, 2019). On top of working well in the flatworms, HCR allowed me to look at up to three genes co-expression, so I decided to look at the expression of the cluster 30 markers together with the cerebral eye marker R-opsin. Figure 76 shows that cells from cluster 30 indeed seem to be in close contact with R-opsin positive cells of the cerebral eyes.

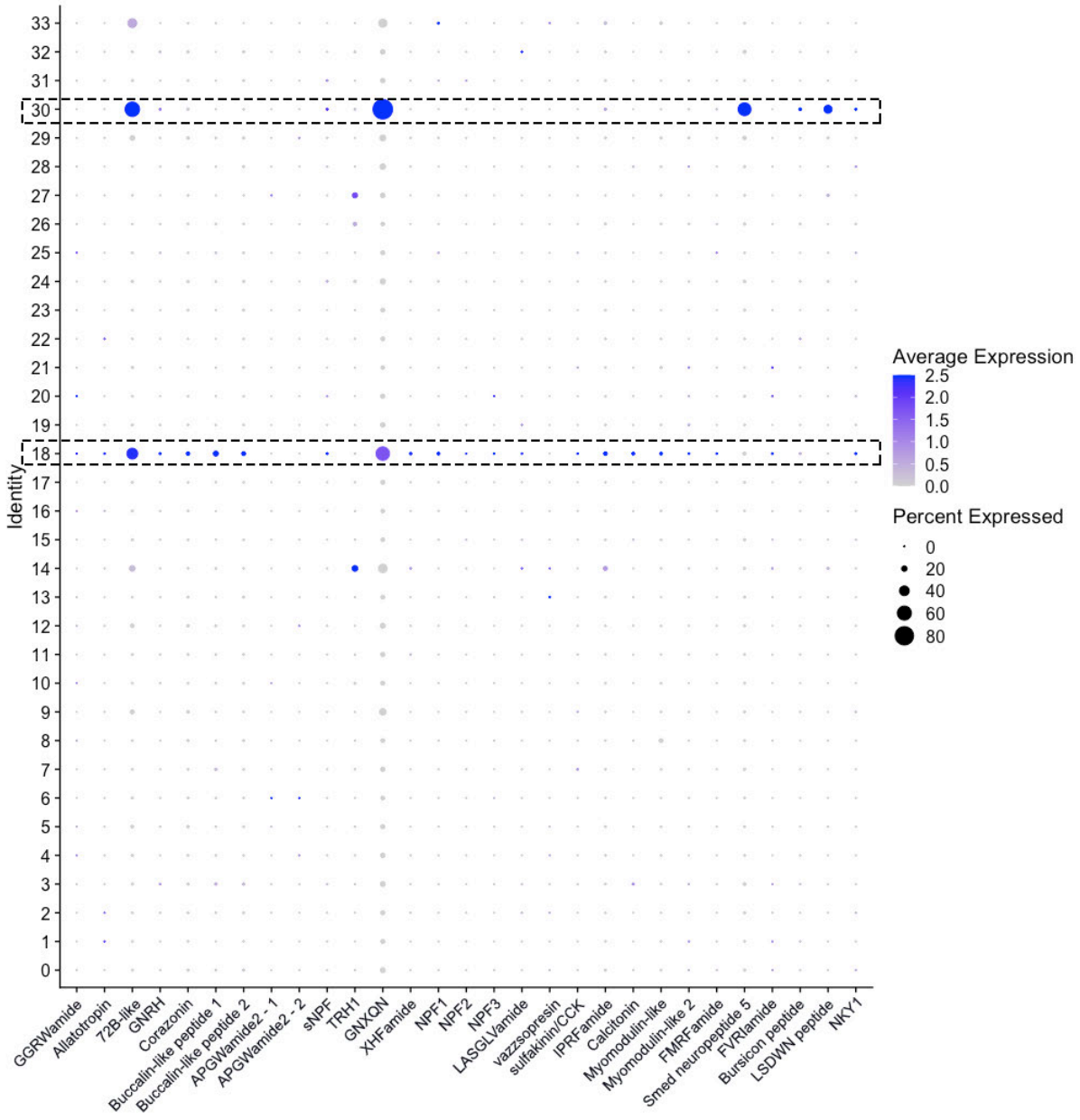


Figure 74. Most neuropeptides show expression in cluster 18 and 30.

Dotplot showing expression of *P. crozieri* predicted neuropeptides in different single cell clusters.

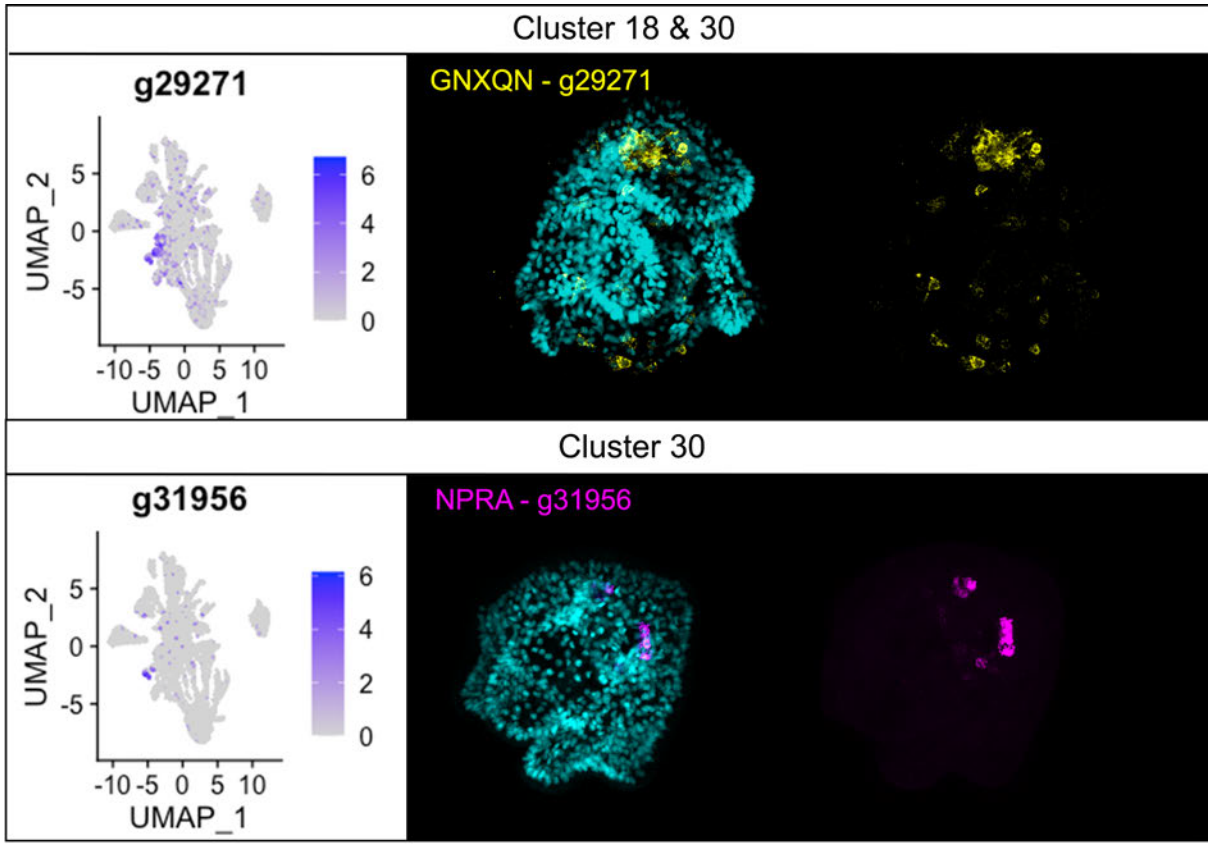


Figure 75. General neuronal marker GNXQN (yellow) shows expression in the apical organ and some scattered neurons.

NPRA (pink), specific marker for cluster 30, shows expression in two patches near the cerebral eyes of the larva. Small UMAPs show expression of the genes in scRNA seq data. DAPI in light blue stains the nuclei.

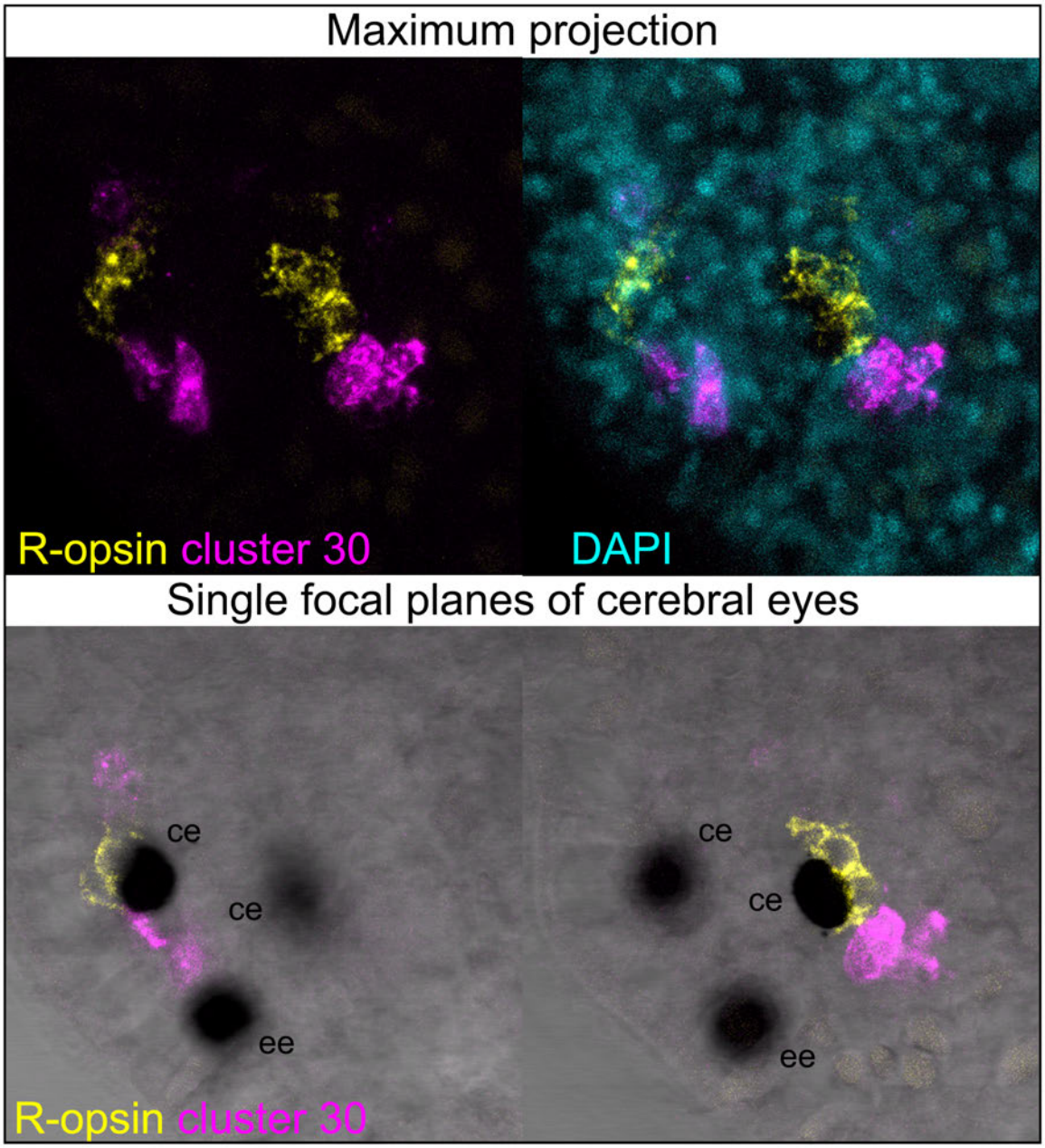


Figure 76. Neuronal cells from cluster 30 (magenta) appear connected to R-opsin positive cells (yellow) of the cerebral eyes (ce).

Gene marker for cluster 30 is NPRA-g31956, ce= cerebral eye, ee=epidermal eye. Yellow= R-opsin, Magenta= NPRA-g31956, Cyan=DAPI. Bottom figures were imaged with brightfield to show the eye pigments.

5.5.4 Protonephridia

A common structure found in most Lophotrochozoan larvae are protonephridia which are the larval excretory organs. They have been previously described in Mueller's larvae and for this reason I was interested in identifying them. When I tried to identify protonephridia using orthologs of markers from adult *S. mediterranea* I didn't get a very clear result (see left of figure 77, light red). However, a few specific protonephridial markers (POU2/3, Hunchback and Six1/2-2) seemed to be consistently expressed in cluster 28 (see right of figure 77, light blue) (Scimone et al, 2011). The cluster marker g10760 clearly highlighted the protonephridia confirming that cluster 28 indeed contains protonephridia (see figure 78). This result was exciting as it meant I could characterise the expression of protonephridial cells and compare this to that of other lophotrochozoan larvae. In the oyster larva, however, I could not find any protonephridia (likely because they're not present) and this will be discussed in detail in the next chapter.

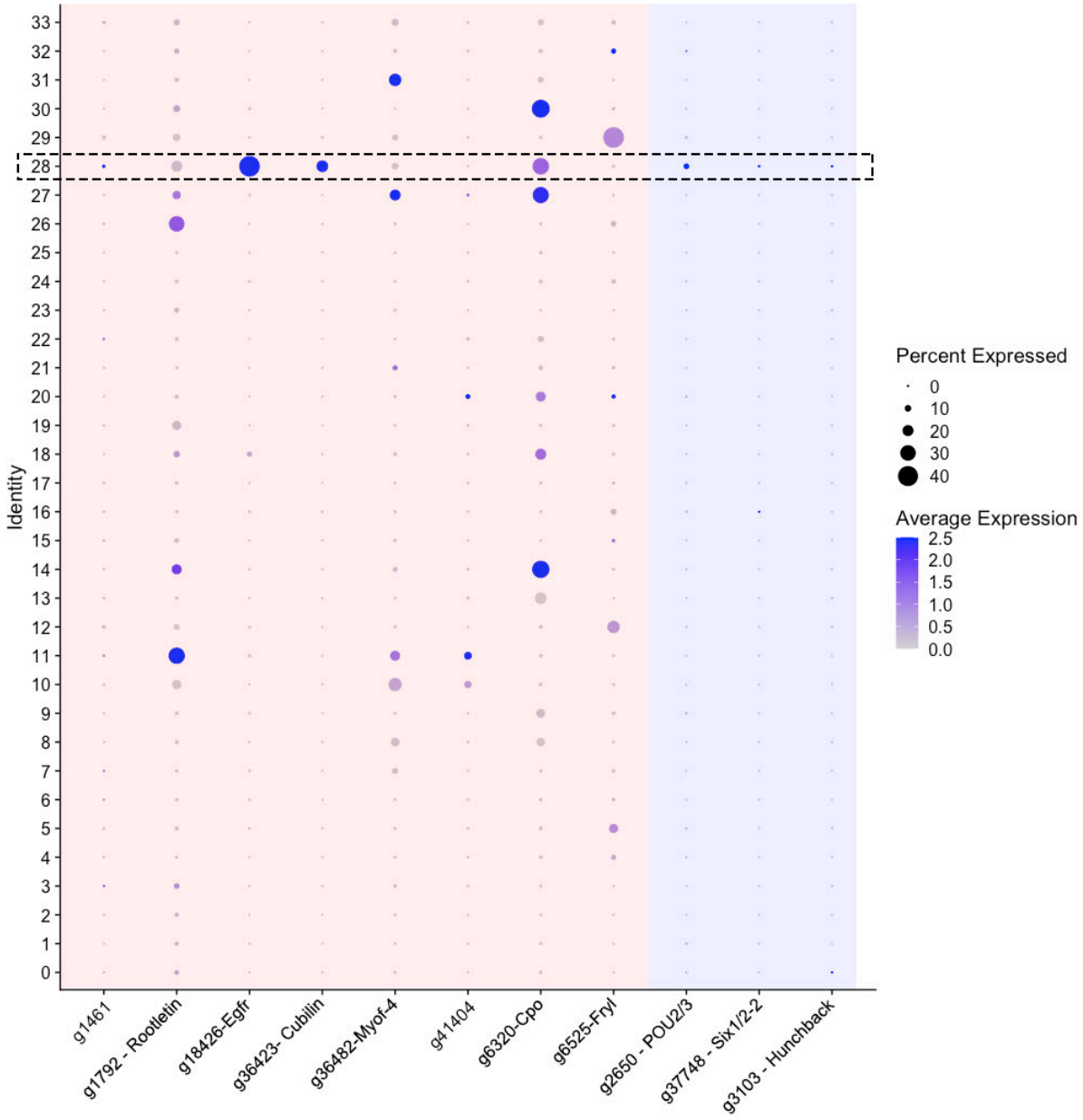


Figure 77. Typical protonephridial markers POU2/3 Six1/2-2 and Hunchback are clearly expressed in cluster 28.

Dotplot showing the expression of orthologs of *S. mediterranea* protonephridial markers in the flatworm SCS data (light red) and of three well known markers (in light blue).

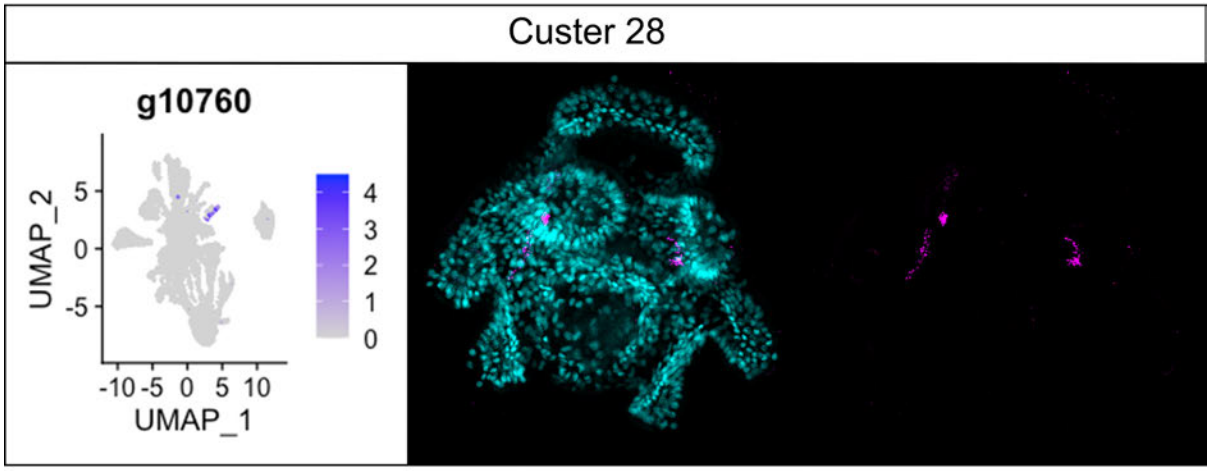


Figure 78. Expression of a cluster 28 marker (g10760) is in the protonephridia.

Small UMAP on the left show expression of the gene in scRNA-seq data, on the right HCR staining in pink of a larva oriented posteriorly. DAPI in light blue stains the nuclei.

5.5.5 Gut

Planarian guts have been described as being made up of two different cell types, phagocytes and goblet cells. When I looked for the clusters expressing *Prosthecereaus* orthologs of the phagocytes and goblet cell markers from *S. mediterranea,* I found that markers of phagocytes (and their precursors) showed expression in clusters 2, 15, 19, 23, 25, 26 and 32 whilst goblet cells markers were not consistently expressed in any clusters (see figure 79). I first selected a general marker for clusters 2, 19, 23 and 32 (Cathepsin-L) and, using HCR, confirmed it is expressed in the gut of the larva (see figure 80). Then I picked more specific markers for clusters 32 and 19 which, respectively, show expression in scattered cells of the gut (g14931-Pol3-93 yellow) and the gut lining (g10547-Vit-1 magenta) (see figure 80). HCR of marker genes specific for cluster 15 show expression in two cells on either side of the oral hood, some scattered cells in the two ventral lateral lobes and in the posterior region of the larva (see figure 81). Finally, HCR of marker genes for cluster 25, show

expression in the anterior part of the gut as well as in a few scattered cells in the apical region (see figure 82). The expression pattern of marker genes for clusters 15 and 25 is quite mysterious as it seems that not all cells from these clusters are necessarily in close contact with the gut. It is possible that some of these cells are localised near the gut in the adult but they have (also) other roles in the larvae.

The gut is another feature that the oyster larva appears to lack since it is only starting to form at the trochophore stage. As for the protonephridia, this will be further discussed in the next chapter.

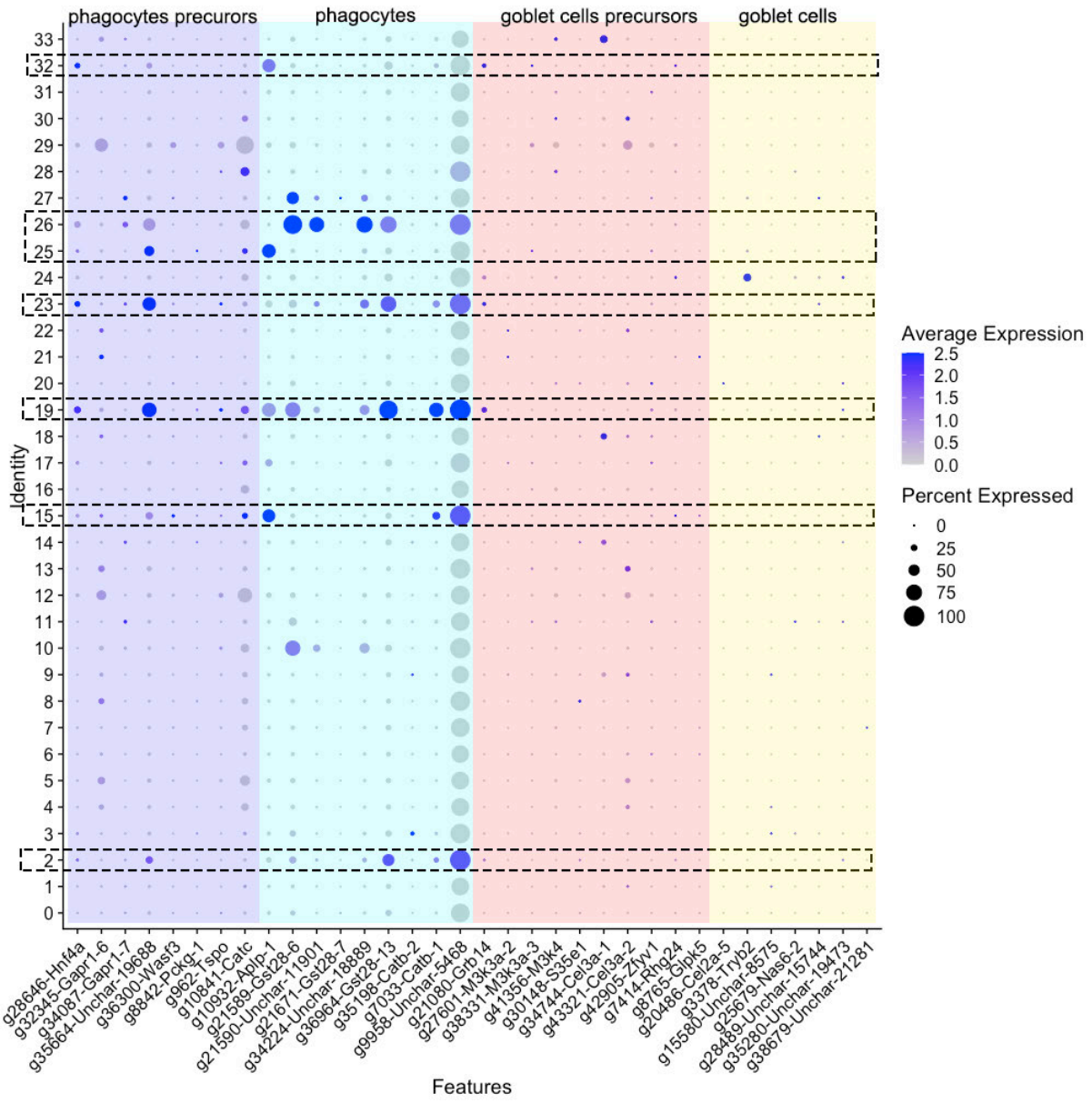


Figure 79. Markers of phagocytes and phagocyte precursors are expressed in clusters 2, 15, 19,23, 25 and 26 and 32.

Prostheceaeus orthologs of *S. mediterranea* markers of phagocytes and phagocyte precursors are expressed in clusters 2, 15, 19,23, 25 and 26 and 32 whilst no clear signal is detected for goblet cells markers.

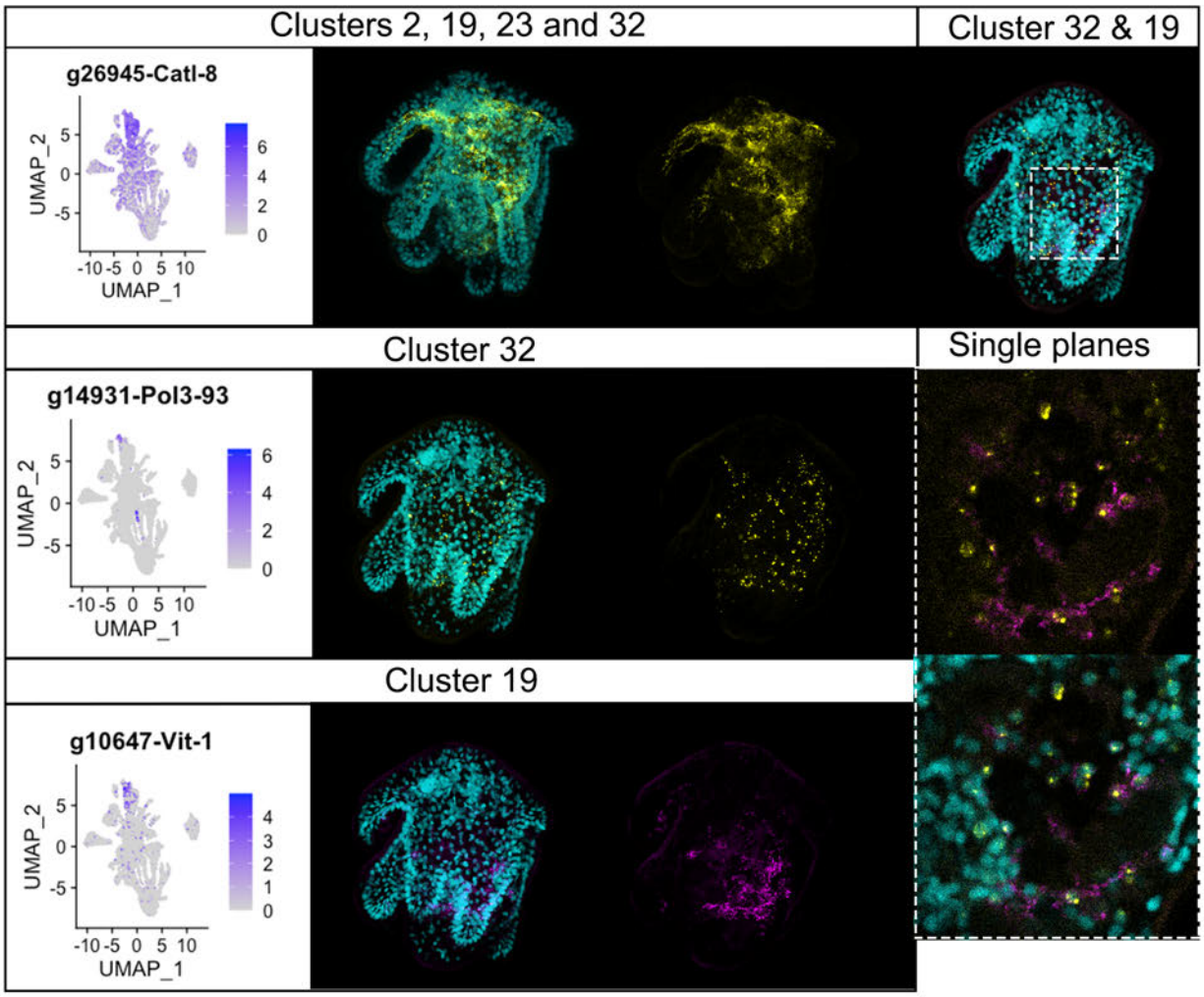


Figure 80. HCR of the general marker for clusters 2, 19, 23 and 32, Cathepsin-L (top, yellow) shows expression in the gut of the larva.

Specific markers for cluster 32 and 19 show expression, respectively in scattered cells of the gut (g14931-Pol3-93 yellow) and the gut lining (g10547-Vit-1 magenta). Small UMAPs on the left show expression in scRNA-seq data. Images are maximum projections unless otherwise stated. DAPI in light blue stains the nuclei.

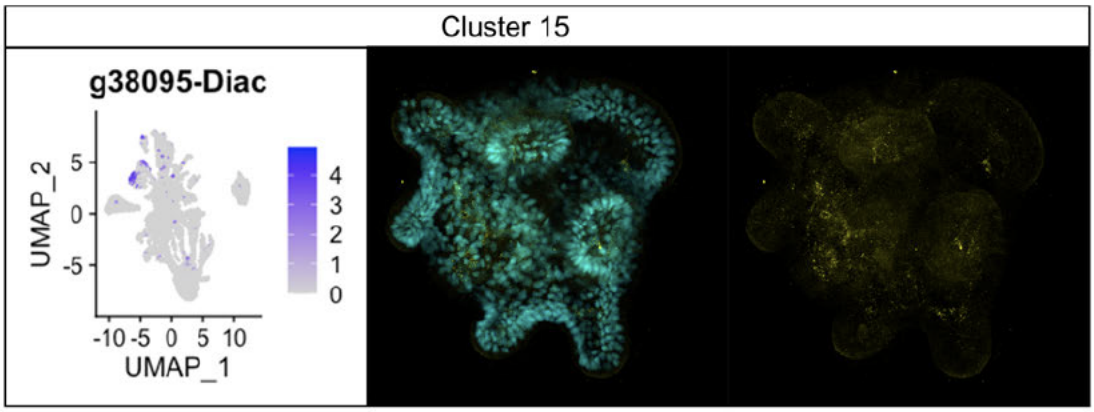


Figure 81. HCR of the gene marker Di-N-acetylchitobiase (g38095) (yellow) for cluster 15 shows expression in a few cells scattered around the larva.

Specifically, expression is localised in two cells on either side of the oral hood, a few in the ventral lateral lobes and a few in the posterior region of the larva. DAPI in light blue stains the nuclei.

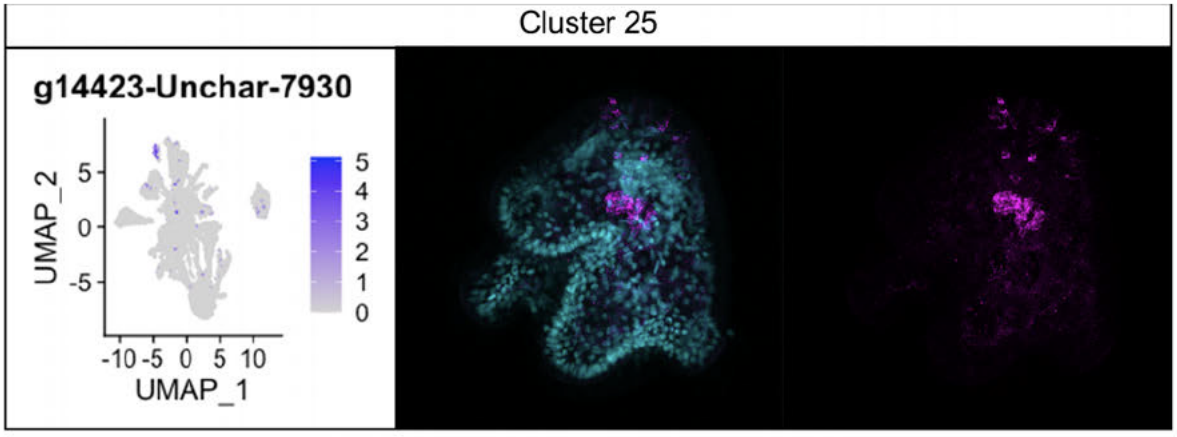


Figure 82. HCR of a gene marker for cluster 25 shows expression in the anterior part of the gut and in a few scattered cells in the apical region.

The gene marker g14423 (pink) did not return any annotated blast hit. DAPI in light blue stains the nuclei.

5.5.6 Secretory cells

Many different secretory-like cell types have been described in flatworms such as rhabdites, cement glands and adhesive cells. For example, in *Macrostomum lignano* a complex adhesive organ has recently been described which is composed of three cells: and adhesive cell, a releasing cell and an anchor cell. Here, authors have also identified a specific marker

for anchor cells, the intermediate filament protein Macif1, which is necessary for adhesion (Lengerer et al, 2014). However, since a broader review of secretory/ cells in flatworm species is lacking, to try and detect these cells in my scRNA seq data I have pulled together various markers from different cells and species. I found that some markers for secretory cells or cement glands were expressed in clusters 8 and 21 (Lengerer et al, 2018; Zayas et al, 2010) and a couple of adhesive cell markers were expressed in cluster 20 (Zayas et al, 2010; Lengerer et al, 2014) (see figure 83 A). As one can see in figure 83 B, secretory cells belonging to clusters 8 and 21 appear scattered in the epidermal layer, mostly in the apical region with just a few cells localised posteriorly. Their location and shape resemble closely that of rhabdites as described by Rawlinson (2014). Expression of Macif1, the marker of cluster 20, appears instead scattered in the mesenchymal layer of the larva in what appears like a net (see figure 83 C). Although there is no apparent co-expression of these two markers in our single cell data figures 83D and 83E show that cells belonging to these two clusters appears to be close to each other. It is possible that rhabdites and Macif1+ cells act together similarly to what, for example, has been described by Lengerer and colleagues (2014) in the adhesive organ of *M. lignano*.

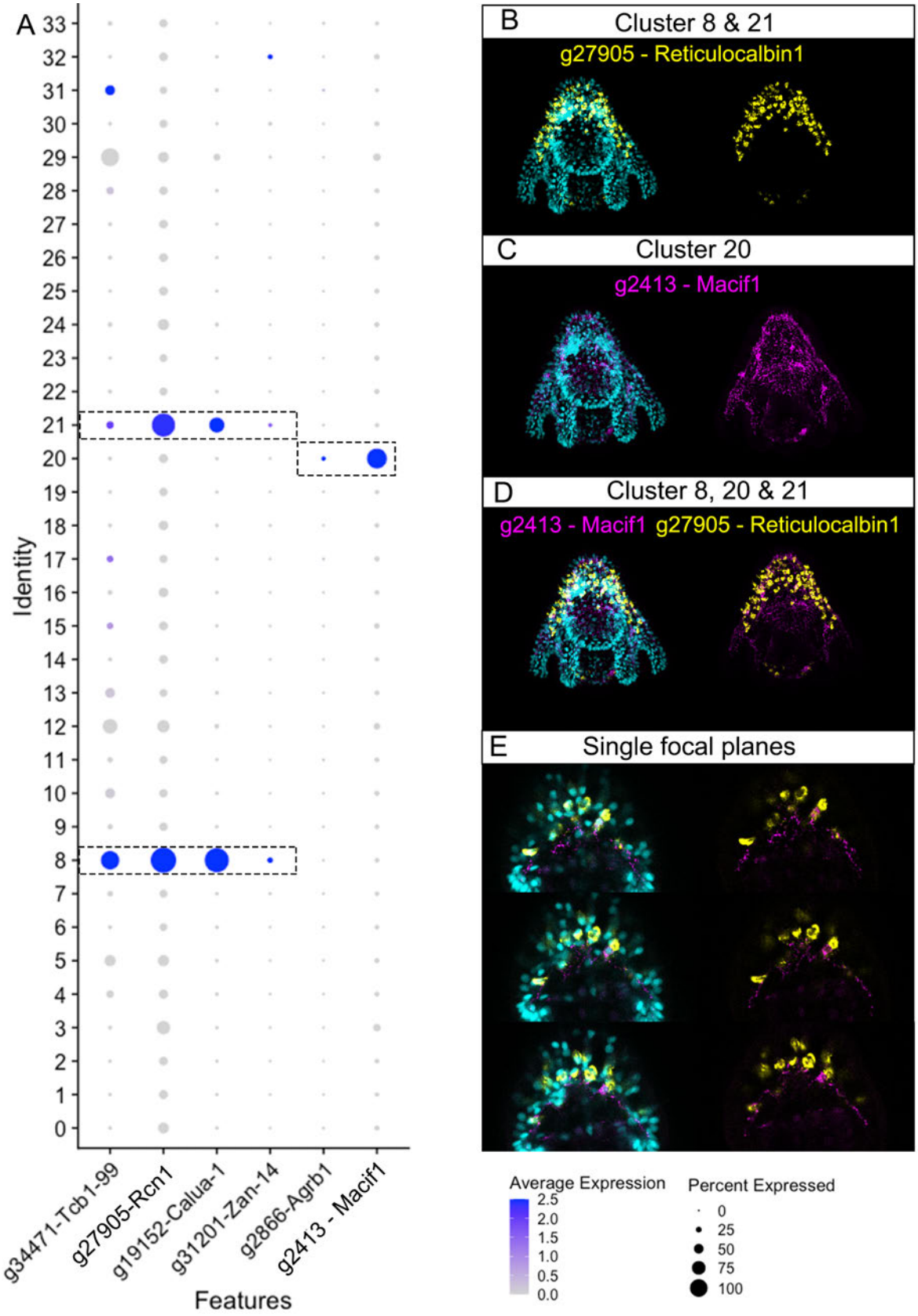


Figure 83. Rhabdites (cluster 8 and 21) and Macif1+ cells are in close proximity.

A) Secretory cells marker show expression in clusters 8,20 and 21. B) HCR of marker genes of cluster 8,21 shows expression in rhabdites C) HCR of Macif1+ cells (cluster 20) show a net like expression around the larva. D) Rhabdites and Macif1+ cells are in close proximity (D, E). Images are maximum projections unless otherwise stated. DAPI in light blue stains the nuclei.

5.5.7 Remaining clusters

I have looked for markers of other *S. mediterranea* cell types as well as for some genes explored in the previous chapters (such as spiralian specific genes for instance) however I didn't see a clear expression pattern. For this reason, I decided simply to design HCR probes for all the remaining clusters in the hope that finding out their distribution might tell me more about their identity.

5.5.8 Clusters 10 and 11

Clusters 10 and 11 were analysed together since cluster 11 shares most of its marker with cluster 10 (as one can see in figure 84). Most markers of cluster 11 did not return any annotated hit (or any hit at all) when blasted on ncbi or uniprot. HCR *in situ* of the genes tyrosinase (g33590) (expressed in both clusters), Otx (g1930) (expressed in cluster 10 only) and collagen type 4 alpha 2 chain (g43312) (expressed in cluster 11 only) show expression in the ciliary bands and in the apical tuft (see figure 85). Specifically, it seems that the ciliary bands are made up of cells from both clusters whilst the apical tuft has only cells from cluster 11. The expression of the transcription factor Otx expression in the ciliary band is extremely interesting as this has been described in other lophotrochozoan larvae as well as in swimming larvae of echinoderms and hemichordates (Marlow et al, 2014).

Since ciliary bands and the apical tuft are one of the main features of lophotrochozoan larvae these two clusters will be dealt with in depth in the following chapter.

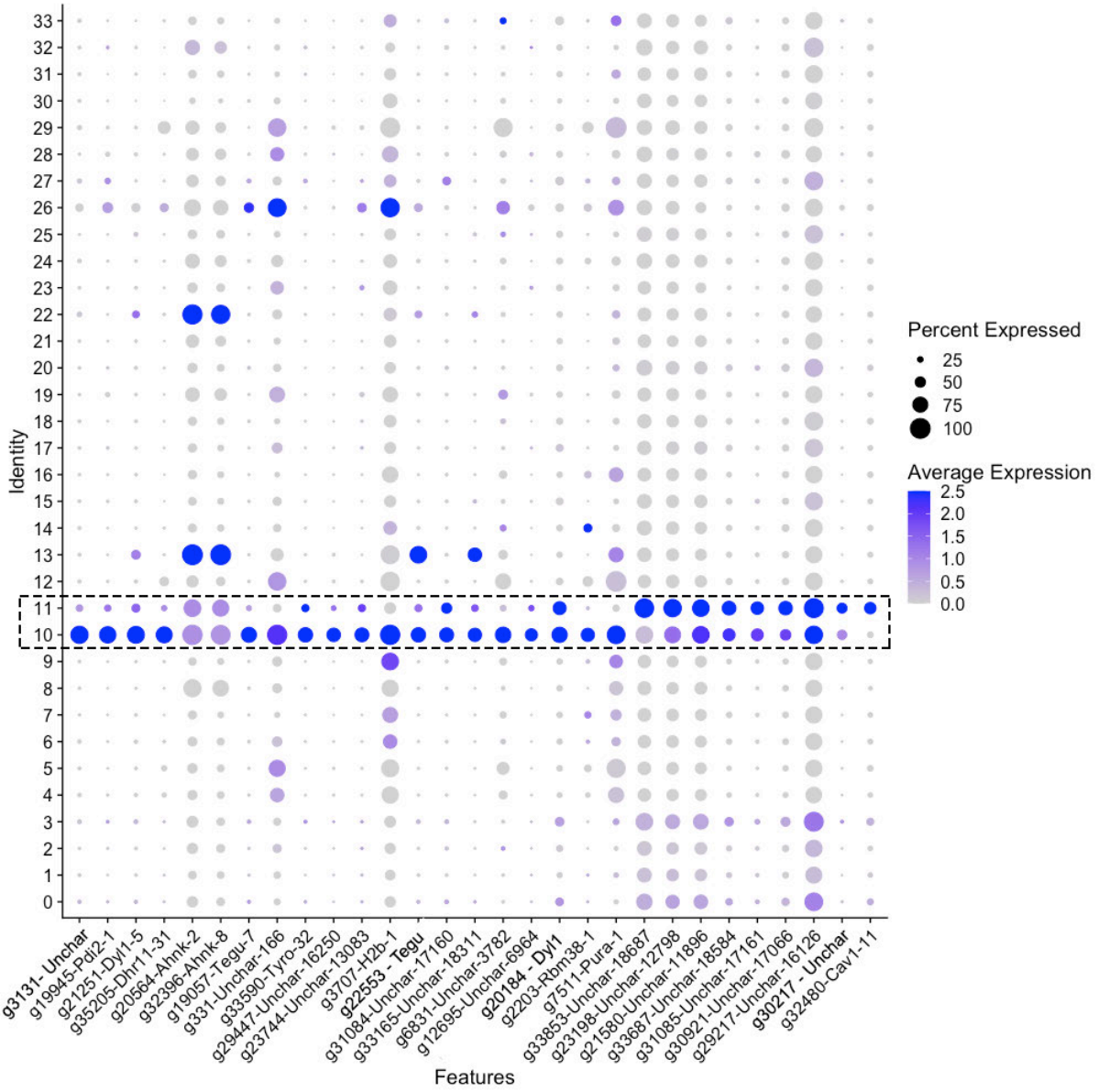


Figure 84. Most markers of cluster 10 and 11 are shared but uncharacterised.

DotPlot showing the expression of the top markers for cluster 10 and 11, most markers of cluster 11 and are shared with cluster 10 and they remain uncharacterised.

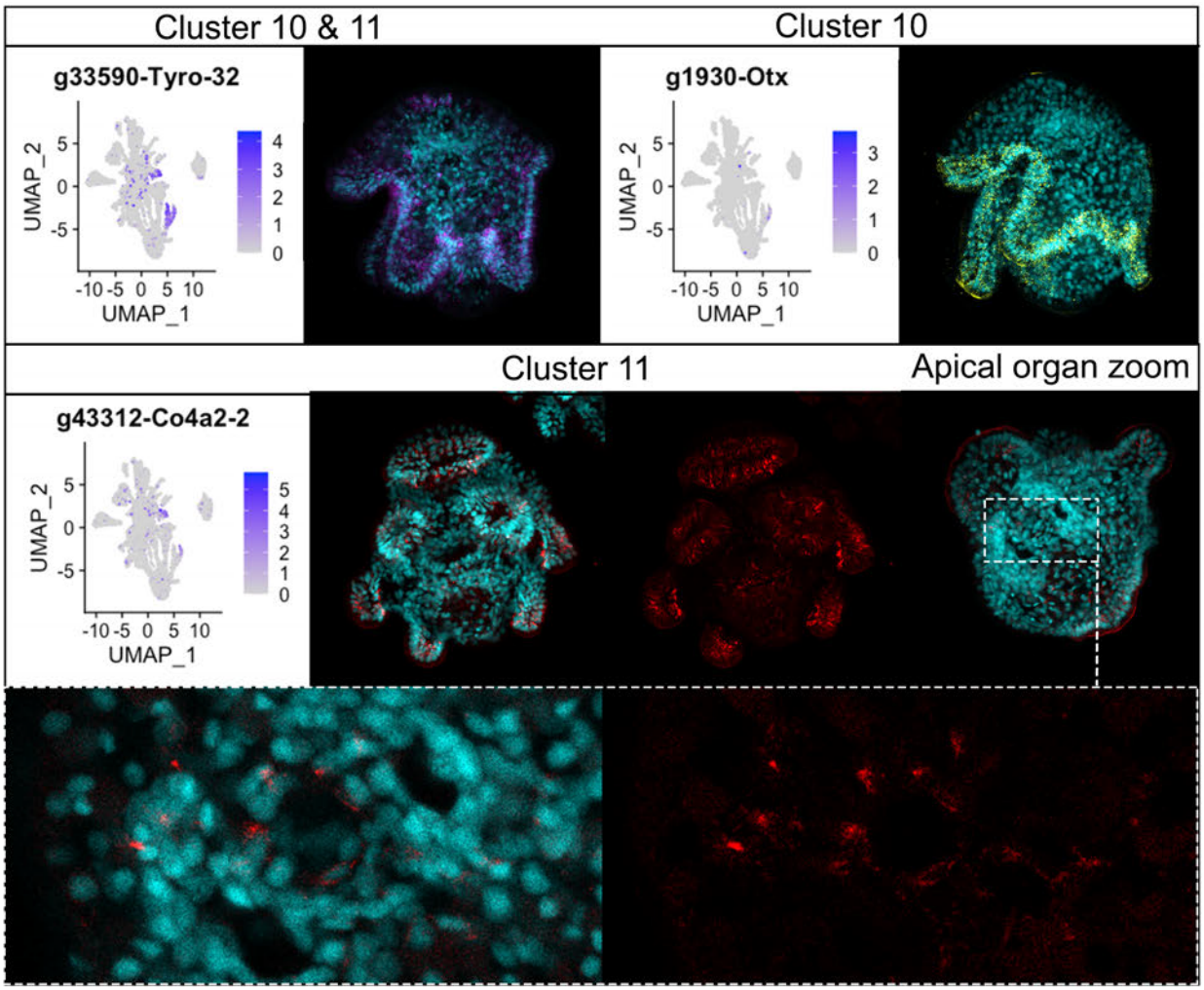


Figure 85. HCR of flatworm larvae of three markers for clusters 10 and 11 show expression in the ciliary bands.

Tyrosinase (magenta) and Co4a2-2 (red) shows additional staining in the apical tuft. Bottom dotted lines are single focal planes near the apical tuft. Images are maximum projections except for bottom row which are single planes near the apical tuft. DAPI in light blue stains the nuclei.

5.5.9 Clusters 22 and 13

Clusters 22 and 13 also appear to share the majority of their cluster markers and for this reason I chose a marker gene for both, Slc-g1340 (see fig. 86). HCR of this gene shows expression in a few scattered cells in the mesenchymal layer of the larva (see fig. 87). The only references I could find in the literature for gene markers of this cluster (namely

aquaporin, stomatin and tegument antigen) were associated with the tegument (Farias et al 2010; Figueiredo 2014; Chienwichai et al 2020) . The tegument is a novel epidermal layer typical of parasitic flatworms belonging to the clade Neodermata and had never been described outside, I believe it's just by chance that I find these genes in the mesenchymal layer of a polyclad larva (which is not a member of the Neodermata). I am not sure what these cells are and they will be from now on referred as tegument antigen + cells. Since their scattered expression resembles somewhat that of rhabdites, I performed double HCR using markers of these two clusters. Figure 88 shows that tegument antigen + cells are in close contact with rhabdites, similarly to Macif1+ cells. This is a further indications that these three clusters could act together in some sort of secretory complex.

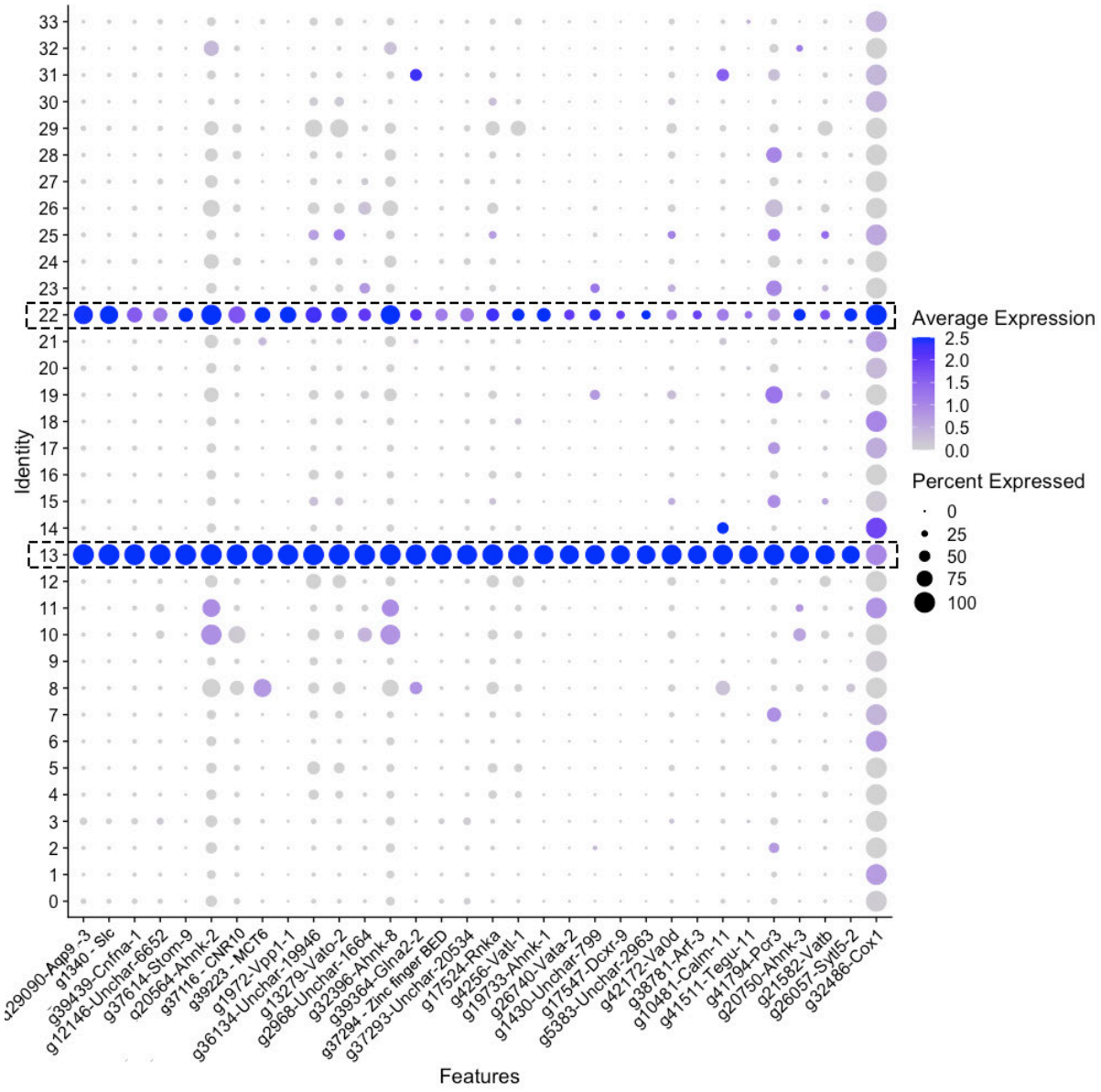


Figure 86. Dotplot of gene expression of the top cluster markers for cluster 13 and 22 showing expression is shared between the two clusters.

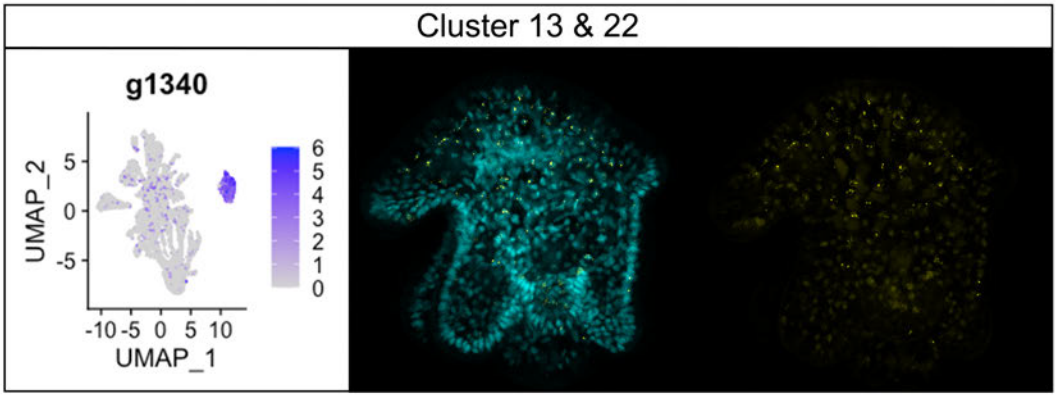


Figure 87. HCR of the gene marker g1340 or Slc show expression in scattered cells of the mesenchymal layer of the larva.

Small UMAP on the left show gene expression in scRNA-seq data. DAPI in light blue stains the nuclei.

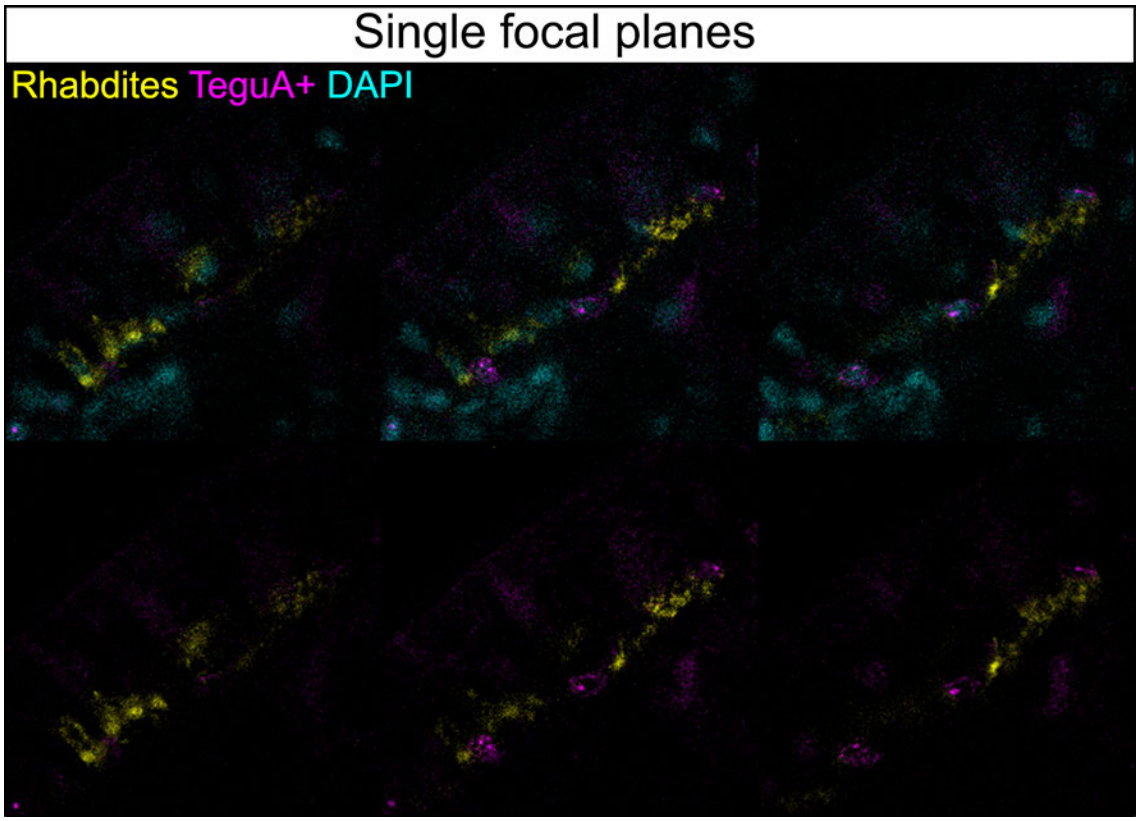


Figure 88. Cells of clusters 13 and 22 are in the same layers as rhabdites and loosely connected.

HCR of the gene marker g1340 (magenta) markers for cluster 13 & 22 and g27905 (yellow) a marker for rhabdites show that these cells are in the same layer and closely connected. DAPI in light blue stains the nuclei.

5.5.10 Cluster 14

Some of the top twenty markers of cluster 14, which can be seen in figure 89 below, appear to be linked to neuronal activity (i.e. sodium channel protein 1 brain like, neuroacetylcholine receptor, synaptotagmin). Moreover, many of these markers are also expressed in clusters 18 and 30 which I have already identified as a neuronal cluster. Indeed, looking more carefully at the expression of neuropeptides in figure 74, a few do seem to show expression

in cluster 14 as well. Finally, HCR for the cluster marker (and neuropeptide) TRH shows high expression in two cells present laterally to the apical organ as well as in the ciliary lobes. The expression the ciliary lobes is more clearly visualized with a second marker (stereocilin - g24624) which also highlights expression in one cell around the mouth (see figure 90).

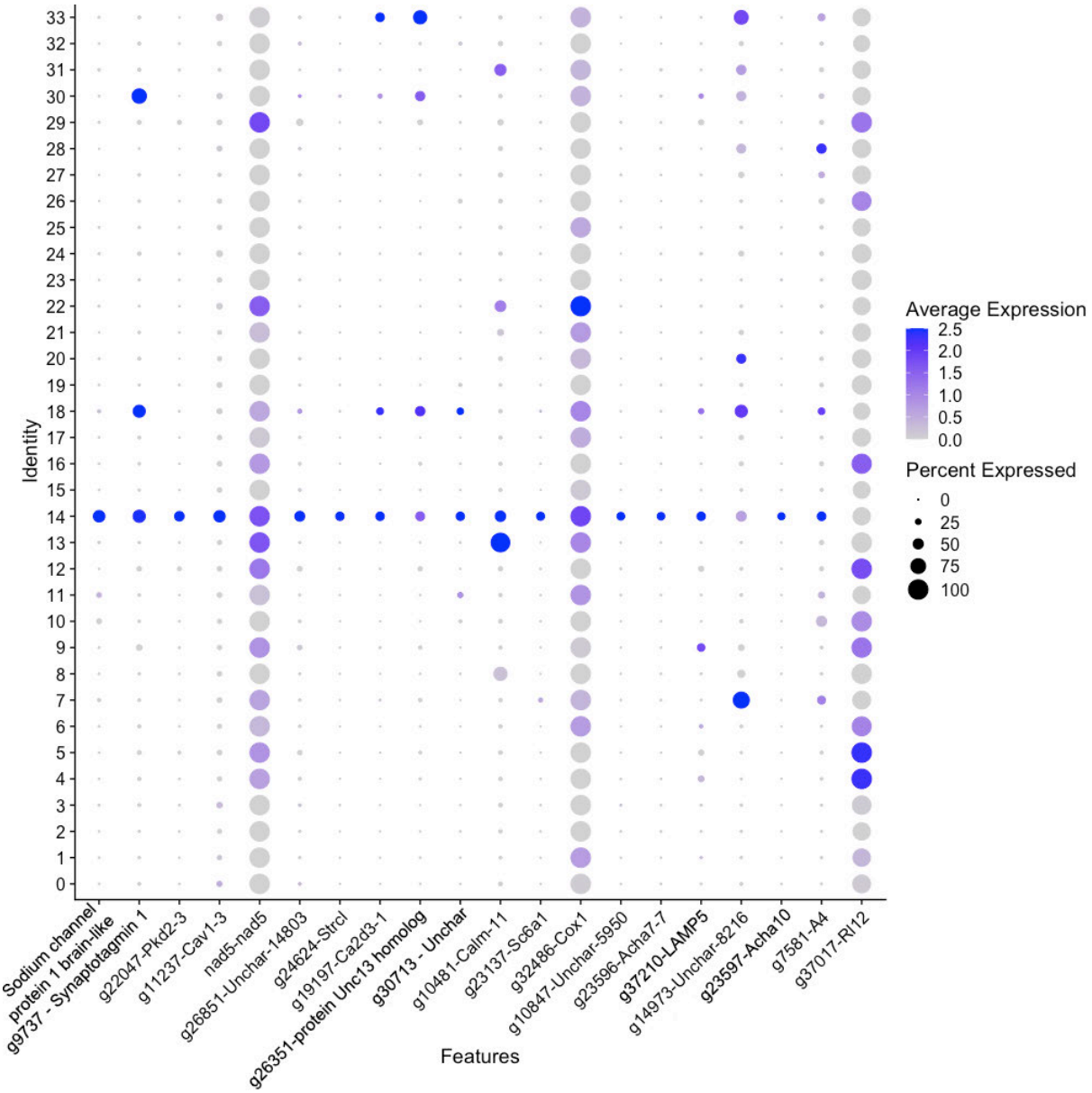


Figure 89. Cluster 14 shares its top marker with neuronal clusters.

DotPlot of top 20 markers of cluster 14 showing many shared markers with the other two neuronal clusters 18 and 30.

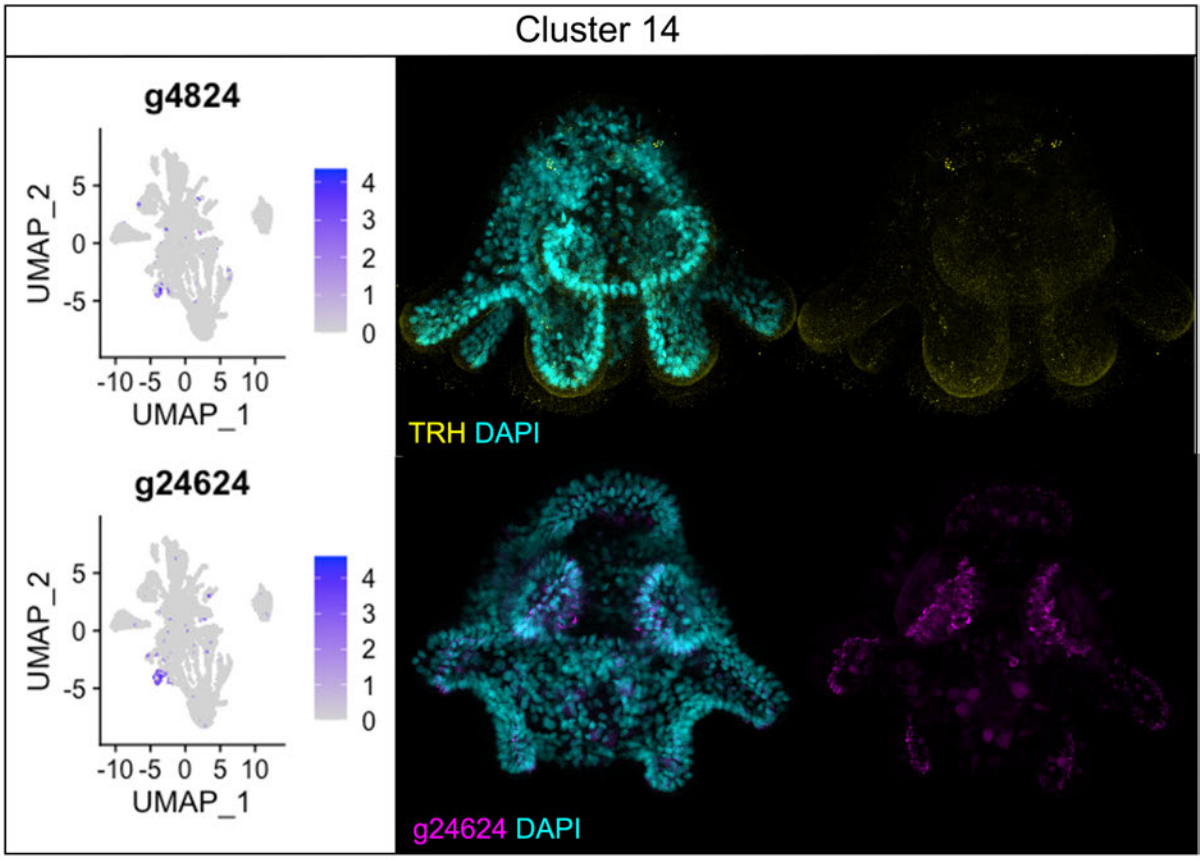


Figure 90. Neuronal cells from cluster 14 are localised in the ciliary lobes.

HCR of the cluster marker TRH (yellow) shows high expression in two cells laterally to the apical organ and weaker expression in the ciliary lobes. HCR of the marker gene g24624 (pink) also shows expression in cells of the ciliary lobes. Small UMAPs show expression of the genes in the scRNA-seq data, DAPI in light blue stains the nuclei.

5.5.11 Unidentified clusters

A few clusters remained unidentified at this point for a number of reasons. Clusters 0, 1, 3 and 6 belonged to the central large blob in the UMAP, and were not explored as they potentially represent lower quality cells/cells undergoing differentiation (this is quite common in developmental SCS experiments). Other small clusters had very few specific markers, among which some were too small for HCR probes, and some failed (this is true for clusters 27, 17, 16, 26, 24 and 31).

5.6 Relationship between cell-types in the flatworm larva

In the previous paragraphs I have assigned cell type identities to as many clusters as possible and found many well-known cell types of lophotrochozoan larvae (such as ciliary bands, the apical tuft, cells from the apical organ cells and protonephridia) some common animal cell types (gut cells, muscle cells) and also some cells never previously described in the Mueller's larva (such as neoblasts, tegument antigen + and Macif1 + cells). Similar to what I have done with the oyster, I was interested in finding out what the relationship between these cell types could be to try and work out, for example, if the different groups of neurons or neoblast or secretory cells were more similar to each other than to other cell types. I performed the same analysis as described in chapter 4, using the average expression matrix to compute the Euclidean distance between the clusters, and obtained the tree shown in figure 91A. Starting from the top we can see that in this tree all my predicted neuronal cell types are grouped together (dark blue) with very high support (100%) however, muscles and cluster 30 (whose identity was not assigned) also belong to this clade with high support. Next, we find a group containing all my predicted "secretory cells" (tegument antigen + and rhabdites). Gut cells group together with high support (orange) and in this clade we also find the two clusters 15 and 25 which were expressing gut related genes (although their expression pattern was not restricted to the gut). Neoblasts also belong in the same clade with high support (light green). Surprisingly ciliary band cells (cluster 10 and 11) group separately even though they appeared to share many cluster markers, this could be due to the fact that cluster 10 contains ciliary precursors, which share genes with neoblasts (see chapter 6.2.1). Since for the flatworm I found some slightly odd results I wondered if some of these could be caused by transitional clusters (i.e. cells that

are still differentiating). For this reason, I tried first to prune out a few un-identified clusters that are present to the middle of the UMAP (cluster 0,1,3,6) (figure 91B) and then the neoblasts clusters (figure 91C). Regardless of the pruning I still obtained very similar groupings, with neurons still grouping together with muscles and cluster 33 (although with lower support) and ciliary bands in separate clades.

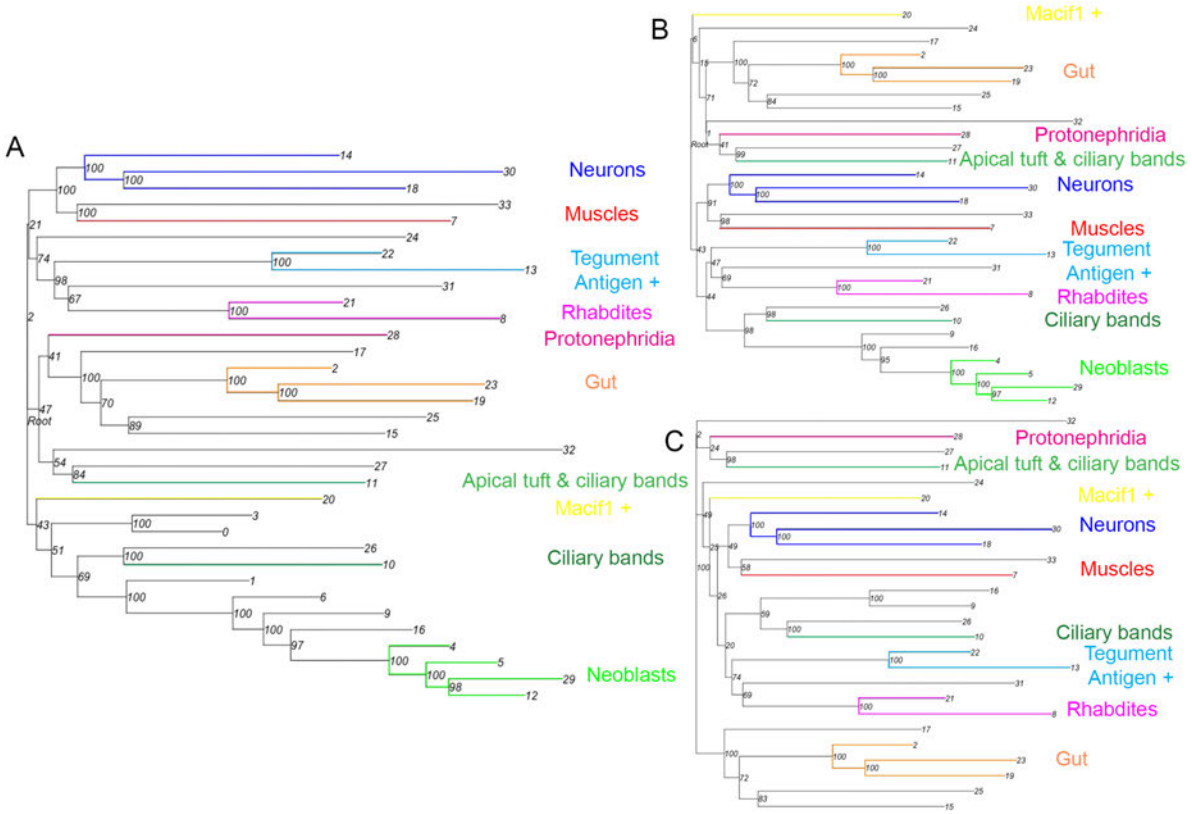


Figure 91. Cell type trees showing relationship between clusters calculated as Euclidean distance showing some similar cell types grouping together, like neurons, secretory cells and neoblasts.

Pruning "middle clusters" (likely representing not fully differentiated cells) (B) and neoblast (C) doesn't substantially change clustering but reduces some bootstrap values. Bootstrap values represent % of time that clade was recovered (10,000 repeats). Tree was rooted via midpoint rooting (roots at the midpoint between the two tips most distant from each other).

5.7 Transcription factor signatures in different clusters

Similar to what I have done with the oyster larva, I was interested in seeing what the transcription factor signatures would be for different clusters of the flatworm larva. As one can see in figure 92 below, strikingly many TFs are expressed in the neoblast clusters (see figure 92 top). Most of the non-neoblast clusters also have specific TFs signatures even though it seems that on average less than 40% of cells in each cluster express them (small dots in dotplots indicate the % of cells in the cluster expressing the gene) (see figure 92 bottom). The most obvious explanation for this result is that, although I re-sequenced samples Pc3 and Pc4 more in depth, I did not achieve the same read per cell depth as for the oyster. This is because when I re-sequenced the samples, I obtained roughly double the number of cells that I previously had, which helped improving the cell coverage of different cell types (more cells per type) but also meant many of the reads were from cells not previously considered (due to low coverage) producing a smaller increase in overall reads per cell (this commonly get calculated as total reads divided by number of cells). Since TFs tend to be expressed at a lower level than other genes, they are also less likely to be captured in scRNA-seq data and it is possible that we randomly pick them up only in a subset of cells. Alternatively, it is also possible that some of these clusters should be further subclustered, however when I tried this, I mostly obtained more neoblasts clusters which didn't seem useful for the scope of this work (data not shown).

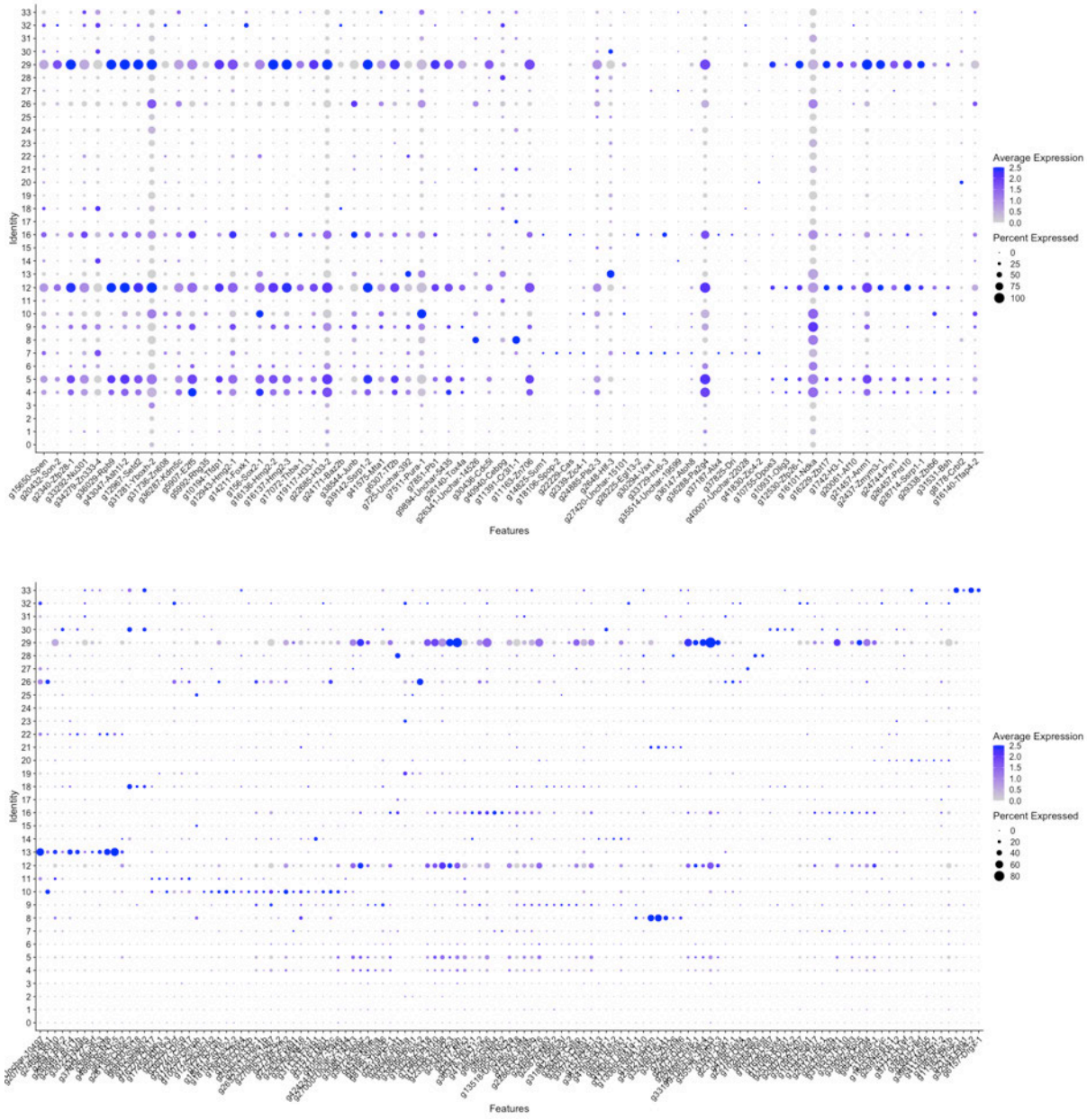


Figure 92. Dotplots of TFs expression in different clusters, divided into neoblast specific (on top) and the rest (on the bottom) shows many TFs are specifically expressed in all neoblast clusters.

However, many other clusters also have distinct transcriptional signatures although these are often expressed in a subset of cells (small dots).

5.8 Transcriptome age index (TAI) for single cell clusters in the flatworm larva

Having worked out the relationship between the different cell clusters, I decided also to look at the transcriptomic age index of these cell types. As explained in the previous chapter, several recent studies have looked at this value in bulk RNA-seq data, although no results have been published on the flatworm larva. In general, these studies appeared to show a peak of expression of "younger" genes at the larval stages for several lophotrochozoan. In the previous chapter I have shown how this result was likely due (at least in the mollusc larva) to the expression profile of the shell gland cells and I have also highlighted the main drawbacks of the phylostratigraphy approach. However, I still thought it would be interesting to look at the results of this analysis in the flatworm larva. Using the same approach as described for the oyster I first calculated the number of genes per phylostratum (see figure 93). Similar to what I observed with the oyster I found very few genes belonging to the Protostomia and Lophotrochozoa clades. However, for the flatworm I found many more species-specific genes. This could be partially because there are fewer flatworm genomes available compared to molluscs and that many belong to parasitic flatworms that could be very derived. Once again it is good to keep in mind that this is the overall distribution of genes used for the TAI calculation.

Figure 94 shows the TAI per cluster obtained for the flatworm larva. At first glimpse we can notice how most clusters seem to have similar TAI value with only a few outliers, one neuronal cluster (18) with a slightly higher (younger) TAI and neoblasts having on average a lower (older) TAI. Interestingly, I had also found that the gut "neoblasts" of the oyster presented a lower (older) TAI than most other cell types, I would hypothesize that this is

because neoblast have a cell type signature similar to proliferating cells and that this is quite an ancient feature of animals (see also figure 95 for contribution of each phylostrata).

Although TAI values themselves are not easily comparable across species (since each species will have different phylostrata), we can see that the overall variance in TAI of the flatworm (2.6-3.5) is lower than what observed in the oyster (3.2-4.1) and that in general the outliers are not so striking.

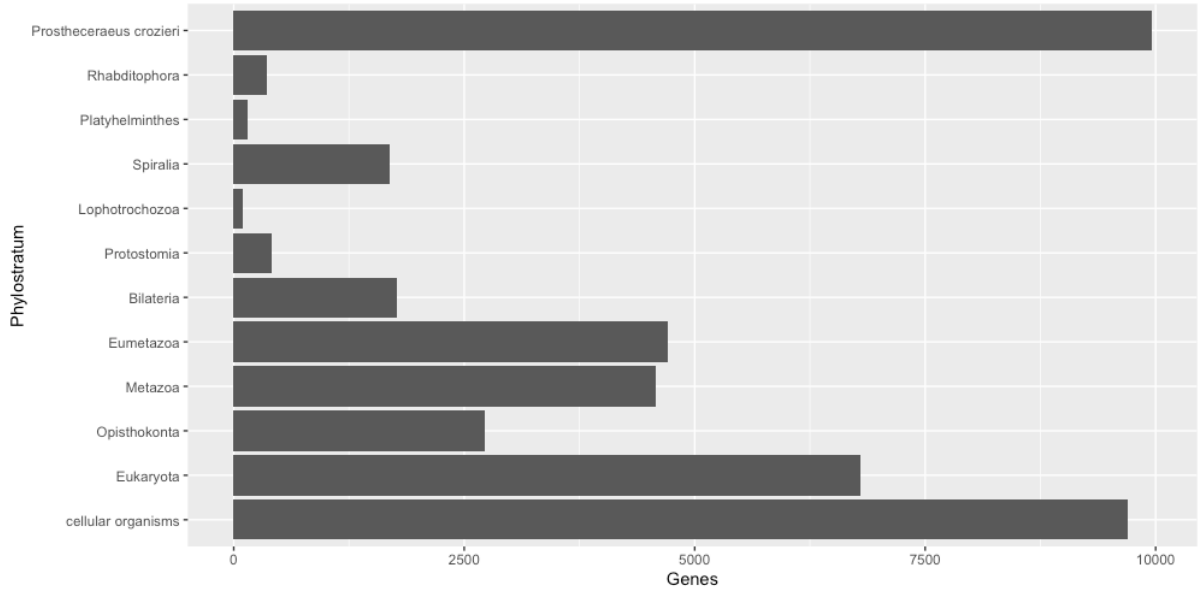


Figure 93. Number of genes of *P. crozieri* belonging to each phylostratum are higher in older PS, however many appear species-specific.

Similarly to what observed in the oyster, few genes belong to the Protostomia and Lophotrochozoa PS.

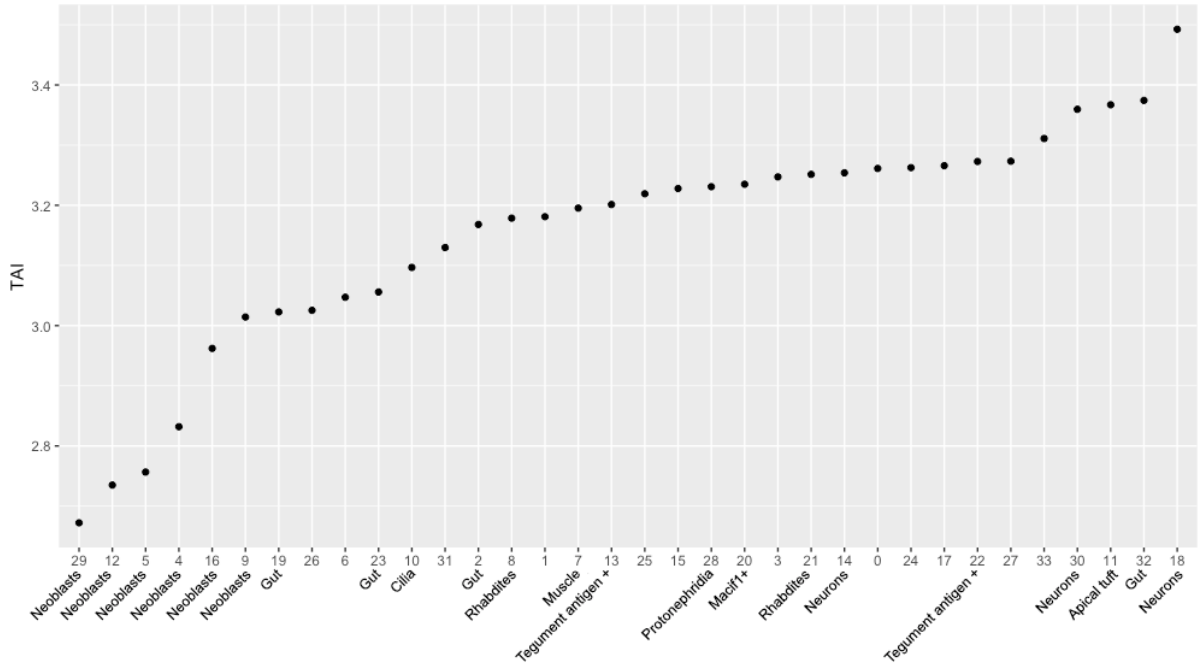


Figure 94. TAI of *P. crozieri* larval clusters is similar except for the neuronal cluster 18 appears to have a "younger" transcriptional signature and neoblasts appear "older".

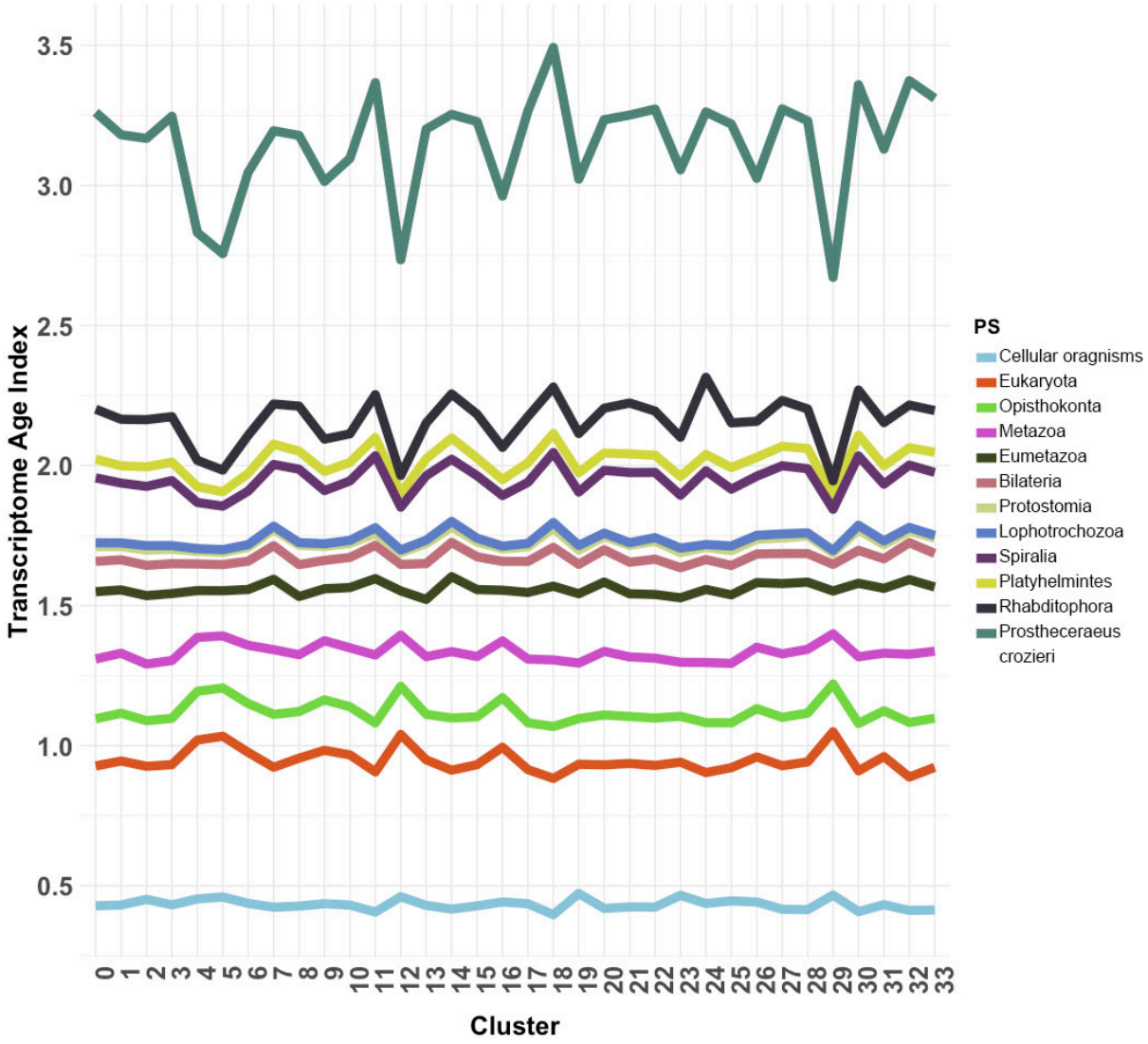


Figure 95. Gene contribution for neoblast (4, 5, 12, 29) older TAI is driven by pre-metazoan genes.

Contribution of different phylostrata to the total TAI of the flatworm divided by cluster

5.9 Conclusions

5.9.1 Summary of results

In this chapter I have introduced the single cell sequencing experiments carried out on the flatworm larva, I have shown how our newly sequenced genome improved the quality of

our single cell sequencing datasets and highlighted differences between our experimental replicates. I have dealt with the problem of the slightly higher overall mitochondrial gene content in the flatworm scRNA-seq compared to the oyster and proved that this does not seem to impact the overall quality of the data and in fact could be a species-specific trait. I have then used the best quality replicates to perform clustering analysis and identified the main cell types present in my data using a combination of literature searches, comparison with other flatworm scRNA-seq data and newly established HCR *in situ*. These analyses allowed me to assign cell type identity to most of the clusters in my data. Moreover, I established similarity relationship among these cell types and explored their transcriptomic age index showing how neoblasts appear to express older (or slower evolving) genes than other cells in the Mueller's larva. The analysis carried out in this chapter will help me in the comparison between the polyclad flatworm larva and the oyster larva which I will discuss in the next chapter.

5.9.2 Traits of the Mueller's larva that are commonly found in other larvae of lophotrochozoan

The aim of this thesis is to compare the transcriptional profile of the cell types present in a trochophore larva with those of the more derived larva of polyclad flatworms. As I mentioned in the previous chapter the trochophore larva we chose in this study, although morphologically very similar to other trochophore larvae, lacks some quite standard larval structures such as a differentiated gut and protonephridia. On the contrary the flatworm larva presents both, and in fact I found up to 5 different cell clusters that are part of or likely related to the gut. The protonephridia in particular could be of great interest to carry out

comparison with other larvae in the future, in fact a recent study by Gasiorowski and colleagues (2021) showed that a handful of genes are co-expressed in protonephridia of many animals. However, in this chapter I show how the transcriptional signature of larval protonephridia appears quite different from that of planarian adult protonephridia with only few co-expressed marker genes. It is possible that is due to the evolutionary distance between the two worms (planarian worms specifically are on a long branch) but it is also possible that larval and adult protonephridia could have distinct molecular expression. To understand whether this stage specific differences exist and whether they can tell us something about the origin of these cells and of larvae it would be especially interesting to compare protonephridia across stages and species.

Moreover, the flatworm larva appears to possess a very complex nervous system with an apical organ (and various neuronal subtypes) connected to larval eyes and an apical tuft (although cells from the apical tuft cluster together with ciliary cells of the lobes). Larval eyes and apical tuft were not identified in the oyster larva so it would be noteworthy to also compare these with that of other larvae as soon as scRNA-seq data should become available.

5.9.3 Flatworm specific cell types

Apart from well-known larval cell types, such as ciliary cells, apical organ neurons and larval eyes and protonephridia, I also identified many cell types that either are flatworm specific and/or never previously described in the flatworm larva. For example, I identified several clusters of secretory cells, some of which are likely rhabdites, that seem to be in close contact with one another and may act together similarly to what observed in the adhesive

organ of *M. lignano* (Lengerer et al, 2014). Additionally, I found several neoblast clusters whose marker genes resemble that of planarian neoblast, I did not however explore the differences between these different clusters as it was beyond the scope of this work.

Overall, the scRNA-seq of the polyclad flatworm larva generated here together with the working HCR protocol I devised represent a great contribution to the study of marine larva and will certainly be useful to other researcher beyond the comparison presented in the next chapter.

6 Comparing cell types across species

The ultimate aim of my PhD is to compare the classical trochophore larva of a mollusc - the pacific oyster - with the (possibly) more derived Mueller's larva of polyclad flatworms to try and understand if these two larvae are homologous. To enable this comparison, I have gathered single cell sequencing data from larvae of both species, checked the quality of the data, used the transcriptional profile of single cells to predict cell clusters (or cell types) and then screened the literature to try and assign cell type identity to each cluster. Moreover, I have validated the cellular identity of most of the clusters using *in situ* hybridisation (see chapter 4 for the oyster larva and chapter 5 for the flatworm larva).

In this chapter I will compare the two datasets to try and find homologous cell types. I will first try to find cell types expressing several homologous genes across species. Then, using the data gathered in the previous chapters, I will try and work out if cells expressing several homologous genes appear to have similar function. The fact that similar cells express similar genes would be a strong indication of their common origin/homology. Finally, to answer the question of whether the larvae themselves are homologous, I will look for larval specific cells (i.e. cells that are lost during metamorphosis). This is because larval specific cells are unlikely to have been co-opted from adults (if they were they would have had to be lost in adults multiple times). The existence of homologous larval specific cells in mollusc and flatworm larvae would mean that these cell types existed in the larva of the common ancestor of mollusc and flatworm and suggest therefore that these larvae are homologous.

The main challenge of comparing single cell data across species is the fact that different species have different genomes and different sets of genes. The first task when performing any cell type comparison across species is to establish the evolutionary relationships of the genes involved (Tanay & Sebé-Pedrós 2021). After establishing the relationship between each gene in both species we have two possibilities: either using only genes that have a perfect one to one match (single copy orthologs) or using all orthologs. Using 1:1 orthologs is the simplest method, since it allows us to create direct correspondences between the genes in both species, however, the strict requirement of 1:1 orthology greatly reduces the complexity of the data (in the case of the oyster and the flatworm I only have ~5000 pairs of single copy orthologs). On the other hand, using all homologs requires more complicated methods to integrate the data. In the next few paragraphs I will present these two different approaches that I used to compare cell types across species using only single copy orthologs (Seurat Integration) and all orthologs (SAMAP).

6.1 Comparison of cell types using Seurat integration

6.1.1 Exploring the use of single copy orthologs for cell type clustering

The first method I tried to compare scRNA-seq data across species is the Seurat Integration method (Butler et al, 2018). This approach requires two gene expression matrices that have the same genes and for this reason I had first to calculate 1:1 orthologs between the two species. To find single copy orthologs, I ran Orthofinder on the proteomes of my two species and obtained a total of 5004 single copy orthologs. Since these are a considerable subset of my original data (both original scRNA-seq data sets contained over 20K genes) I first wanted

to check how different my two datasets would look when using only these small genes subset – for each species independently, are these 5004 genes sufficient to recapitulate the clusters I observe with the total set of genes? The result of performing clustering analysis using only single copy orthologs reduces the complexity of the UMAP considerably (i.e. going to >30 clusters to ~10) (see figure 96). However, it is hard to tell much else from simply looking at the UMAP. What would be useful is to see how many cells stay in the same clusters across the two datasets (i.e. the original one with more than 20K genes and the single copy ortholog one with ~5K genes).

For this reason, I decided to look at the percentage of cells that remain in the same cluster and I show the result of this analysis in the heatmaps of figure 97. For the oyster dataset (top heatmap), 7 out of 31 clusters (clusters 12, 15, 1, 25, 5, 23 and 3) remain almost identical when using only single copy orthologs versus all genes. Several other clusters, such as two ciliary band clusters (2 and 22), two shell gland clusters (6 and 20), haemocytes, gut neoblasts and neurons end up grouped together when using only single copy orthologs but remain separate from other cell types. About 8 clusters are completely absent, with cells being scattered across different clusters (24, 17, 16, 26, 28, 31, 27, 30). These lost clusters are either very small (31, 26) or are clusters that I was unable to identified in chapter 4, and which may represent transitional clusters or those made up of low-quality cells. Altogether I observed that a reasonable number of clusters remained correctly grouped when using only 5000 genes these data might give us some insight in the comparison between the two species.

For the flatworm, the result of clustering using only single copy orthologs appears a bit less successful, in fact, only 5 clusters remain very similar between the all genes vs single copy

orthologs analyses (clusters 8, 7, 13, 24, 20). Cells originally belonging to separate gut clusters were slightly re-shuffled into two different clusters, and the same is true for cells of the ciliary band, apical tuft and neoblasts. Similar to what I observed with the oyster, neurons belonging to clusters 18 and 14 end up together. Approximately 6 clusters get completely lost (16, 27, 28, 29, 30 and 33). These are all slightly smaller clusters, however among these we have some very distinct cell types such as protonephridia (28) and one cluster of neurons connected to the cerebral eyes of the larva (30). In general, the result obtained in the flatworm is less clear than what was seen with the oyster data, and highlights how a comparison using only single copy orthologs is likely to cause problems due to a considerable loss of complexity. Similar cell types (such as neurons or ciliary cells) often get grouped into a bigger more general cluster which could prove problematic for downstream analysis. This is because, ideally, we would like to find larval specific cells (such as for example specific neuronal types, or ciliary band cells) since we already know that in all Bilateria (at least) general neuronal cell types are homologous.

To see whether the problem stems from the reduction of genes or is something specific to the restriction to single copy orthologs I checked whether a random subset of 5000 genes would behave better or worse than single copy orthologs for cell type clustering. The result of this analysis can be seen in figure 98 for the oyster (on top) and the flatworm (on the bottom). As one can observe by comparing figure 2 with figure 3, using a random selection of 5000 genes appear to be better for clustering than single copy orthologs. I here show results for only one random selection but the result was consistent in 5 different repeats. The likely explanation is that many of the 1:1 orthologs are housekeeping genes expressed

in all cells and this is a strong indication that using a subset of 1:1 orthologs for comparison across phyla is not a very strong approach.

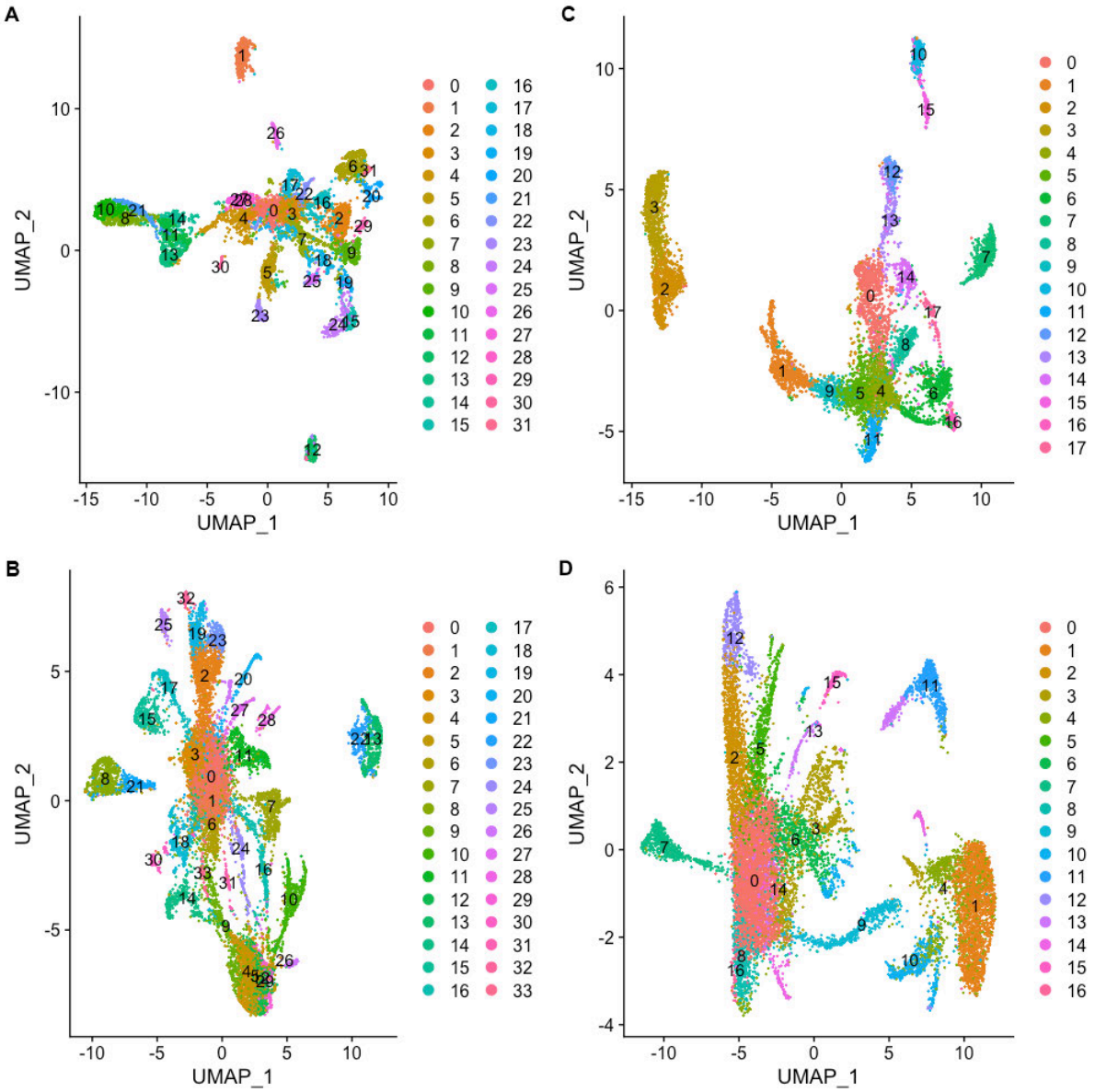


Figure 96. Using only single copy orthologs reduces complexity of scRNA-seq data.

UMAPs of all genes (as shown in previous chapters) for the oyster (A) and flatworm (B) as opposed to using only single copy orthologs between the two species (C – oyster and D – flatworm) show an overall reduction in complexity and cluster number.

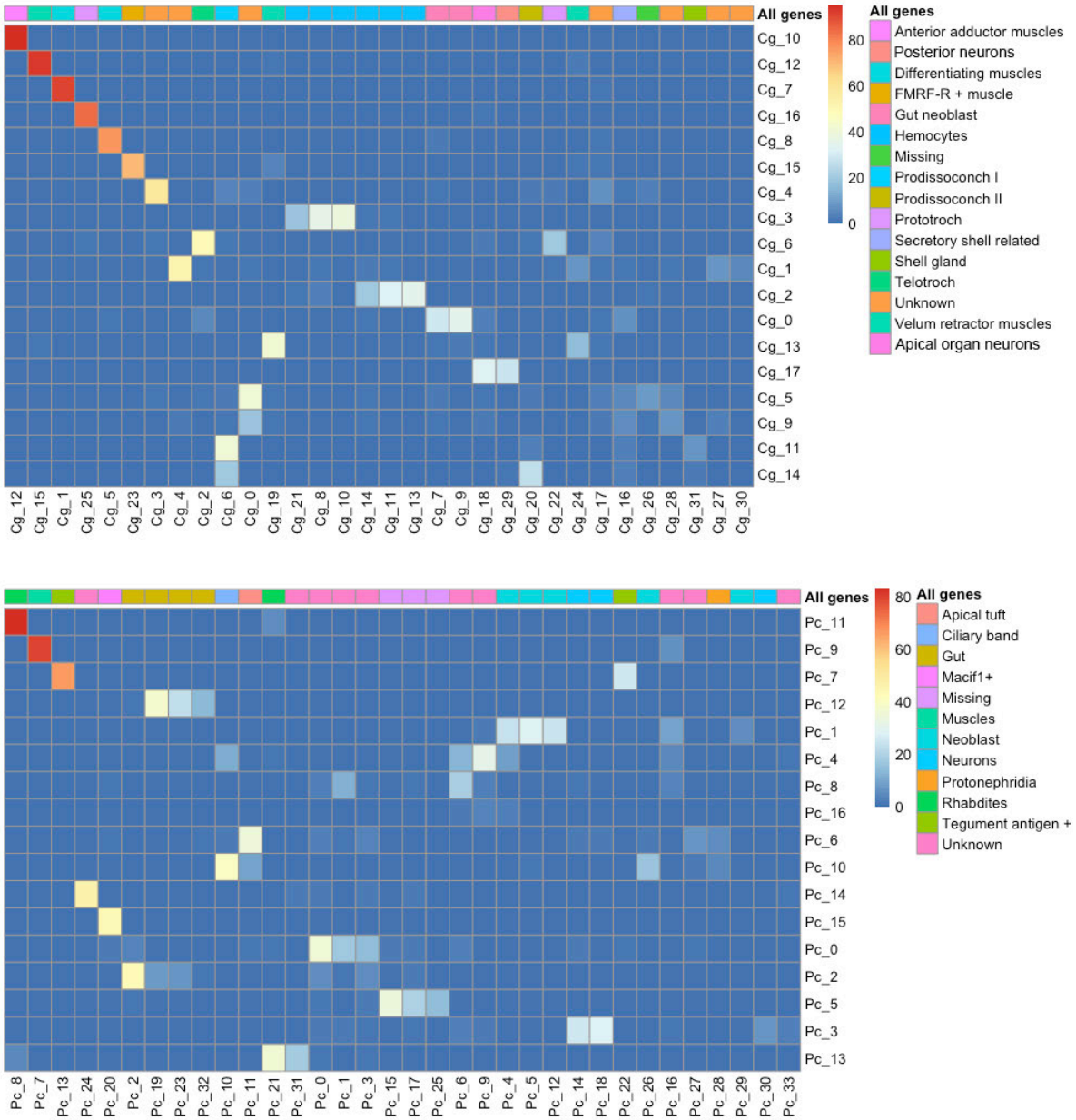


Figure 97. Using only single copy orthologs pulls cells from different clusters together.

Heatmaps showing the percentage of cells that remain in the same cluster when using all genes (columns) vs only single copy orthologs (rows). The oyster datasets on top appear more similar in the two clustering treatments than the flatworm, however, for both, we observe an overall loss of complexity with many clusters being pulled together and a few getting completely lost.

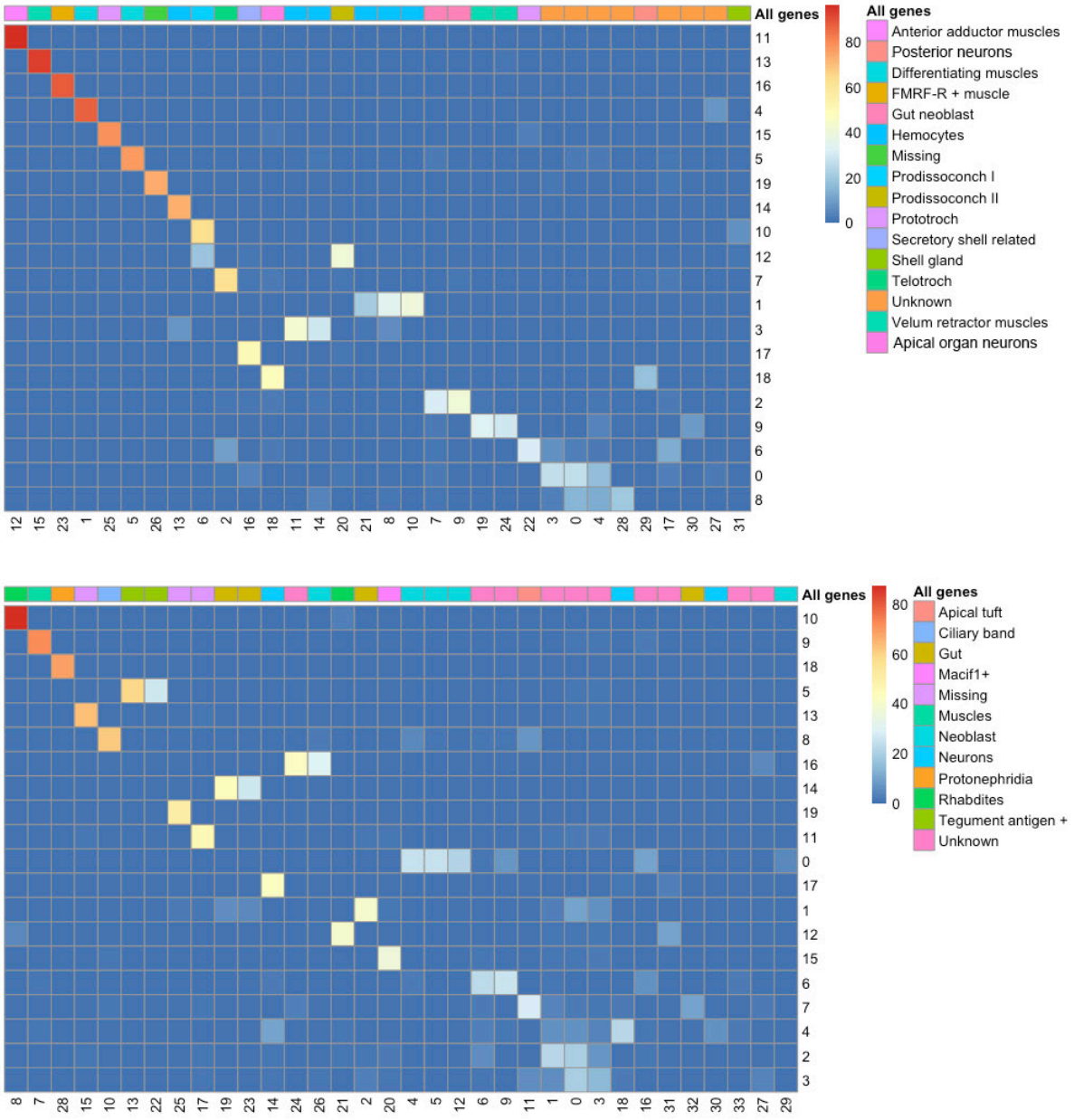


Figure 98. Single copy orthologs are less informative than the same number of randomly selected genes.

Heatmap showing the percentage of cells that remain in the same cluster when using all genes (columns) vs a subset of 5000 randomly selected genes(rows) in the oyster (on top) and in the flatworm (on the bottom). By comparing these two heatmaps with those of figure 2 we can see how randomly selected genes seem to recover more clusters than the same number of single copy orthologs.

6.1.2 Integration of cell types across species using Seurat integration

In the previous section I explored the effect of using single copy orthologs to perform cell type clustering. I showed that this leads to an overall decrease of complexity in my original datasets. However, I have also shown how most major cell types (i.e. muscles, ciliary bands, neurons, neoblasts, haemocytes etc) remain correctly identified, although often subclusters were pulled together. These results gave me some hope that integration using single copy orthologs could retain some biological sense. For this reason, I proceeded and attempted to integrate the data using the Seurat integration method.

Firstly, I made subsets of the original matrices for both species that contained only single copy orthologs and substituted the species-specific gene names with the Orthogroup name (this way both species have the same gene reference). I then performed the usual read normalization and computed the 2000 most variable feature for each dataset separately for each species as described in Chapter 3. The next step was to choose features that are variable across both datasets and use these to find “integration anchors” which are pairs of cells from each dataset that are nearest (most similar) to one another. These anchors are then used to align the two datasets in the same multidimensional space and the values of this integration can then be used for standard downstream analyses such as scaling, running the PCA, finding neighbours and finally inferring clusters (as previously described in detail in chapter 3).

The result of the integration can be seen in the UMAPs of figure 99: in total I obtained 19 clusters (with a resolution of 0.5) of which ~12 mostly contain cells from only one of the two species. To try and understand what these clusters correspond to in my old annotation I decided to use the same method as before and look at how many cells were in the same clusters before and after integration. Starting with the oyster clustering (see top of figure 100) one can see that the best matching clusters (those where more than 60% of cells remained together) after integration are clusters 13, 15 and 16 of the integration; these are clusters that only contain oyster cells. More interestingly we can see that cells originally from ciliary band clusters (Cg 2, Cg 22 and Cg 25) group together in integrated clusters 7 and 9 which include cells from both species. Similarly, neuronal cells (Cg 18 and Cg 29) cluster together in integrated cluster 8 and some muscle cells (Cg 23, Cg 5 and Cg 12) cluster together in integrated cluster 5; both these integrated clusters (5 and 8) also contain flatworm cells. Gut neoblasts are scattered between integrated clusters 2 and 10 which contain cells from both animals. Finally, all haemocytes cells are grouped together in cluster 4 which is mostly made up of oyster cells.

When comparing the original flatworm clustering with the integrated clustering I found that the best matching cluster is the muscle cluster (Pc 7) which corresponds to integrated cluster 5 together with some previously mentioned oyster muscle clusters (see bottom of figure 100). This match between the two species is reassuring but not very interesting since all Bilateria have muscles and we expect muscle cells in general to be homologous. The next three best scoring clusters are rhabdites, tegument antigen + cells and cells from cluster 24 (which remain unidentified), however, these original clusters all match integrated clusters that are made up mostly of flatworm cells. More interestingly, I found that neuronal cells

(Pc 14, Pc 18 and Pc 30) of the flatworm group together in integrated cluster 8 together with oyster neurons. As mentioned before it would have been more interesting to find matches between specific neuronal (and ideally larval-specific) types; however, as seen in the previous paragraph, we lost resolution on the different neuronal subtypes when using single copy orthologs. Another interesting finding is that cells from the apical tuft and ciliary band (Pc 10 and Pc 11) also group together with ciliary band cells from the oyster in integrated clusters 7 and 9. Finally, some neoblast clusters cells (Pc 4, Pc 5 and Pc 12) end up together with oyster gut neoblasts in integrated cluster 2.

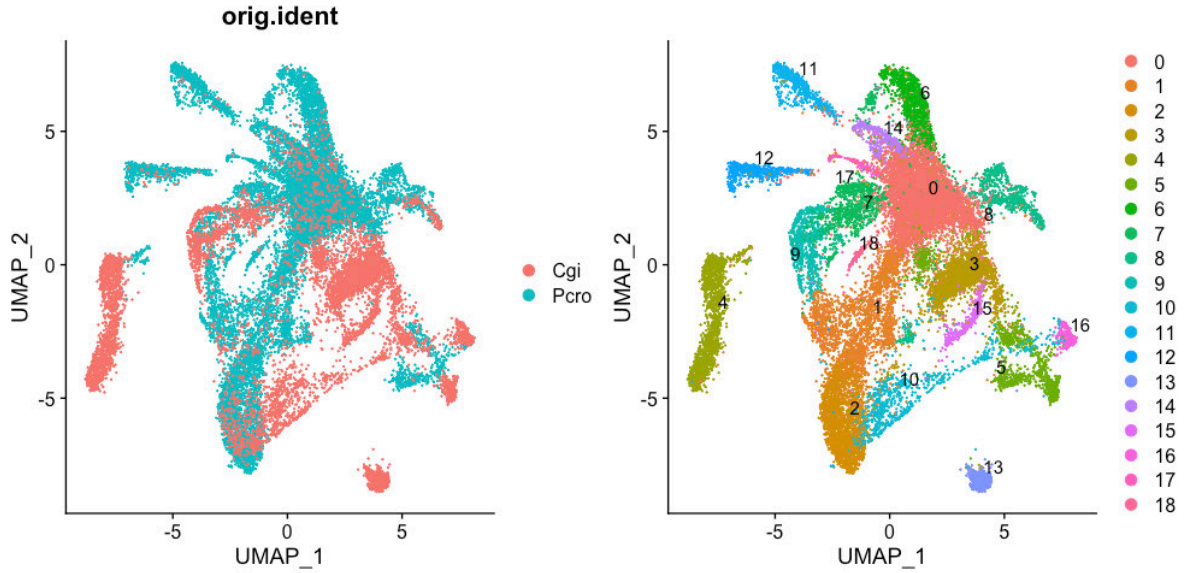


Figure 99. Cross species integration using single copy orthologs shows many specie specific clusters.

UMAPs of flatworm and oyster SCS data integrated using only single copy orthologs and Seurat Integration pipeline show many clusters are species specific (all red or all blue) with only a few clusters containing cells from both species.

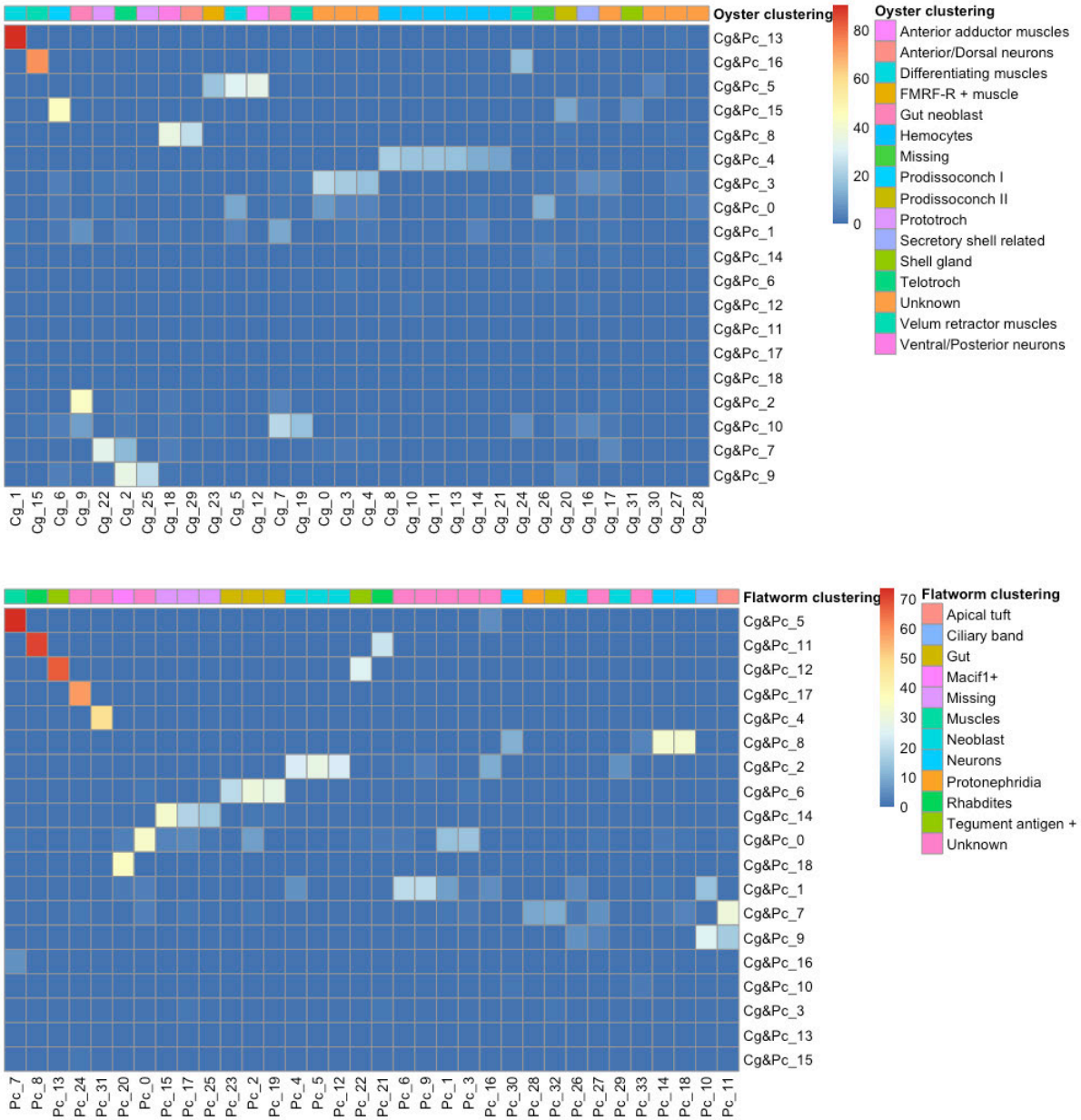


Figure 100. After cross species integration using single copy orthologs cells that remain together in the same clusters are mainly from species specific clusters.

Heatmaps showing the percentage of cells that remain in the same cluster when using original species-specific clustering vs integrated clustering of both species (rows). The highest scoring matches (where most cells remain together across clustering) correspond to integrated clusters with cells only belonging to one species except for: muscles (integrated cluster 5), ciliary bands (integrated clusters 7 and 9), neurons (integrated cluster 8) and neoblasts (integrated cluster 2). Suffix Cg indicates oyster clusters, Pc indicates flatworm clusters whilst Cg&Pc indicates integrated clusters.

6.1.3 Genes supporting common integrated clusters

In the previous sections I have shown the effects of using only shared single copy orthologs to cluster single cell sequencing data from my two larvae and used these genes to integrate the two datasets. I found only five clusters that contain cells from both species and these are: muscles (integrated cluster 5), ciliary bands (integrated clusters 7 and 9), neurons (integrated cluster 8) and neoblasts/proliferative cells (integrated cluster 2). As briefly mentioned above, these matches are in a sense not very surprising since we expect common cell types such as muscles, neurons, ciliary cells and proliferative cells to be homologous across all Bilateria (at least). In fact, even If these two larvae had evolved independently, they would likely have co-opted these cell types from the adult stage. To show these larvae are homologous I would need to find distinctively larval specific cells matching across species rather than very general cell types. Since using single copy orthologs seems to reduce the complexity of my original datasets greatly (for example pulling together all neurons) it is likely that this result is a limitation of the method itself rather than a real biological signal.

However, It is still possible that the transcriptional profile of these shared clusters can tell us more about the evolution of these cells. To try and understand what sort of genes were driving these similarities I decided to extract markers from the integrated dataset for these four matching clusters. Figure 101 shows the expression of the top markers of integrated clusters of interest in the original SCS datasets. Looking at the expression in the original clustering in each species tells us that these genes are actually expressed in muscles, neurons, ciliary bands and neoblasts of both animals before proceeding with any further

analysis. Since most of these are indeed expressed specifically in both species I decided to try and see if A) any of these shared genes are transcription factors and B) what phylostrata all shared genes belong to. I wanted to look at transcription factors because these would tell us whether similar cell types are regulated in the same way (which is less likely to arise convergently) and at the phylostrata to see if any genes were lophotrochozoan specific. To look for TFs I ran Orthofinder on the oyster and the flatworm together with a list of TFs from human and *Drosophila*. I then retained all genes of my species that belong to an orthogroup that contained at least one TF from either animal. I only found two TFs for the muscle cluster (IRX and VSX) and one for the neoblast (histone-lysine N-methyltransferase ASH1L-like). *Irx* genes are common across Metazoa and have possibly undergone several lineage specific duplications, for this reason it can be hard to establish orthology across different phyla (for more information see the tree in figure 102 and see Kerner et al, 2009). In general, *Irx* genes encode transcription factors involved in many developmental processes, ranging from muscle to sense organ development however, their role has not yet been described in any lophotrochozoan. *Vsx* (visual system homeobox) genes, are usually expressed in neuronal or eye cells, as their name suggests and in fact they appear expressed in the neuroepithelium of *Platynereis* larvae (Denes et al, 2007) (for gene tree of *Vsx* see figure 103). Finally, ASH1-like is described in the literature as being expressed in neuronal progenitors in both *Capitella* and *Platynereis* (Meyer & Seaver 2009) (for gene tree of ASH1-like see figure 104). Altogether, so few TFs appear to be specifically expressed in the same cell types across these groups of matching cells that it's hard to draw any conclusions but it would be interesting to see if, for example, *Irx* expression was concentrated in the muscles of other larvae. As previously implied, I believe that these results are limited by looking only at single copy orthologs and that a more complex method using all orthologs could

potentially be more informative especially since, among single copy orthologs, I only recover about 250 TFs (out of a total of ~700 in each species).

Phylostratigraphy analysis of marker genes for these four different cell types (muscles, cilia, neurons and neoblast) shows that most genes are ancient with the most recent belonging to the Bilateria phylostratum (see figure 105). In a sense it is not surprising to find no lophotrochozoan specific genes considering that there are only 3 lophotrochozoan and 21 spiralian single copy orthologs between my two species according to my phylotratigraphy analysis (see figure 106). In chapters 4 and 5 I had already shown that very few genes in each species seem to belong to lophotrochozoan, spiralian and protostomes clades so it makes sense that even fewer would be shared across the two species. I am not sure what is causing this, but I can hypothesize that it could be a technical limitation due to the lack of good transcriptomes and genomes for lophotrochozoan clades or that it could be caused by a relatively short branch leading to lophotrochozoan clade. In general, I think that the main result of this integration analysis is that using single copy orthologs oversimplifies the complexity of the single cell sequencing data and is not useful to use when comparing species across such evolutionary distances.

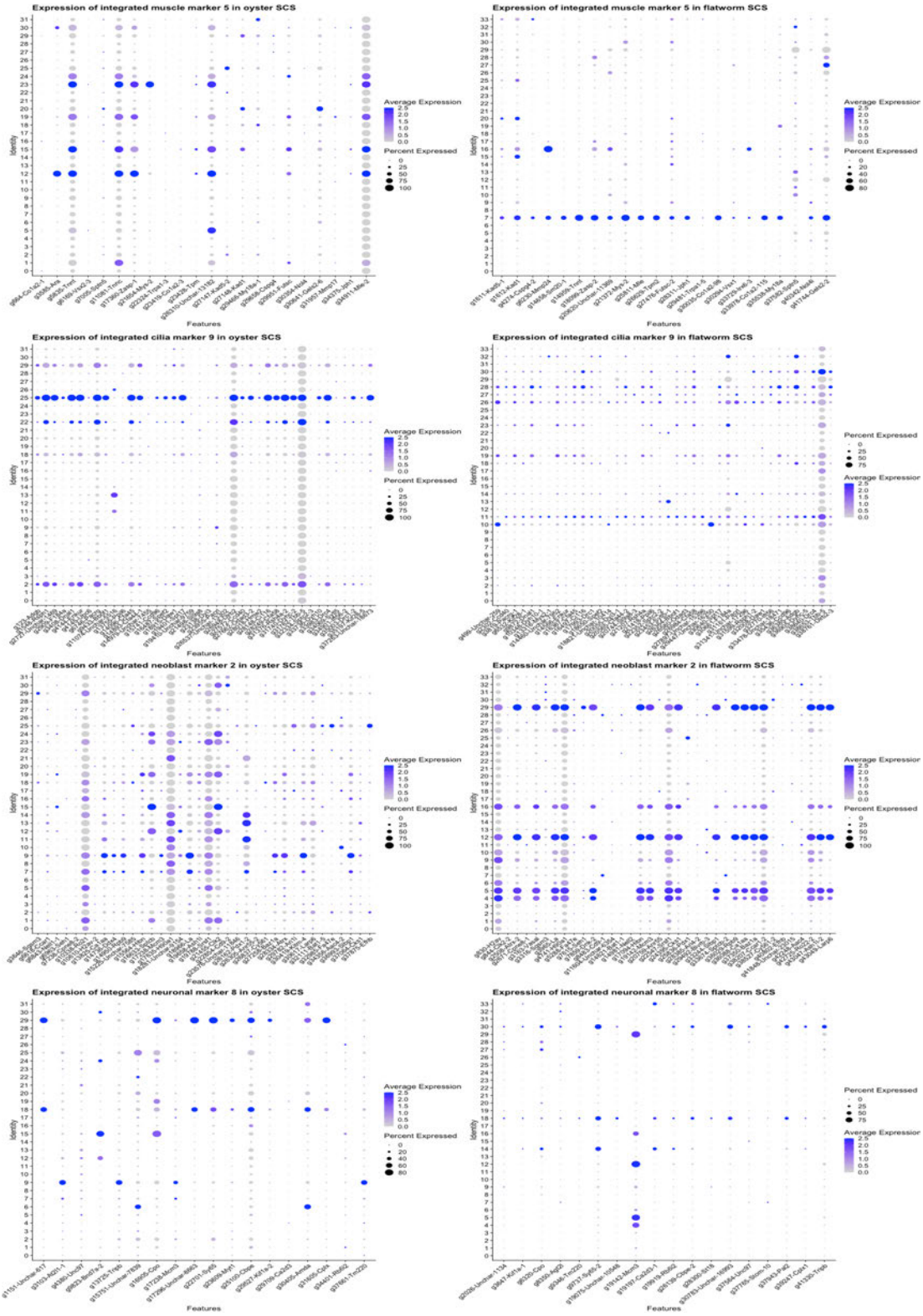


Figure 101. Shared single copy marker for muscle, cilia, neurons and neoblast show specific signature.

Dotplots showing the expression of shared single copy orthologs in different cell types (muscle, cilia, neurons and neoblast) in both species (oyster on the left and flatworm on the right) demonstrate the specificity of these genes.

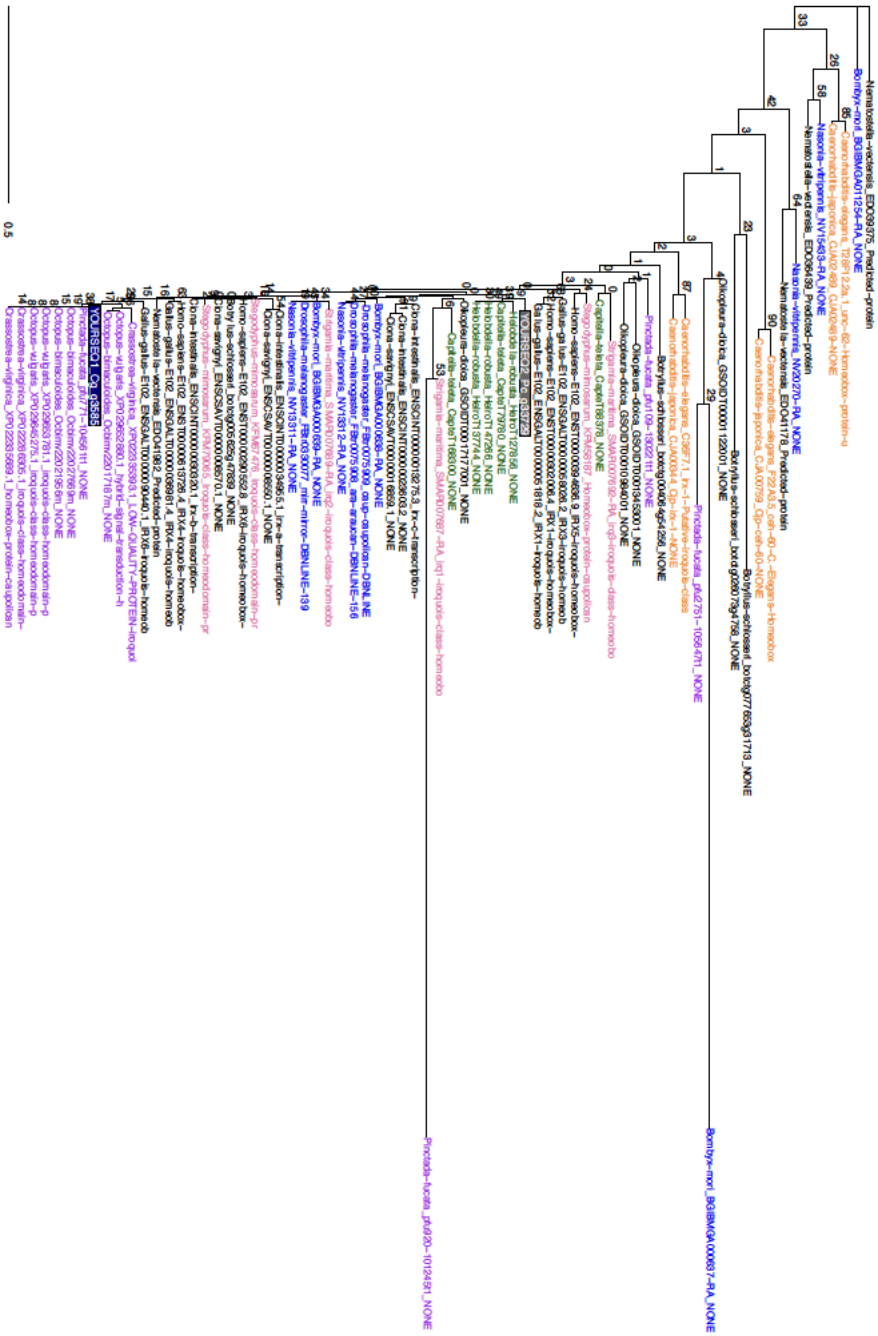


Figure 102. Irx genes show lineage specific duplications.

Irx tree showing single copy orthologs from the oyster (Cg_g3585) and flatworm (Pc_g33729), Irx genes tend to have linage specific duplication so it is hard to tell what Irx paralog my sequences are most likely to be orthologous to.

Figure 103. Vsx gene tree showing the position of the oyster (Cg_g6169) and the flatworm (Pc_g30294) genes.

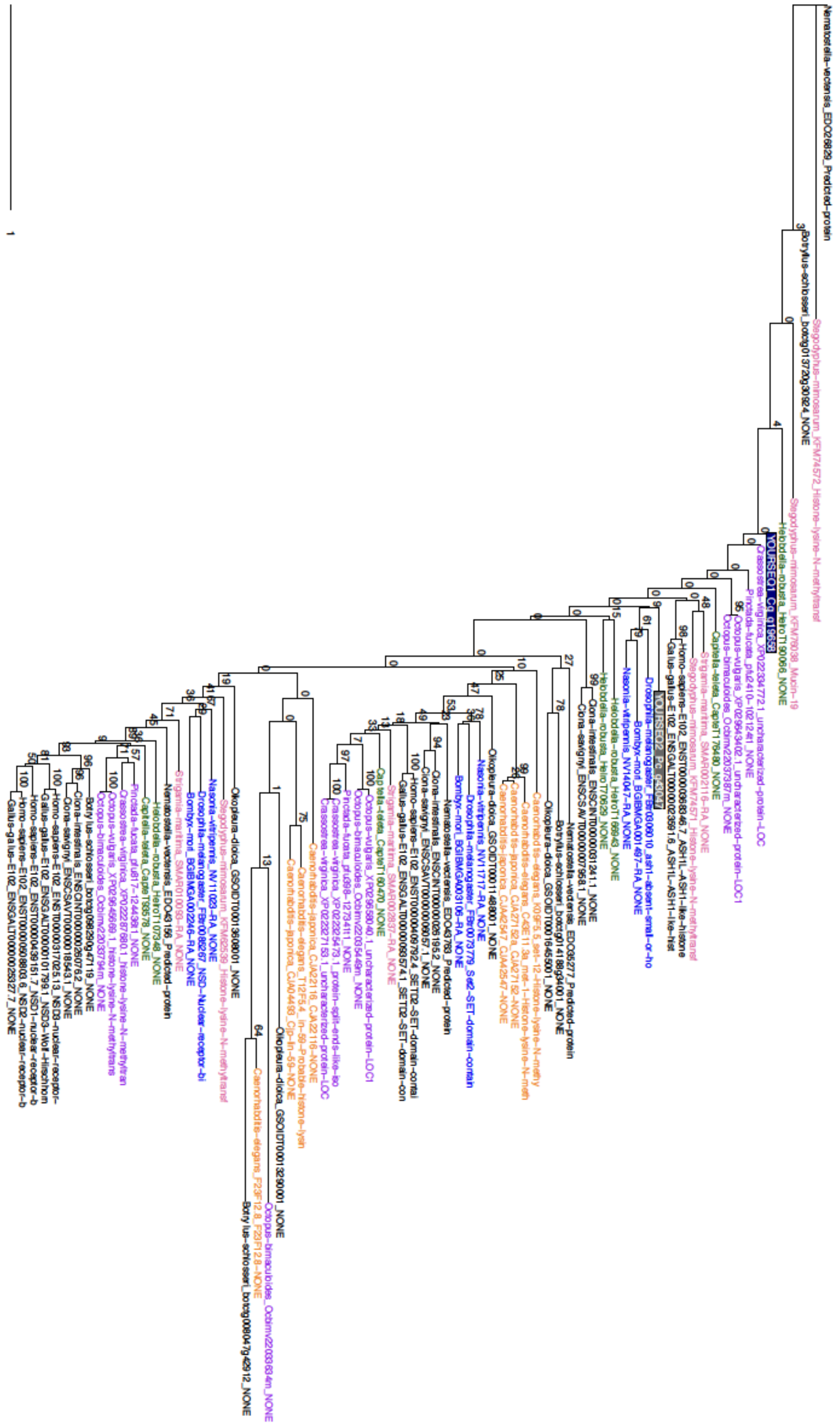


Figure 104. Ash1-like gene tree showing the oyster (Cg_g19658) and the flatworm (Pc_g43047) genes.

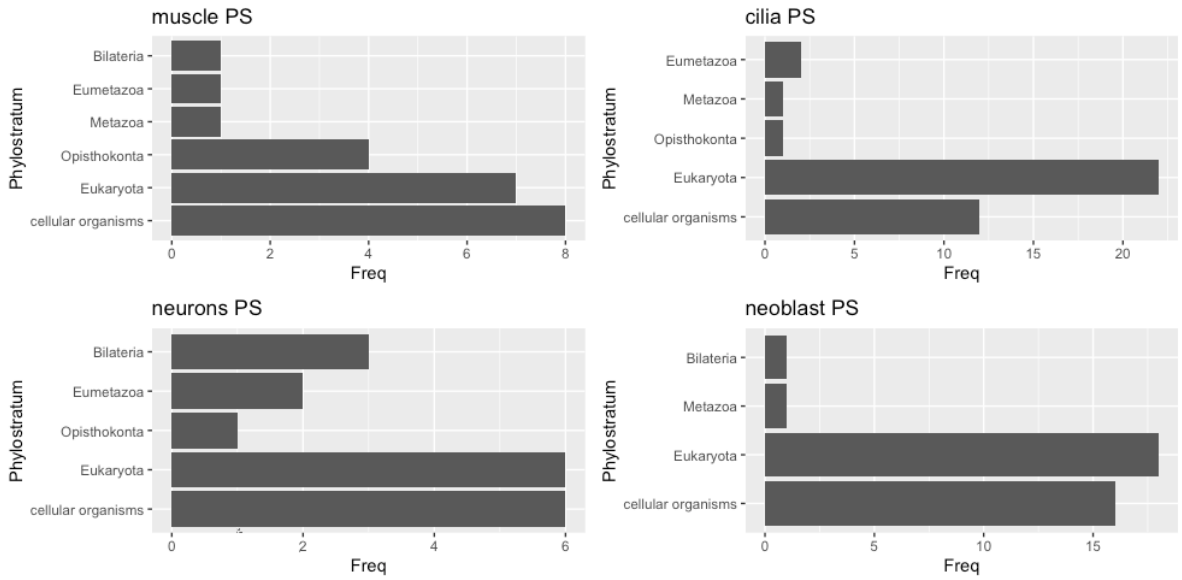


Figure 105. Marker genes for each group of matching cells divided by phylostrata showing most genes are quite ancient with the most recent being Bilateria specific.

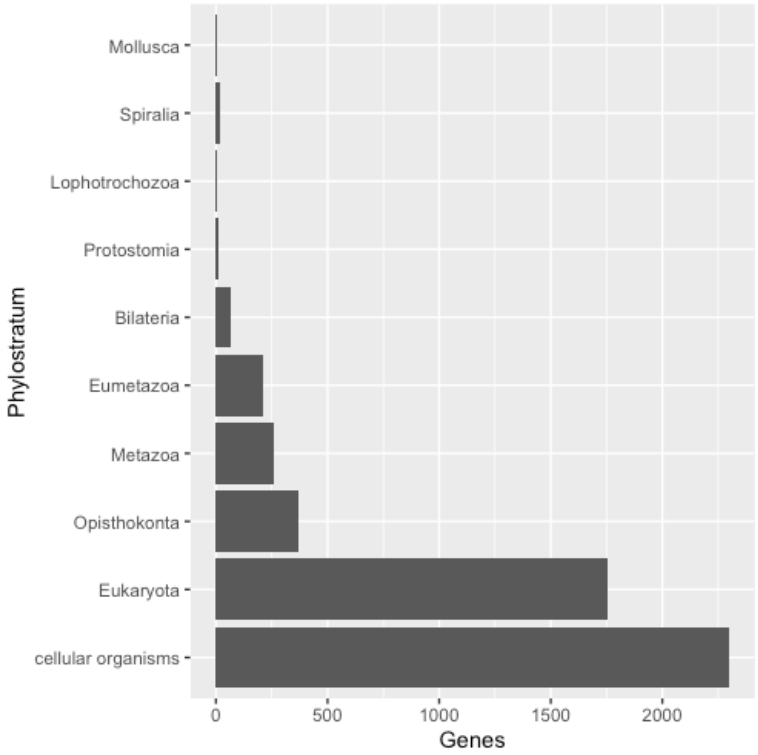


Figure 106. Distribution of single copy orthologs among the different phylostrata highlights that the vast majority of genes belong to older phylostrata

Clearly there is a discrepancy between our phylostratigraphy analysis (which uses blast) and the single copy orthologs list between the oyster and the flatworm (produced using Orthofinder) since some single copy genes appear to belong to the Mollusca phylostratum, however these are only 2 genes.

6.2 Comparison of cell types using SAMAP

In the previous sections I attempted to carry out a comparison between my two single cell sequencing datasets using only single copy orthologs. However, as discussed above, this approach oversimplifies both the overall cell clustering and the transcriptional signature of different cell types and so doesn't give very meaningful results. This loss of complexity leads to grouping together of cells with similar identity (i.e. all neurons, all cilia, all muscles etc) and drastically reduces the pool of genes differentially expressed per cell type, leading to recovery of very few TFs and lophotrochozoan specific genes which could potentially give us more information on cell type homology. For all these reasons, I was eager to try a different method that would allow me to use a larger number of genes hopefully to retain more complexity in cell types and transcriptional signatures. The second method I tried is a newly developed tool called SAMAP (Tarashansky et al, 2021). The SAMAP algorithm works in two steps, first it creates a graph connecting genes from species A to similar genes in species B, weighting these connections according to protein sequence similarity (using BLAST). This gene-gene graph is then used to project each single cell sequencing dataset onto a joint manifold space which has lower dimensions than the two original data sets. At this point SAMAP looks for similar neighbourhoods of cells across the two species to anchor the two datasets together. The use of cell neighbourhoods (which are basically similar to cell clusters) as anchors as opposed to using single cells (which is what the Seurat Integration does) makes sense when comparing cells across not only different species but, as in our case, different phyla. It is likely that the two animals will have similar cell types but not very likely that they would have the exact same cells, which, in my opinion, makes the Seurat

integration more useful when comparing different conditions or experiments in the same species.

Once this joint space is produced, SAMAP calculates the expression correlation between homologous genes and uses this to reweight the edges of the gene-gene graph which will be used a subsequent SAMAP iteration. The process continues until the cross-species mapping is not modified by the next iteration. Among the many outputs of the SAMAP algorithm, the most useful to me are an alignment score between the clusters in each species and a list of gene pairs (one per species) that are co-expressed in the matching clusters.

A heatmap of alignment scores are shown in figure 107 which shows several blocks of "aligned" clusters across the two species. The alignment score is calculated as the average number of cells from the other species that are neighbors to a particular group of cells relative to the maximum possible number of neighbors, so it's a measure of similarity between two cross-species clusters.

Among these "blocks" we can find some similar patterns to the previous integration such as a match between muscle clusters (B), two blocks of proliferative cells/neoblasts (C and D), one block of matches between ciliary bands (J) and two separate neuronal matches (H and I). This integration appears to maintain more complexity in the comparison compared to the previous one (for example we see two different neuronal matches). Moreover, we recover many additional matches that we did not see with the previous analysis such as several blocks of matches between haemocytes and secretory cells (E, F), a match between the flatworm gut and cluster 26 of the oyster (A) and one between the flatworm protonephridia

and oyster cluster Cg 16 (K). However, to get a real idea of whether any of these matches make biological sense we need to take a look at the genes that are expressed in the potentially equivalent cell types.

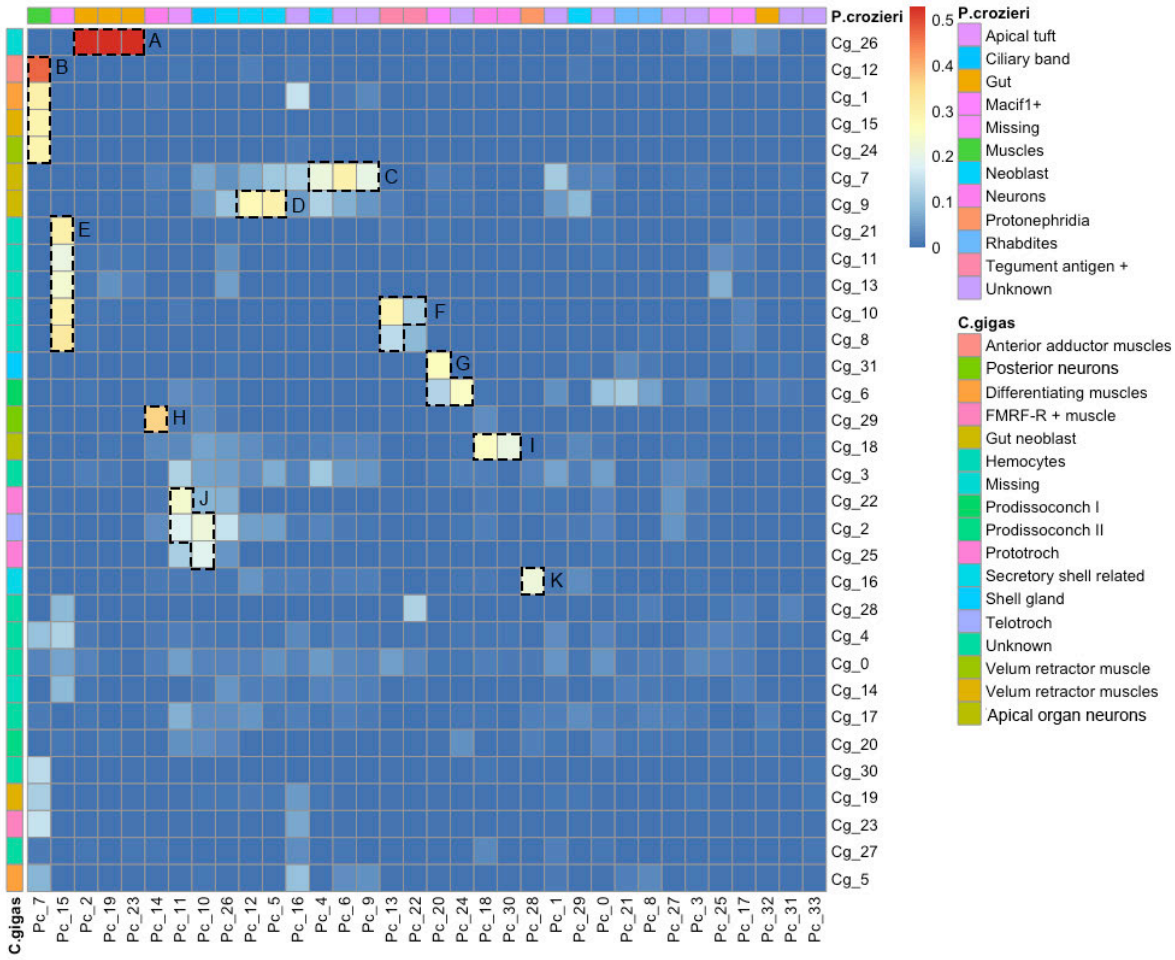


Figure 107. Alignment scores for matches between the oyster clusters (rows) and the flatworm clusters (columns) shows several blocks of matching clusters.

These include: muscles (B), neoblasts (C & D), ciliary bands (J), two distinct neuronal groups (H,I), haemocytes (E,F,G), a match between the flatworm gut and cluster 26 of the oyster (A) and one between the flatworm protonephridia and Cg 16 (K). Cg_ indicates oyster original clusters (as described in chapter 4) Pc_ indicates flatworm original clusters (as described in chapter 5).

6.2.1 Number of co-expressed genes per each match

The first feature of the gene pair list given by SAMAP I looked at is the overall number of genes supporting a match. This number can be useful to get rid of spurious matches that could be caused by chance by the co-expression of a small number of very similar genes in two cell types. When looking at the numbers of genes co-expressed in each match (see figure 108) we can see that there is no obvious correlation between the alignment score and the number of supporting genes. For example, the group with the highest alignment score, group A (a match between the gut cluster of the flatworm with the, presumed neuronal, cluster 26 of the oyster), appear to have few co-expressed gene pairs compared to others with lower alignment scores (such as B or D), however it is worth nothing that group A still has between 30 and 90 co-expressed gene pairs. It could be that, despite its localisation, the oyster cluster 26 is composed of the first differentiated gut cells and that for this reason it matches the gut clusters in the flatworm. More likely, this result could be caused by a spurious co-expression of some genes (for example cathepsin-L is a marker for both species), made worse by the low coverage for cluster 26 of the oyster (see chapter 4) and possibly by the fact that cluster 26 has very few cells.

Similarly, group G has very few genes supporting the matches; in this case as well it is likely that the match was caused by spurious genes considering that the oyster clusters in these matches are shell gland related cells that we do not expect to find in the flatworm.
Lastly, group K, matching the protonephridia of the flatworm with a not very well characterised cell group in the oyster also present a smaller number of genes supporting the match. In this case, similarly to group G, it is possible that the match is caused by a few

similar genes considering that the oyster larva does not appear to have protonephridia at the trochophore stage.

For the remaining matches, it is striking to see that we find as many as 200 genes (sometimes even more) co-expressed between cell types of the two species, but we cannot directly tell whether any of these cell types are homologous based simply on that information. The genes are so many that it is not feasible to explore the list one gene at the time, especially since lophotrochozoans are largely understudied and most genes lack any functional description in any member of this group (let alone at the larval stage).

For this reason, I decided to look at how many of the co-expressed genes made up the original transcriptional profile of a certain cell type. That would tell me whether the shared genes are just a small subset of what makes up the identity of a certain cell type or not and ultimately could indicate whether any of these cell types are homologous. I looked at the percentage of marker genes for a specific cell type (in each of the two species) that were also co-expressed in the matching cell type of the other species. Most strikingly, for both species and for most groups, more than 20% of marker genes were co-expressed. This was not true for groups A, G and K which I have described as likely spurious matches earlier on (see figure 109).

However, I found a clear difference in overall percentages between the oyster and the flatworm with the flatworm showing higher percentages for any given match (in figure 109 notice how most percentages for the oyster on top are lower than the flatworm on the bottom). This could in part result from the oyster sample having a higher median number of

genes per cell (~1100) than the flatworm (~400) which impacts on the total number of marker genes per cluster in the flatworm (median 24) compared to the oyster (median 110).

Since supporting genes are in pairs (one per species) but there are many more marker genes in the oyster the percentages will be lower than in the flatworm. I have tried to reduce this difference by sequencing the flatworm samples in greater depth (see chapter 5) however this caused the number of cells to double, but led to only a slight increase in the number of genes per cell. Assuming that this is indeed the reason for the differences in overall percentages between the two species, we can imagine that, with a higher number of genes per cell detected in the flatworm, we would expect to see a similar reduction in percentages.

Regardless, it is clear that for certain cell types such as muscles (B) and proliferative cells/neoblasts (C-D) the co-expressed genes make up a high portion of the overall transcriptional signature of the cell types. As mentioned in previous sections, we would expect those cell types to predate Lophotrochozoa and to likely be homologous in all Bilateria.

Integrated groups E and F both represent a match between oyster haemocytes clusters with the flatworm gut cells clusters 15 and the tegument antigen + cluster. Seeing how more than 50% of co-expressed genes are markers for cluster 15 (gut) and 25% are markers for the tegument antigen + cluster it is possible that these two flatworm clusters are indeed made up of hematopoietic cells although not much is known about these cells in flatworms (let alone in flatworm larvae). It is also unclear how many different types of immunocytes,

haematocyes or hematopoietic cells are present in protostomes in general and whether they are homologous to each other and to vertebrate immune and endocrine cells even though some authors believe they could be (Malagoli et al, 2017). Regardless of this debate, it is likely that none of these cells are larval specific and hence not very interesting for our comparison, similar to muscles and proliferative cells.

The more exciting result concerns integrated groups H, I and J which are matches between neuronal clusters and between ciliary clusters. Group H is a match between the neuronal clusters Cg 29 and Pc 14 and for both larvae these clusters are made up of cells located in the posterior part of the larva (the oyster posterior neurons and some neurons on the lobes of the flatworm larva as well as a couple of neurons located apically). 25% of the oyster markers for cluster 29 and 65% of the flatworm markers for cluster 14 are co-expressed between these two cell types and, as one can see in figure 110, all co-expressed genes identified by SAMAP show very specific expression in these clusters. These two clusters are promising candidates for larval specific cells since, as one can see in figure 14, relatively few of the shared, co-expressed genes are also expressed in other neuronal clusters of either species indicating that the shared genes are not part of a general neuronal signature but specific to these cell types.

The other neuronal match between species (integrated group I) is between cluster 18 of the oyster (apical organ neurons) and clusters 18 (apical organ and scattered neurons) and 30 (neurons connected to the cerebral eyes) of the flatworm. For these matches the percentage of marker genes that are also co-expressed are slightly lower than for the previous neuronal match: for the oyster about ~15% and for the flatworm ~25%. Moreover,

when I took a look at the expression of the shared genes I noticed that for the oyster most co-expressed genes of cluster 18 are also expressed in cluster 29 (see top of figure 111) and the same holds true for co-expressed genes in cluster 18 (see middle of figure 111) of the flatworm so it seems that these could be general neuronal markers. However, the co-expressed genes between cluster 18 of the oyster and cluster 30 of the flatworm are quite specific for cluster 30 (see bottom of figure 111). It is possible that these matches are just caused by general neuronal markers or that indeed we see co-expression of several markers in the apical organ of the two larvae.

The remaining matches are between ciliary cells (integrated group J) which contains the three ciliary clusters of the oyster (2 prototroch cell types Cg 22 and Cg 25 and one telotroch cell type Cg 2) and the two ciliary clusters of the flatworm (Pc 10 which contains cells from the ciliary lobes and Pc 11 which also includes cells from the apical tuft). In this last group we see a quite striking shift, with the oyster having a higher percentage of markers being co-expressed genes (between 25% and 50%) than the flatworm (between 5% and 25%). Similar to what was seen with the previous neuronal match, there doesn't seem to be a 1:1 preferred cell match among these ciliary clusters. For example, in the oyster, the transcriptional signature of the prototroch cluster 22 is made up by almost 50% of the genes shared with the apical tuft and ciliary lobes of the flatworm larva. In the flatworm, on the other hand, these genes only make up 5% of the transcriptional signature of apical tuft and ciliated lobes.

These slightly different results in the two species may have been caused by the difference in numbers of gene markers in the two species (for example, Pc 11 only has 16 markers and Cg 25 has 623) which could be sorted with a deeper sequencing.

In an attempt to try and understand what was causing the differences between the matches in the two species I decided to plot the expression of the co-expressed genes between the various matches (see figure 1 and 2 of the appendix). What I found is that genes co-expressed between the oyster ciliary cells and cluster Pc11 (ciliary lobes and apical tuft) were very specific to these ciliary clusters in both species. However, genes co-expressed between the oyster ciliary cells and cluster Pc10 (ciliary lobes) were often also expressed in neoblast clusters of both species. Since cells from cluster Pc 10 appear to branch off from the neoblast clusters (see figure 96B) it is possible indeed that this cluster is made up of differentiating ciliary cells.

What would be exciting in this case, which holds true to all the other general matches we recovered, would be to compare these cell types to adult cell types to see if their transcriptional profile matches are comparable to the larval ones. In short, it would be interesting to know if ciliary cells in the adult (that likely carry out a very different function than in the larva) are similar to the larval ciliary cells and across species. Should we find that larval ciliary cells have a distinct signature to ciliary cells of the adults and that this is shared across phyla, it would be a strong indication for a common origin of larval cells and hence of larvae.

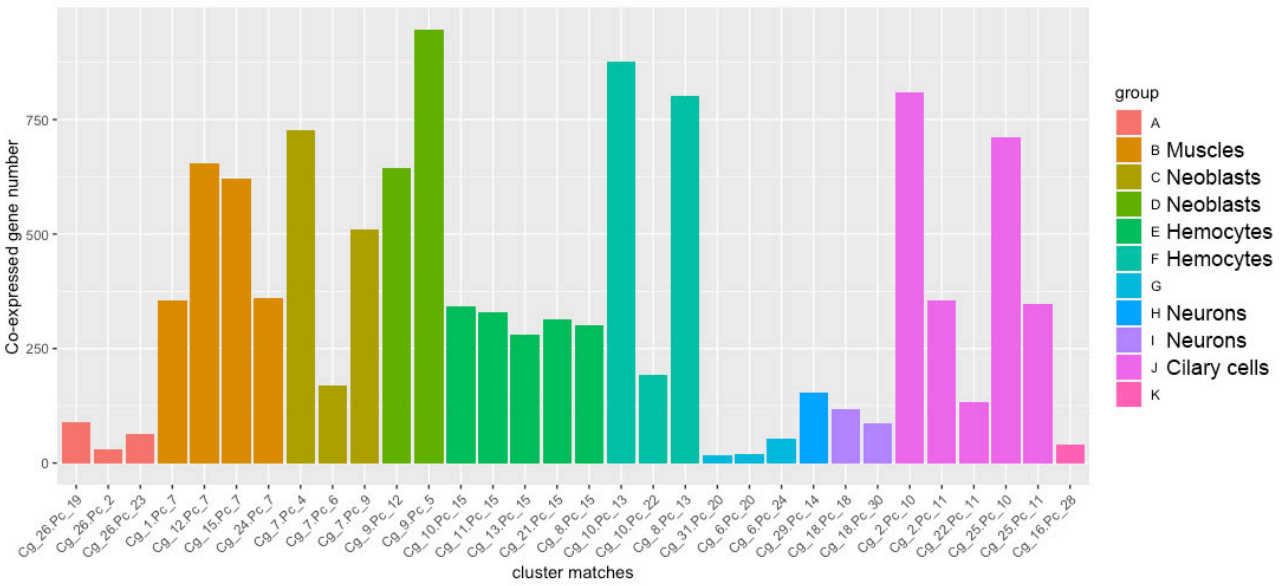
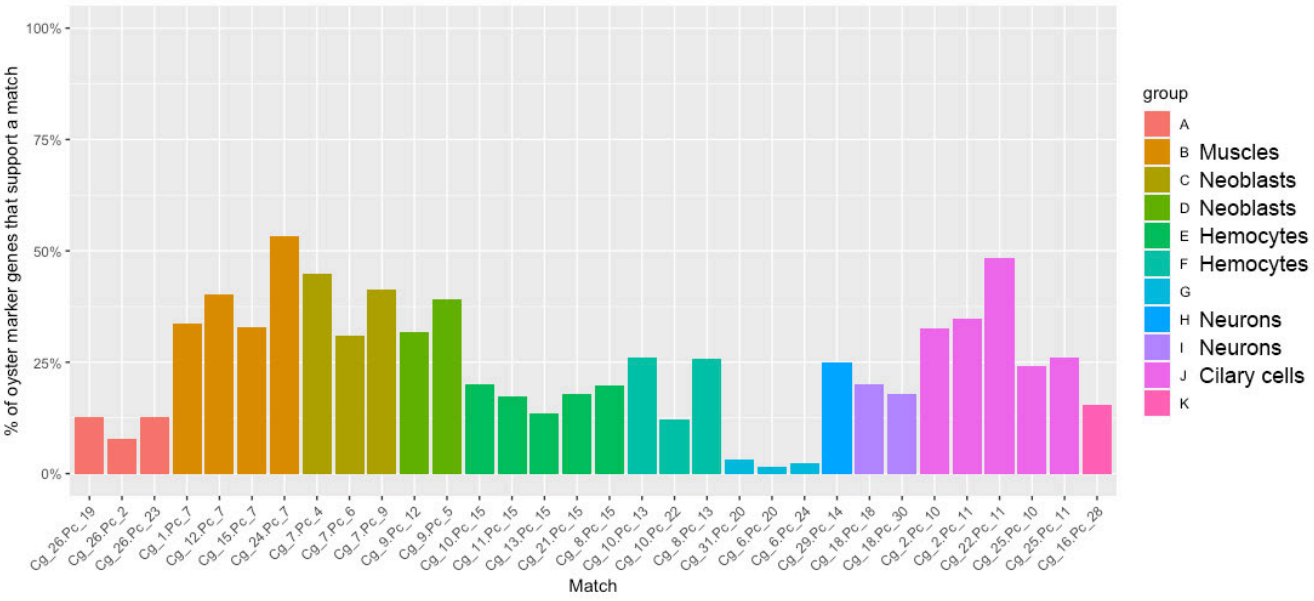


Figure 108. Alignment score is not always correlated with high number of co-expressed genes.

Graphs show number of gene pairs (top) that are co-expressed between different matching clusters of the oyster (Cg) and flatworm (Pc). The graphs show the alignment score is not directly correlated to the number of co-expressing genes and that some groups of matches appear not to be backed up by co-expression of many genes (A, G, K) (compare with heatmap in figure 105).



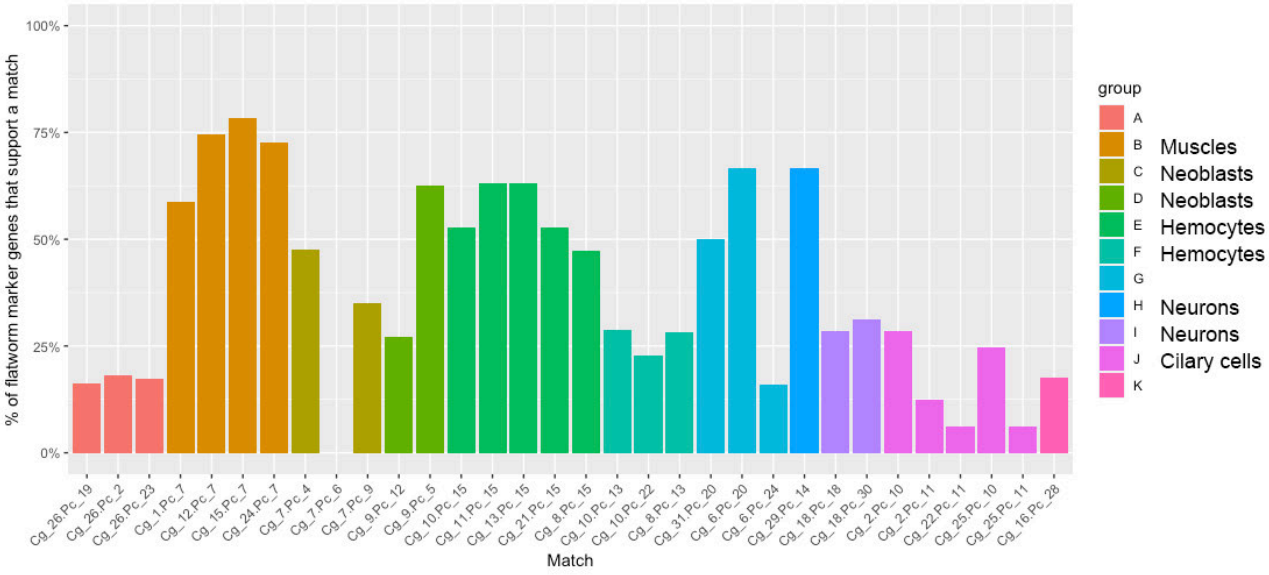


Figure 109. Percentages of gene markers that support any specific match in the oyster (on top) and the flatworm (on the bottom) make up a big proportion of overall markers (especially for the flatworm data).

Differences in overall percentages between the two species (oyster % are lower than flatworms) is likely due to under sequencing of the flatworm dataset.

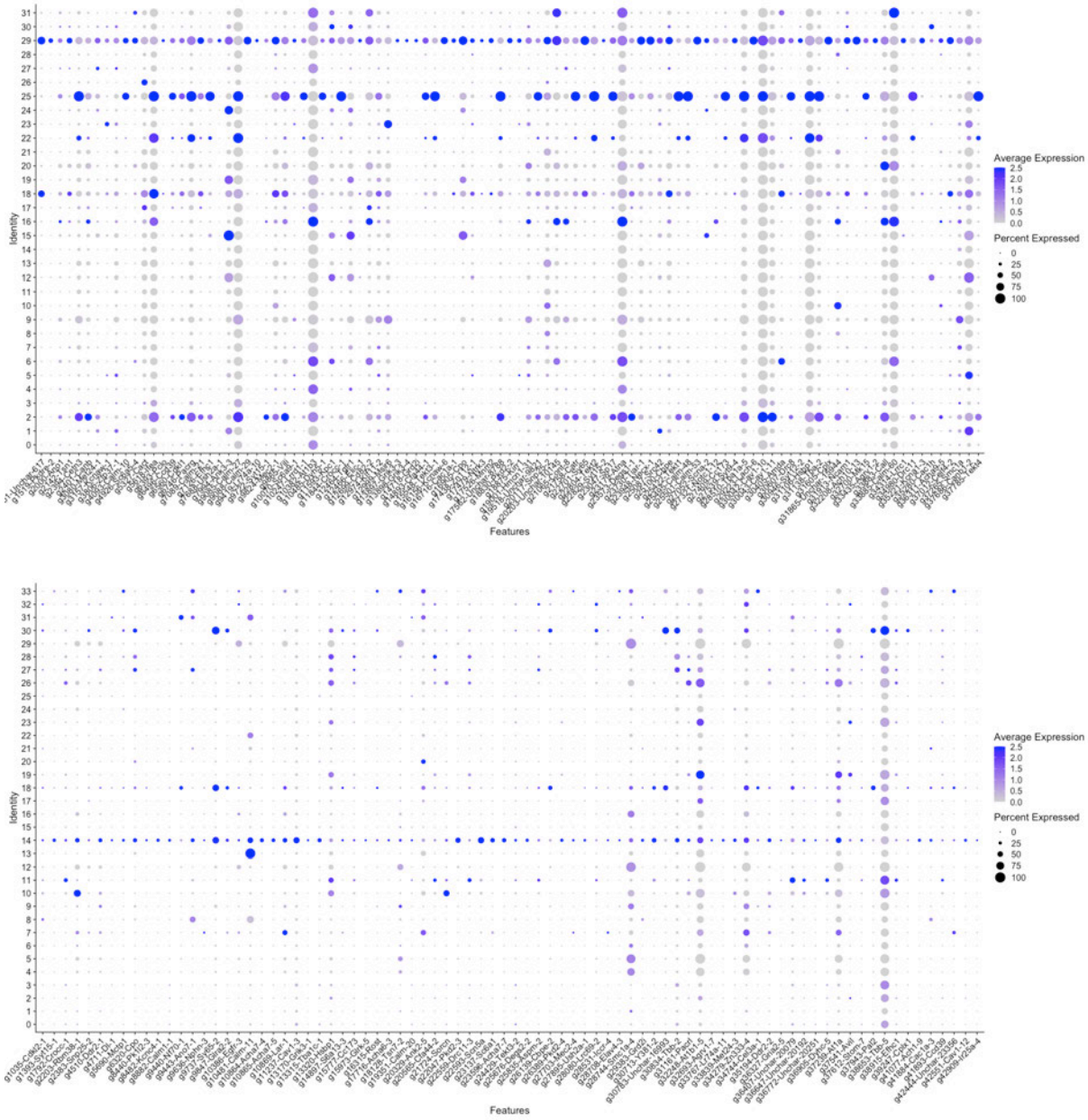


Figure 110. Genes that are co-expressed between cluster Cg29 and Pc14 are very specific to those clusters and do not show specific expression in other neuronal clusters.

Dotplots of co-expressed genes of the oyster neuronal cluster 29 and the flatworm cluster 14 show very specific signatures in cells belonging to cluster Cg 29 (top) and Pc 14 (bottom) in the original datasets and do not represent a general neuronal signature.

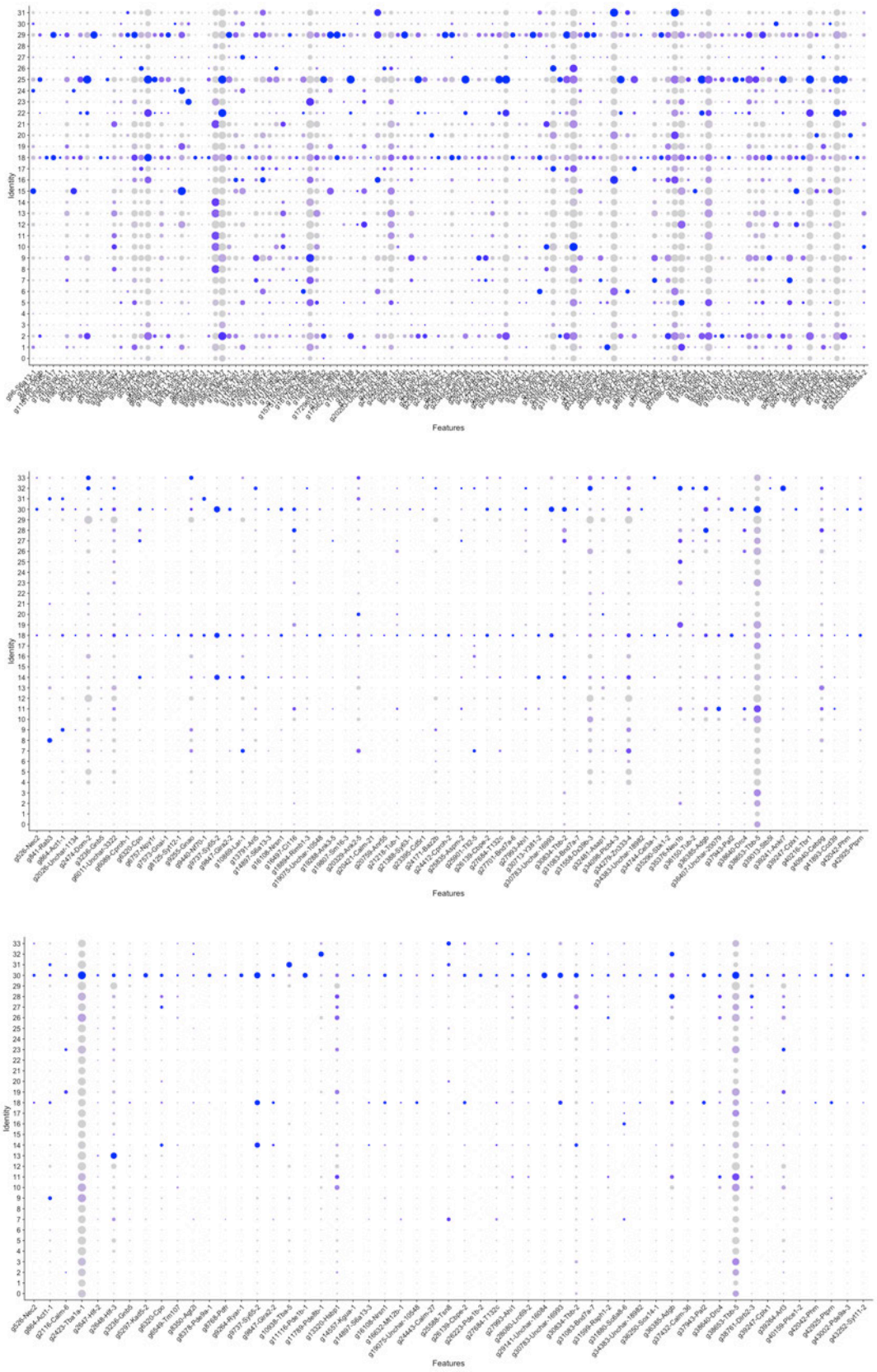


Figure 111. Genes that are co-expressed between cluster Cg18, Pc18 and Pc30, show expression in other neuronal clusters.

Top: Co-expressed genes of cluster Cg18 and Pc 18 in the oyster are expressed in both cluster Cg18 and Cg29. Middle: co-expressed genes in cluster Cg18 and Pc18 in the flatworm also show expression in cluster 30. Bottom: co-expressed genes between cluster Cg18 cluster Pc30 appear specific for cluster 30. These co-expressed genes could represent general neuronal markers.

6.2.2 Number of co-expressed TFs per each match

In the previous section I looked at how many co-expressed genes supported each cell type match and at how specific they were. This helped me understand which matches were caused by a small subset of genes and which might be more biologically meaningful. This highlighted a general match between muscles, neoblast/proliferative cells, neurons and ciliary bands of the larvae but also a very specific match between two posterior neuronal cell types populations. I was interested in seeing how many of these supporting genes were TFs, this is to understand whether matching cell types not only expressed similar proteins but were also similarly regulated after ~500 million years of evolution. Clearly, from the data that I acquired I can only comment on co-expression of TFs across species and not on their concerted regulation nor on the overall gene regulatory network of different cell types although this would be something interesting to explore in the future.

When looking at the number of TFs in each match I noticed a very similar distribution to that of all genes (i.e. groups that co-expressed more genes also co-express more TFs) (compare figure 108 with figure 112). Moreover, both species appear to co-express several TFs (~15) in muscles, neoblasts/proliferative cells and ciliary bands. Fewer TFs appear to be co-expressed across neurons (<10). Since the overall number of co-expressed TFs per match is quite large and SAMAP should have already established orthology across gene pairs I

identify the co-expressed TFs in the oyster only (which is more likely to have annotated transcription factors anyway).

Across all muscles matches (group B) I found 38 oyster TFs. When blasted their best hits (when significant) were: homebox protein Meis-like 1 (g32292), HAND2 (g32928), homeobox protein Mohawk (g33029), PBX1 (g33989), SON (g34076), IRX-6 (g3585), zinc finger protein (g36011, g14947, g16173), zinc finger protein basnuclin-1 (g36634), FoxK2 (g37786), FoxF1 (g15397), FoxG1 (g20400), ZIC-4 (g3925, g3926, g3927), zinc protein Gfi-1b (g629), HMGB-like (g8440), lysine specific demethylase 5a (g9271), protein dead ringer homolog (g10822), paired mesoderm homeobox protein 2 (g12591), histone-lysine N-methyltransferease 2A (g17752), KRAB (g18618), protein atonal homolog 8 (g1980), neurogenin 1 (g20968).

Across all neoblast/proliferative cell matches (group C and D) I found 31 TFs and their best hits were: proliferation-associated protein 2G4 (g12771), SWI/SNF (g15608), ING5 (g15909), ING1 (g88), zinc finger protein (g16173), histone-lysine N-methyltransferease 2A (g17752), Sox-2 (g17930), SALL 1 (g22184), bromodomain-containing protein 3 (g23440), transcriptional repressor YY1 (g25454), bromodomain adjacent to zinc finger domain (g26072), FoxN3 (g26097), putative peptidyl-prolyl cis-trans isomerase dodo (g27212), FACT complex subunit SSRP1 (g29749), helicase domino isoform (g30228), homeobox CDX-1 (g31246), MTA3 (g31878), E2F5 (g33149), enolase-phosphatase E1 (g36294), Nkx-2.1 (g37215), nucleosome-remodelling factor subunit BPTF (g37794), HES-4-A (g5821), nucleoside diphosphatase kinase A-like (g5223), high mobility group-T (g8440), protein arginine N-methyltransferase 1 (g8871), PR domain zinc finger (g9884).

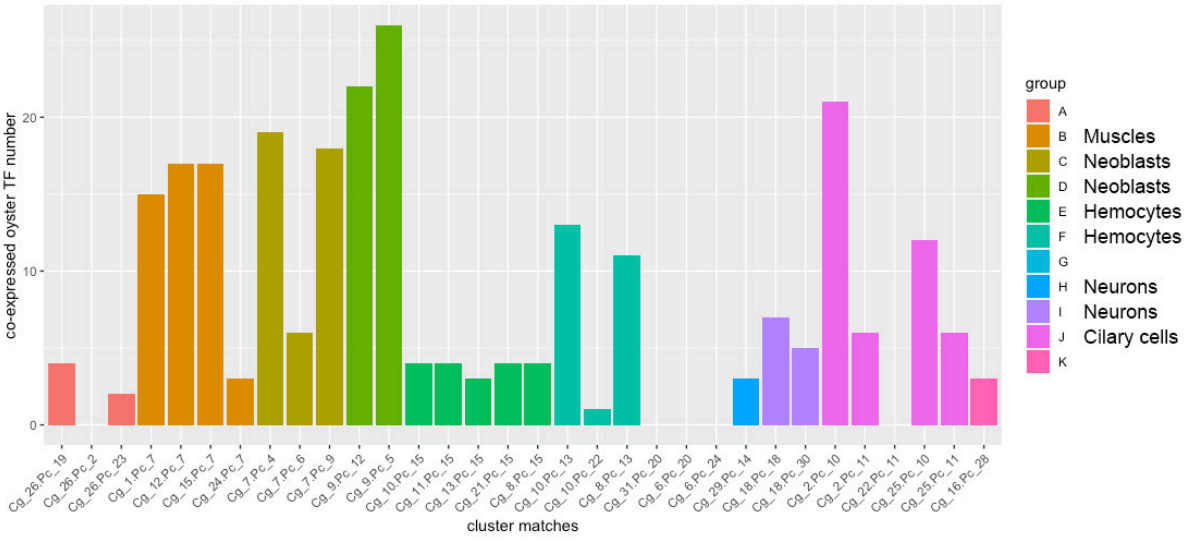
Across the “posterior” neuronal match (group H) I found only 3 TFs, however it is worth noting that cluster Cg 29 has only about 8 TFs as marker genes in total. Since the TFs in this match are only 3 for the oyster and 2 for the flatworm I blasted all 5 and got the following results:
zinc finger protein like (Cg-g17526 – Pc-g34279), carboxypeptidase E (Cg-g25100, Pc-g26139), NF-kappa-B inhibitor alpha like (Cg-g5327) (for this oyster gene the flatworm pairs are Pc-g30713 and Pc-g20329 but none of the two returned any significant annotated hit when blasted).

Across the remaining “apical” neuronal matches (group I) I found 11 co-expressed TFs and they are: Kv channel-interacting protein 4 (g1394), HLF (g1640), Sox2 (g17930), NF-kappa-B inhibitor epsilon like (g21793), NF-kappa-B inhibitor alpha like (Cg-g5327), Sox-11 (g25021), carboxypeptidase E (g25100), bromodomain adjacent to zinc finger domain (g26072), CCAAT/enhancer-binding protein gamma (g31102), CCAAT/enhancer-binding protein beta (g31103), enolase-phosphatase E1 (g36294).

Finally, across all the ciliary cells matches I found 23 co-expressed TFs and they were: proliferation-associated protein 2G4 (g12771), thyroid hormone receptor beta-A (g14856), HLF (g1640), Sox2 (g17930), Sox-11 (g25021), Krueppel-like factor 5 (g18155), zinc finger homeobox protein 4 (g20536), RREB1 (g21729), NF-kappa-B inhibitor epsilon like (g21793), NF-kappa-B inhibitor alpha like (Cg-g5327), SALL1 (g22184), bromodomain adjacent to zinc finger domain (g26072), phospholipid scramblase 1-like (g26301), msx2-interacting protein (g28832), CCAAT/enhancer-binding protein gamma (g31102, g31104), CCAAT/enhancer-

binding protein beta (g31103), EGR 1-B (g31219), MTA3 (g31878), enolase-phosphatase E1 (g36294), Fox-K2 (g37786), Fox-B1 (g9532) and ETS 4 (g875).

In general, except for the neuronal cell match group H which I have shown in the previous section to have a very specific molecular signature anyway, most cell type matches seem to co-express a pretty large number of transcription factor. This result is quite exciting as it is an indication that these matching cells not only express similar downstream protein (which in a sense we expect from cells that have the same function) but are also regulated similarly. Since most of these cell types are also likely to be present in adult Lophotrochozoa, it would be interesting to see whether adult cell types also share a similar TFs signature. Indeed, a comparison between larval and adult cell types could show whether the similarities we see (in both general and in TFs molecular signature) are due to the independent co-option of similar/homologous cell types in each of the two larvae (in which case the adult cell type signature might be even more similar) or due to a common larval ancestor.



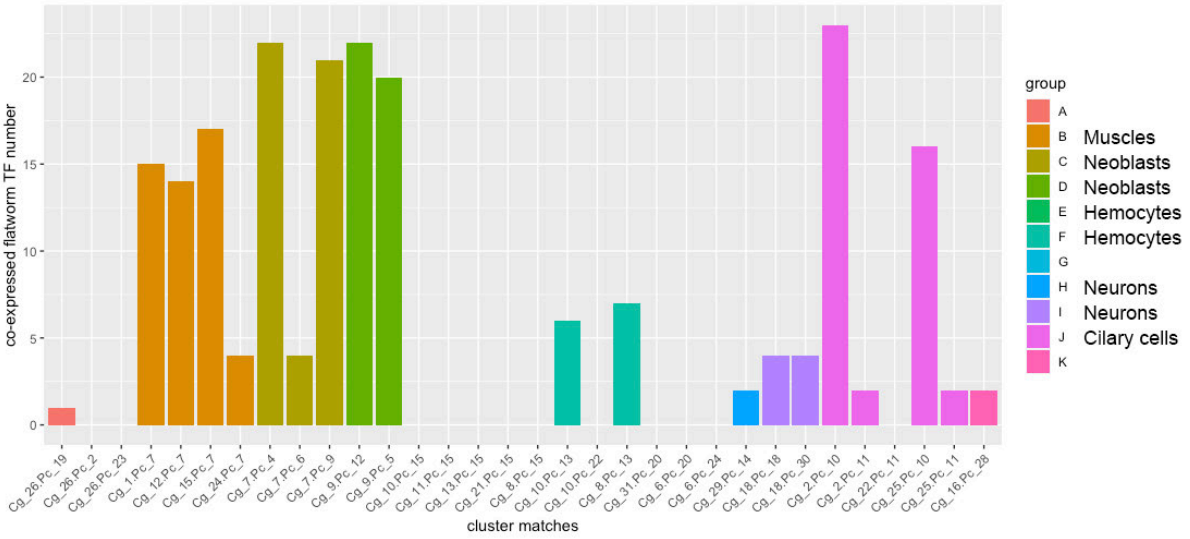


Figure 112. Number of co-express TFs correlates with number of co-expressed genes.

Bar plot graph shows that the number of TFs that are co-expressed between different matching clusters of the oyster (Cg) and flatworm (Pc) follow a similar distribution to that of all genes.

6.2.3 Lophotrochozoan specific genes

Finally, to try to understand whether these matching larval cell types might be homologous I decided to see how many of the co-expressed genes were lophotrochozan specific. This is because I would imagine that evolving a planktonic larva would require significant adaptation and possibly big changes in rates of evolution, gene duplication or appearance of new genes in a previously benthic animal. To see how many genes shared by larvae originated in the ancestor of these animals I decided to use a more sophisticated approach than the previous phylostratigraphy analysis done for the TAI (which showed some incongruences, see figure 106). I built a database of 30 highly complete genomes (for more information see table 3) containing 5 molluscs, 5 annelids, 5 platyhelminths, one brachiopod, one nemertean, one bryozoan, one phoronid, one rotifer, 5 ecdysozoans (4 arthropods and one priapulid) and 5 ambulacrarians (3 echinoderms and 2 hemichordates).

I used this dataset to run Orthofinder to identify orthologs and then picked all orthogroups that contained genes from at least two lophotrochozoan phyla but no gene from Ecdysozoa or Ambulacraria. This analysis was carried out with the help of Paschalis Natsidis from our lab who gathered most of the genomes, carried out Busco analysis for completeness and wrote a script to select the orthogroups. In total we found 2707 apparently lophotrochozoan specific orthogroups; of these only 142 contained both an oyster and a flatworm gene. From these 142 orthogroups I extracted 585 oyster genes and 304 flatworm genes.

In figure 113 I show the distribution of co-expressed lophotrochozoan specific genes in the different cell type matches. In general, we didn't find many co-expressed lophotrochozoan specific genes (<7 per match) however we did find a few in muscles, neurons and in ciliary bands clusters. Numbers here are small so it is hard to tell whether this is a real biological signal but it is exciting to see that the majority of the shared lophotrochozoan genes are indeed expressed in the ciliary bands of the larvae (see figure 114) which are the most common feature of all invertebrates' marine larvae.

Table 3. Database of 30 highly complete genomes used to find Lophotrochozoa specific genes.

We collected the most complete genomes we could find (approximately >80% Busco score) for 5 mollusc, 5 annelids, 5 platyhelminths, 5 other lophotrochozoans, 5 arthtropods and 5 ambulacrarians.

Species	**Superphylum**	**Phylum**	**Subphylum**	**Complete**	**Single-copy**
Capitella teleta	Lophotrochozoa	Annelida	Polychaeta	97.2	92.3
Hirudo medicinalis	Lophotrochozoa	Annelida	Clitellata	79.6	73.9
Lamellibrachia luymesi	Lophotrochozoa	Annelida	Polychaeta	93.2	90.8

Eisenia andrei	Lophotrochozoa	Annelida	Clitellata	89.9	82.8
Helobdella robusta	Lophotrochozoa	Annelida	Clitellata	93.3	90.3
Aplysia californica	Lophotrochozoa	Mollusca	Gastropoda	98.2	85.6
Crassostrea gigas	Lophotrochozoa	Mollusca	Bivalvia	84.6	81.1
Elysia chlorotica	Lophotrochozoa	Mollusca	Gastropoda	96.9	95.6
Octopus bimaculoides	Lophotrochozoa	Mollusca	Cephalopoda	95.6	74.1
Pecten maximus	Lophotrochozoa	Mollusca	Bivalvia	98.4	87.1
Echinococcus multilocularis	Lophotrochozoa	Platyhelminthes	Cestoda	82.9	79.7
Macrostomum lignano	Lophotrochozoa	Platyhelminthes	Rhabditophora	92.1	6.7
Prostheceraeus crozieri	Lophotrochozoa	Platyhelminthes	Rhabditophora	89.7	87.1
Schistosoma mansoni	Lophotrochozoa	Platyhelminthes	Rhabditophora	80.2	78.1
Hymenolepis microstoma	Lophotrochozoa	Platyhelminthes	Cestoda	81.1	74.1
Lingula anatina	Lophotrochozoa	Brachiopoda		97.8	75.4
Notospermus geniculatus	Lophotrochozoa	Nemertea		95	65.1
Bugula neritina	Lophotrochozoa	Bryozoa		79.5	69.8
Phoronis australis	Lophotrochozoa	Phoronida		97.3	94.3
Adineta vaga	Lophotrochozoa	Rotifera		90.6	17
Bombus terrestris	Ecdysozoa	Arthropoda	Lepidoptera	99.4	67.4
Daphnia pulex	Ecdysozoa	Arthropoda	Crustacea	97.9	70.3
Parasteatoda tepidarorium	Ecdysozoa	Arthropoda	Chelicerata	93.8	81.3
Trigoniulus corallinus	Ecdysozoa	Arthropoda	Myriapoda	96.1	84.6
Priapulus caudatus	Ecdysozoa	Priapulida		89.2	72.3
Asterias rubens	Ambulacraria	Echinodermata	Asteroidea	98.7	88.5
Anneissia japonica	Ambulacraria	Echinodermata	Crinoidea	97.3	84.4
Lytechinus variegatus	Ambulacraria	Echinodermata	Echinoidea	98.9	91.1
Saccoglossus kowalevskii	Ambulacraria	Hemichordata	Enteropneusta	95.5	92.3
Rhabdopleura recondita	Ambulacraria	Hemichordata	Pterobranchia	87.1	75.8

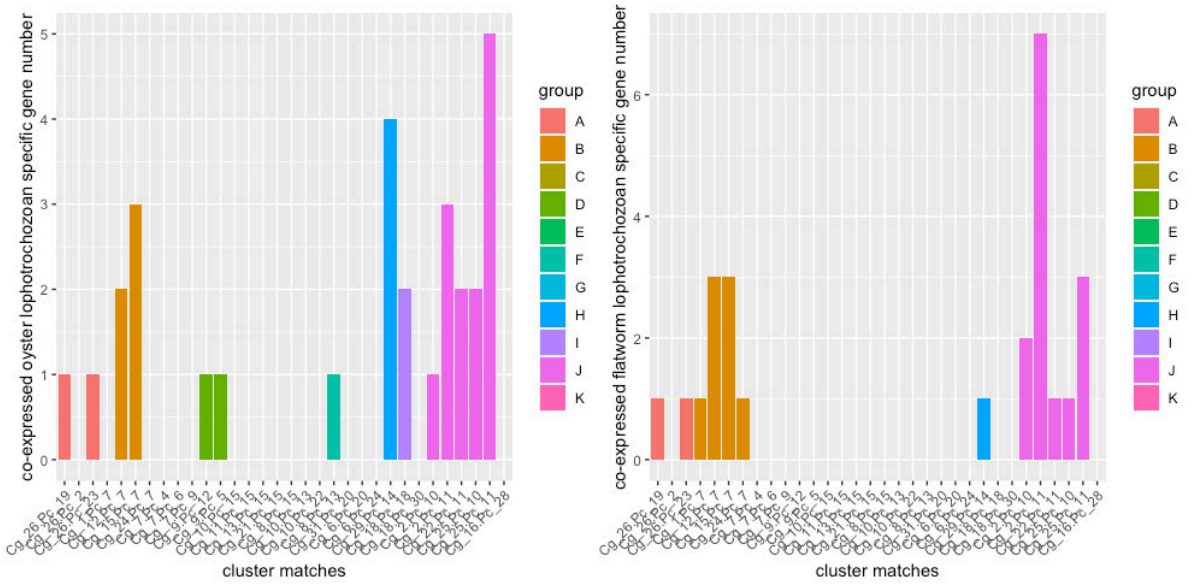


Figure 113. The Majority of co-expressed Lophotrochozoan genes are found in ciliary clusters.

Bar plot graphs showing number of lophotrochozoan specific gene (left oyster, right flatworm) pairs that are co-expressed between different matching clusters of the oyster (Cg) and flatworm (Pc). In general, we found very few co-expressed lophotrochozoan specific genes (<7).

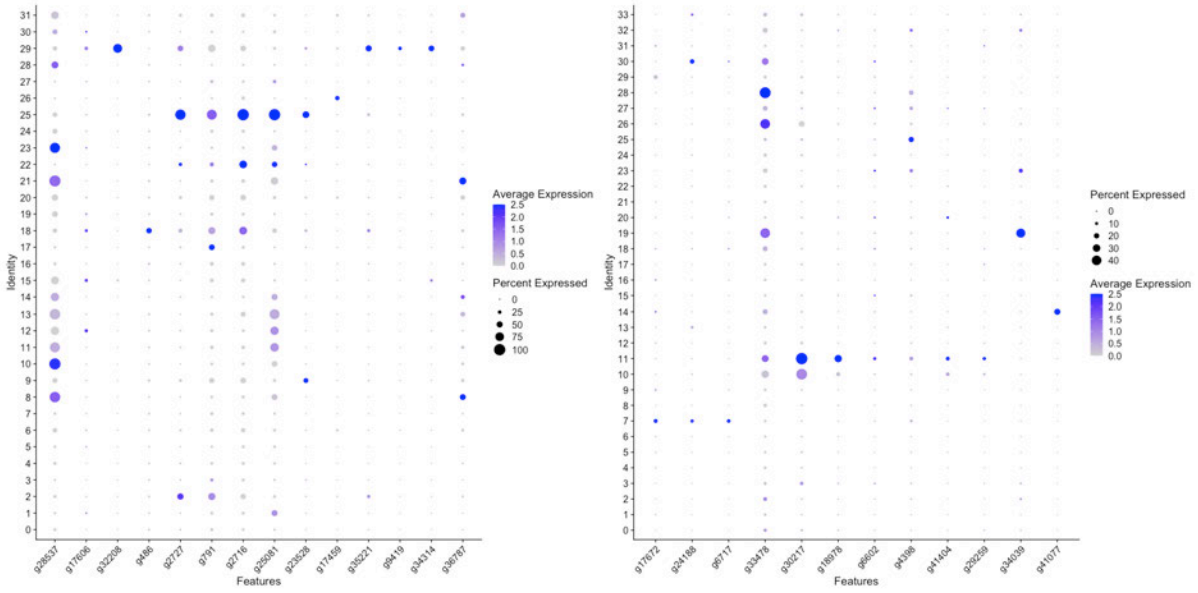


Figure 114. Lophotrochozoan specific co-expressed genes are specific to ciliary clusters.

Dotplot of the expression of the Lophotrochozoan specific co-expressed genes in the oyster (left) and flatworm (right) showing the majority of lophotrochozan genes are expressed in the ciliary bands (cluster Cg 2, 25, Cg 22 and Pc 10, Pc 11)

6.3 Conclusions

6.3.1 Summary of results

In this chapter, I attempted to perform an automated comparison of two Lophotrochozoan larvae using two different tools (Seurat integration and SAMAP) to try and assess their homology. I showed how the use of single copy orthologs of the Seurat integration is not very informative when comparing cell type across phyla as it greatly reduces complexity and causes loss of specific sub-cell types. For this reason, I decided to use a newly establish method called SAMAP (Tarashansky et al, 2021) which allowed me to utilise all orthologs for the comparison. Both methods highlighted similarities between muscles, neoblasts/proliferative cells, neurons and ciliary cells of the two larvae. Moreover, looking at the co-expressed genes produced by the second tool (SAMAP) I showed how these cells share a high percentage of co-expressed genes among which I found several transcription factors and even a few lophotrochozoan specific genes. Together with these general broad cell types similarities I also identified two matching population of neuronal cells with a very distinctive molecular signature (i.e. different from the other neuronal population in the larvae) present in both larvae. These cells could indeed represent a larval specific cell population and together with the high degree of matching molecular signature of the rest of the larva could indicate a common larval ancestor. In this paragraph I will critically discuss the strengths and limitations of the approach taken and draw some conclusions on the results obtained.

6.3.2 Computational limitations of scRNA seq comparisons across species

As mentioned above, in this study I presented only two different methods for cell type comparison across species. The first method I tried (the only method available when I began my PhD) was Seurat Integration, and clearly proved to be not very informative, as it uses a very a small subset of genes and drastically reduces the complexity of the original datasets. The second method I tried to apply is a newly developed tool called SAMAP (Tarashansky et al, 2021) which uses a gene to gene graph to align the datasets where multiple connections between genes are allowed, this way a considerably larger number of genes can be used for the comparison. Overall this method allowed me to identify several matches across similar cells and most of these appeared to make biological sense. Moreover, using this method I detected a considerable number of co-expressed genes for most of these matches (several of which were TFs) and these genes make up a large fraction of the cell original marker gene set. This means that the genes that are co-expressed across these distantly related cells are genes that contribute to the specific transcriptional signature of those cell types and not just a random subset of commonly expressed genes.

There are three main critical points I want to make about the results obtained with SAMAP:

1) The first is that SAMAP is obviously actively looking for co-expressed genes that are specific for each cell type, so in sense my observation is tautological: I look for co-expressed specific genes and then check that the genes I found are specific. What would be interesting to see is how specific co-expressed genes are in clearly homologous cells (i.e. what % of the original signature they make up). This would provide an idea of what similarity to expect, which is an especially hard bar to set in

our case where the two species have had 500 million years of divergence time. For instance, in this work we initially had planned to produce three cell type datasets, two for molluscs (one bivalve and one gastropod) and one for flatworms. Ideally this would have given us an idea of what similarity we could recover from clearly homologous larvae (the two mollusc ones) before embarking on a cross phyla comparison. Sadly, due to the Covid pandemic and subsequent closure of UCL for several months we could not produce these data, but this is surely something that would be extremely useful to investigate and that may be possible with the advent of more scRNA-seq dataets to explore.

2) The second point to make is that due to the large number of co-expressed genes recovered I could not confirm the orthology of each gene, so it is possible that the number of co-expressed genes is inflated by paralogous genes. Ultimately, we still do not have a perfect method to establish gene orthology (Natsidis et al, 2021), but hopefully more tools will be available to tackle the cross phyla comparison of cell types in the future allowing us to investigate how different methods to establish gene orthology (Orthofinder or OMA to name a few) affect the comparison and to possibly test them across evolutionary distances. In general, it is also possible that different cell type comparative methods could work better at different evolutionary distances; as we shown for Seurat Integration, which may only be useful when comparing closely related species.
3) Finally, SAMAP did identify some matches with very strong alignment scores that were not backed up by many genes, nor made much biological sense. This happened in three instances: A) for the match between the flatworm gut clusters and the oyster “neuronal” cluster 26; B) for the match between the flatworm protonephridia

and the oyster cluster 16; and C) for the match between the oyster shall gland and the flatworm Macif1+ cluster and cluster 24. When I looked at how many genes were co-expressed between these clusters I found very few (compared to the rest of the matches) and for this reason I ignored them. However, it is worth noticing that these matches still have between 20-70 co-expressed genes which is not necessarily such a small number: either these matches are caused by biological similarities or SAMAP can be skewed by a reduced number of very similar genes (especially when one of the clusters has poor coverage or very small cell number, as for cluster 26 of the oyster).

6.3.3 Differences in larval cell types

Another limitation of the comparison presented in this study is the extent of morphological differences found in the two larvae compared. More specifically, the oyster larva is extremely small (~50um) compared to the flatworm larva (~150um), is made up of around 200/300 cells compared to the ~1000/1500 cells of the flatworm larva and appears overall simpler. These features made it easier for us to get a very good coverage of all cell types and with a very high number of reads per cell and genes per cell for the oyster (see chapters 3,4 and 6) but at the same time made it harder to obtain a similar result on the flatworm. Moreover, as presented in detail in chapter 4, the small oyster larva at the trochophore stage doesn't have a gut or protonephridia and only has few neuronal cells. In comparison the nervous system of the flatworm larva is considerably more complex with several neuronal clusters that form a large apical organ and connect to the larval eyes (see chapter

5). That is not to say that the oyster larva is not complex in its own way, in fact, it presents several distinct muscles and shell gland clusters but it lacks some common features of trochophore larvae (i.e. a fully developed gut, protonephridia, an apical tuft and complex nervous system). For this reason, it would be extremely interesting, in the future, to extend our comparison to other lophotrochozoan larvae with more complex attributes similar to the flatworm larva.

Even though we found some differences in the cell types present in the two larvae, this doesn't necessarily conflict with the hypothesis of a common larval ancestor, in fact other mollusc trochophore larvae possess some of these characters. Moreover, even if we assume that larvae evolved once in the ancestor of Lophotrochozoa it is possible that each phylum retained or evolved different larval characters such as for example the shell gland of mollusc larvae. Indeed, as shown in chapter 4 I found that the shell gland of the oyster larva appears to express many novel or rapidly evolving genes which could indicate its recent origin. This result was particularly interesting because it also confirmed that the young transcriptomic age indices some authors recently found for larval stage of bivalves are likely skewed by the expression of cells of the shell gland and not, as suggested by some, evidence of the recent evolution of larval stages.

7 Conclusions

7.1 Summary of the main findings

The aim of my PhD was to try and establish whether the classical trochophore larva of mollusc was homologous to the (presumably) more derived Mueller's larva of a polyclad flatworm. Establishing the homology of these very distantly related larvae would help to understand whether larval stages are ancestral to Lophotrochozoa, providing an important piece to the puzzle of reconstructing the last common ancestor of animals. Our approach to this problem is novel and consists in identifying cell types *a priori* using single cell sequencing and then using the transcriptional profile of these cells to assess their similarities. To carry out this comparison I first gathered single cell sequencing data for both animals (see chapter 3 and 4 for the oyster and chapter 5 for the flatworm), used the transcriptional profile of each cell cluster to try and identify cells based on their function and validated their presence and location in both animals using *in situ* hybridisation. To date these are the first single cell sequencing atlases generated for larvae of either phylum and the first ever extended expression profile study on the polyclad flatworm Mueller's larvae for which we didn't have a working *in situ* hybridisation protocol until now. Regardless of the comparative aspect of this study, the data I generated on the flatworm larva will prove very useful in determining the origin of larvae in any future study.

After generating scRNA-seq data for both species and characterising most cell types I perform an automated comparison using two different tools (see chapter 6). First, I

attempted to use only single copy orthologs (which is a commonly used approach for cross species comparison) and showed that is performs relatively poorly across phyla; it greatly reduces complexity and causes loss of specific sub-cell types which could be informative for our comparison, such as larval specific cell types. For this reason, I tried a newly establish method (SAMAP, Tarashansky et al, 2021) which allows the user to keep all orthologs (1:1 and 1: many) for the comparison. In general, both methods highlighted similarities between broad cell families such as muscles, neoblasts/proliferative cells, neurons and ciliary cells of the two larvae. However, a more thorough analysis of the co-expressed genes generated by SAMAP highlighted how many of these cell types share a high number of co-expressed genes (up to 652), multiple transcription factors (up to 38) and even a few lophotrochozoan specific genes (up to 5) (for details see figure 115).

Of particular interest, I identified a population of “posterior” neuronal cells with a very distinctive molecular signature (i.e. different from the other neuronal populations in the larvae) that were present in both larvae. These cells could indeed represent a larval specific cell population and, together with the high degree of matching molecular signature of the rest of the larvae, could indicate a common larval ancestor. In this last chapter I will expand on the possible homology of lophotrochozoan larvae, outline how scRNA-seq could help us in finally answering this question and propose some future research ideas on this subject.

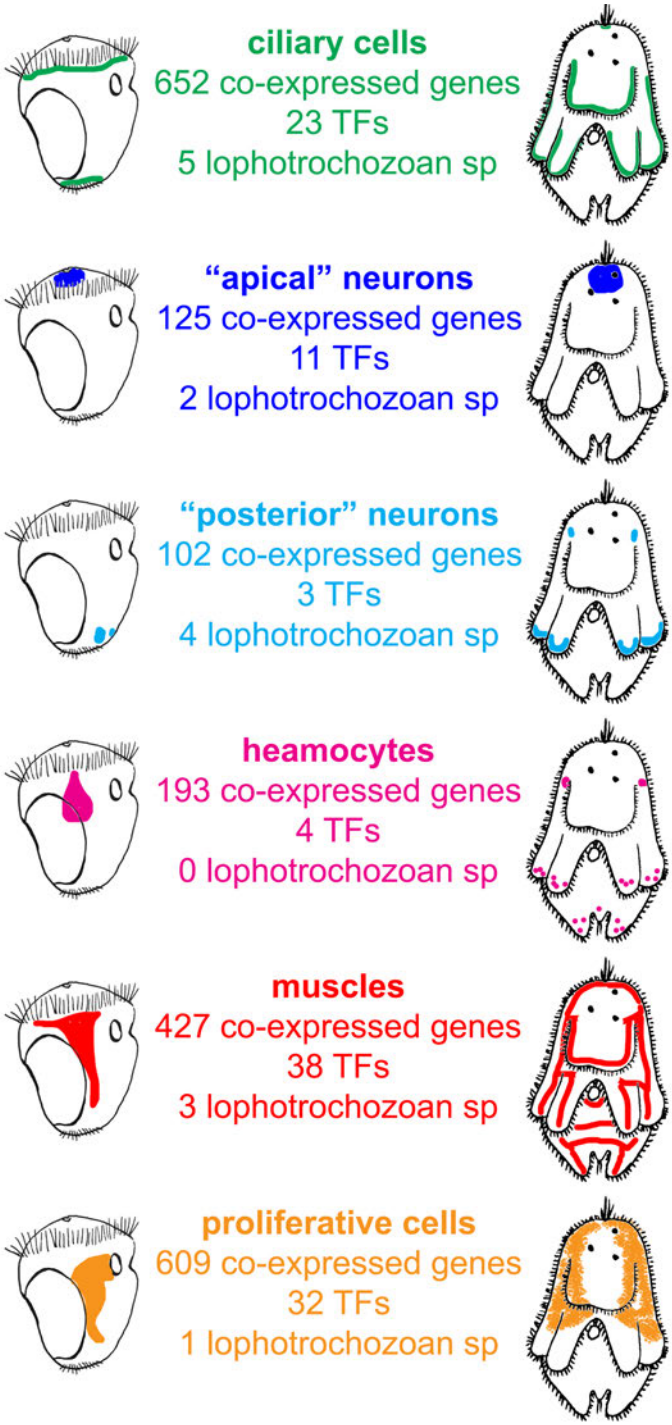


Figure 115. Cell types of oyster trochophore larva and flatworm Mueller's larva share many orthologous genes (up to 652), TFs (up to 38) and a few Lophotrochozoan specific genes.

On the left oyster larva schematic with approximate position of cells indicated in colour, on the right flatworm larva schematic with approximate position of cells indicated in colour. The number of co-expressed genes refer to the oyster genes (since 1: many matches are allowed by SAMAP the number of orthologs co-expressed is different in the two species). Lophotrochozoan sp indicates lophotrochozoan specific genes (for details on analysis see chapter 6).

7.2 Are lophotrochozoan larvae homologous?

Many authors have tried to tackle the question of the origin of marine larvae with an array of different approaches (see chapter 1). Initially, comparing the morphology and ultrastructure of cells that made up similar organs (such as ciliated bands, apical organ, protonephridia, larval eyes…). Then following the development of early blastomeres which showed a general conservation of cell fates across Lophotrochozoan larvae but also a certain plasticity in some species. Finally, more recently, looking at the similarity of molecular expression of candidate genes between different species. The general idea behind these methods was that the more complex the character shared across phyla (i.e. not only a similar structure but a similar structure that develops in the same way, or that expresses the same genes) the less likely it would be for it to have evolved in the exact same way twice. Importantly, these characters ought also to be larval specific to avoid the possibility that they have been incorporated convergently from homologous adults.

However, there are some major drawbacks to these approaches. First, it is hard to scale them up, for instance cell tracing is extremely time consuming and it can be hard to expand to many species. Similarly, performing *in situ* hybridisation requires an *a priori* knowledge of what genes to look for and it is again time consuming. Secondly, it seems that many authors have often given more importance to similarities rather than differences. The most notable example of this is the super imposition of spiralian nomenclature of blastomeres to non-spiral cleaving embryos including arthropods, which were long thought to be the sister group of the annelids. Finally, some authors have disagreed with the concept of a primitive

"trochophore" like larva suggesting that lecitotrophic larvae are ancestral to molluscs and annelids and that bi-phasic life cycle would have evolved secondarily in platyhelminths and nemerteans (Haszprunar et al, 1995).

At the same time, our knowledge of animal phylogeny has greatly improved and showed that the majority of ciliated marine larvae of animals are concentrated in the protostome super clade of Lophotrochozoa. So not only do these larvae look similar, often develop in the same way but we know now that most of them are closely related to each other. All this evidence hints at a single origin of Lophotrochozoan larvae (at least) and yet most authors would only agree that larvae of molluscs and annelids are homologous. Our work aimed to bring a new perspective to this century long debate by using the novel technique of scRNA-seq. This approach allowed us to unravel and compare the full transcriptional profile of cells of two lophotrochozoan larvae.

All the analyses described in chapter 6 seem to indicate homology of several cell types found in larvae of both phyla, such as muscles, neoblasts/proliferative cells, neurons and ciliary cells, as they appear to co-expresses hundreds of orthologous genes and several TFs. However, with the data currently available, we cannot unambiguously establish whether they originated in a common larval or adult ancestor and hence comment on the overall homology of the two larvae. In fact, all bilaterians possess muscles, proliferative cells, neurons, and ciliary cells and so it is possible that these were co-opted multiple times from the adult stage in both lineages and that this could cause the similarities we observe.

However, I have also identified a match between two subpopulation of neurons which appear to present a very distinct molecular signature which could be potentially larval specific. Moreover, it is also possible that some other cell types (such as ciliary cells, apical organ neurons or even muscles) could present a different transcriptional signature in the adult and in the larva. For example, authors have shown that blue mussel larvae possess a distinct larval shell proteome from the adult one (Carini et al, 2019). Likely this unique larval shell gland serves specific developmental needs, and in this sense, it wouldn't be surprising to find, for example, that ciliary cells, neuronal cells, gut cells or even muscle cells in the larva have a different molecular signature than in the adult.

A recent study found several TFs expressed in the excretory organs of several major protostome and non-vertebrate deuterostome lineages (Gasiorowski et al, 2021), some of which were also present in the flatworm protonephridia (see chapter 5). However, because of the small subset of genes analysed in the paper we do not yet know whether a more specific transcriptional signature is shared across protostomes, Lophotrochozoa and/or larval protonephridia nor whether they are expressed in adults (in the study they only show results for adult planarians). As a small example, in chapter 5 I performed a single cell comparison between flatworm larval and adult protonephridia by looking at the expression of *S. mediterranea* protonephridial markers in my flatworm larval data and found that only a few were co-expressed. This could be caused by the evolutionary distance between the two flatworms but could also indicate a difference in adult versus larval transcriptional profile of protonephridia. To test whether this is the case it would be important to carry out a comparison of cell types across species and across stages (i.e. adult cells vs larval cells). This would inform us on whether A) any matching cell types identified in this study are larval

specific and B) how similar larval or adult cell types are to ultimately help us to unravel the origin of lophotrochozoan larvae.

7.3 Can we use scRNA-seq to assess cell type homology?

Something that this study clearly highlighted is the sheer complexity of larval body plans and larval cell types and, even more, of their molecular signatures. In fact, throughout this work I found several cell sub-types never previously described (such as the multiple muscle clusters of the oyster larva) and a large number of genes not previously annotated in any species. This is likely largely due to the fact that Lophotrochozoa is the least studied of Bilaterian clades and adds to the usefulness of generating scRNA-seq data for these least studied phyla. Furthermore, this largely unexplored complexity also proves the importance of characterising cell types *a priori* using their full transcriptional signature rather than concentrating on a handful of well characterised genes.

For example, in the past, authors in favour of the homology of trochophore and Mueller's larvae have suggested that the ciliated lobes of the latter were derived prototrochs (based mostly on cell tracing) (Nielsen, 2005). In this study however, I found that A) cells of the prototroch and telotroch of the oyster and cells of the apical tuft and ciliary lobes of the flatworm were not specific to their structure of origin and B) a varying number of genes were supporting matches between the ciliary clusters depending on which species gene pair I looked at. Overall it appears that SAMAP picked up a general ciliary cell signature and not a very specific match between the prototroch and the ciliated lobes as we would have expected. This could be due to a number of reasons:

1) It could be that these different ciliated structures transiently express specific genes and transcription factors during their development but once fully differentiated they all revert to a general ciliary cell signature. In fact, the apical tuft of the flatworm doesn't appear to present a distinct signature. This could be tested in the future using earlier developmental stages scRNA seq data.
2) It could be that the absence of an apical tuft in the oyster larva confuses the analysis, and that if we chose a mollusc larva with an apical tuft then we would have had a more convincing match between those cell types.
3) Finally, it is possible that the ciliary bands are either not homologous or that we cannot establish their homology using scRNA seq data because their expression diverged so much in the past 500 million years that we all we pick up is a similarity in the general ciliary cell signature

Although I didn't find a specific one-to-one match between cells of the prototroch and those of the ciliary lobes but rather a general match between ciliary clusters, it is bewildering to think that more than 650 orthologous genes are co-expressed in these cells between the two larvae, of which approximately 20 are TFs and 5 are lophotrochozoan specific (see figure 115). The lophotrochozoan gene analysis expands on a recent paper, which indeed inspired it, that found two Lophotrochozoan genes co-expressed in ciliary bands of several spiralian larvae (Wu et al, 2020). Moreover, in the oyster larva I found a total of 14 out of 37 of the spiralian specific genes described in the Wu et al. paper in the ciliary clusters (see figure 32 in chapter 4). I think those results taken together strongly point to a possible common origin of ciliated larvae in the ancestor of Lophotrochozoa. In conclusion, the example of the ciliary cells highlights how some of our expectations on cell type diversity

may not hold true when looking at their full transcriptional signature. Yet it also proves the importance of using scRNAseq as an unbiased approach to explore cell type homology.

As mentioned before, systematically exploring cell type homology across different evolutionary distances will ultimately tell us whether any biological signature of such ancient events (such as the origin of Lophotrochozoa, Bilateria or Metazoa) can still be identified in transcriptional signatures of cells. Overall, I believe that using scRNA-seq datasets as a proxy for *a priori* cell type delimitation and comparison is a very powerful tool and it is highly preferable to the analysis of small gene subsets which could be biased by our current knowledge. I think many of the limitations of this study will be overcome by the development of new methods (both for data collection and data analysis) as well as generation of more datasets which will bring exciting new insights in the field of Evo-Devo.

7.4 Future directions

There are several limitations to the work presented in this thesis that I wished to have tackled if I had had more time. Firstly, as mentioned in chapter 5 and 6, a more in-depth sequencing of the flatworm datasets would be needed. This is because the smaller number of molecular markers and TFs identified in the flatworm data opposed to the oyster could be masking some further similarities/differences between the two larvae. This is certainly something that we plan to explore in the near future but couldn't be included in the thesis due to time restrictions. My hope would be that a more in-depth sequencing would return more cells and more markers per cell type which would allow for a more specific clustering. For instance, I would hope to find a more substantial molecular signature between the two

ciliary clusters Pc 10 and Pc 11 which at the moment show mostly overlapping genes, which could potentially allow us to identify the biological differences between these two ciliated cell types. Indeed, we may even be able to disentangle the ciliary cells of the apical tuft (which is a very interesting feature of marine larvae) from the rest of the ciliary cells. Moreover, it may be possible to further sub cluster the neuronal cells of Pc 18 to specifically identify sub-neuronal cells of the apical organ.

In fact, performing sub clustering of certain cell types, specifically neurons, could be useful for both species. However, whenever I attempted this I always found it hard to strike a balance between meaningful biological clusters and small cell to cell variations. It is certainly easier to perform such sub clustering when there is an *a priori* knowledge of different neurons type, as shown in a recent paper by Paganos and colleagues (2021). Certainly, with a more in-depth sequencing of the flatworm dataset I would like to attempt this again.

Apart from these small technical issues, the most exciting outlook of this project is to expand it to more marine larvae. The most obvious would be to try and compare the scRNAseq produced in this work with the echinoderm larva scRNAseq, which should become available soon (Paganos et al, 2021). It would be especially exciting to see whether we can find similar co-expression results for the ciliary cells and apical organ of these very distant marine larvae. If we did we could either conclude that what we are seeing is a common ciliary cell type/neuronal signature or that indeed all marine larvae are related. As mentioned previously, they only way to discern between the two would be to expand the comparison to adult or juveniles of the same species. This could highlight a difference in larval vs adult cells transcriptomic profile and potentially indicate whether larval cells retain

a higher similarity than adults' cells, which would strongly favour the hypothesis of the homology of marine larvae.

Furthermore, it would be very interesting to expand the comparison to other Lophotrochozoan larvae. In fact, as mentioned in the previous chapter, the oyster trochophore larva turned out to lack many classical features such as protonephridia, a developed gut and an apical tuft. For this reason, it wasn't possible to compare these structures to those of the flatworm larva. Since in many marine larvae these features are lost during development (for details see chapter 1), and could hence represent larval specific traits, it would be exciting to compare their transcriptional profile.

Ultimately, I believe that by expanding the single cell comparison to more marine larvae and to their adults/juveniles we would be able to identify whether: 1) there is a general co-expression of orthologous genes in larval structures 2) this can be explained by multiple co-option from adult stages or 3) it could represent evidence of larval homology.

REFERENCES

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215 (3): 403–10.

Arenas-Mena, Cesar, Kimberly Suk-Ying Wong, and Navid Arandi-Forosani. 2007. Ciliary Band Gene Expression Patterns in the Embryo and Trochophore Larva of an Indirectly Developing Polychaete. *Gene Expression Patterns* 7 (5): 544–49. https://doi.org/https://doi.org/10.1016/j.modgep.2007.01.007.

Arendsee, Zebulun, Jing Li, Urminder Singh, Arun Seetharam, Karin Dorman, and Eve Syrkin Wurtele. 2019. Phylostratr: A Framework for Phylostratigraphy. *Bioinformatics* 35 (19): 3617–27. https://doi.org/10.1093/bioinformatics/btz171.

Arendt, Detlev, Ulrich Technau, and Joachim Wittbrodt. 2001. Evolution of the Bilaterian Larval Foregut. *Nature* 409 (6816): 81–85. https://doi.org/10.1038/35051075.

Arendt, Detlev. 2008. The Evolution of Cell Types in Animals: Emerging Principles from Molecular Studies. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2416.

Barnett, Derek W, Erik K Garrison, Aaron R Quinlan, Michael P Strömberg, and Gabor T Marth. 2011. BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files. *Bioinformatics* 27 (12): 1691–92.

Bernt, Matthias, Alexander Donath, Frank Jühling, Fabian Externbrink, Catherine Florentz, Guido Fritzsch, Joern Pütz, Martin Middendorf, and Peter F Stadler. 2013. MITOS: Improved de Novo Metazoan Mitochondrial Genome Annotation. *Molecular Phylogenetics and Evolution* 69 (2): 313–19.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15): 2114–20.

Budd, Graham E, and Richard P Mann. 2020. Survival and Selection Biases in Early Animal Evolution and a Source of Systematic Overestimation in Molecular Clocks. *Interface Focus* 10 (4): 20190110.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. BLAST+: Architecture and Applications. *BMC Bioinformatics* 10 (1): 1–9.

Cameron, R Andrew, Kevin J Peterson, and Eric H Davidson. 1998. Developmental Gene Regulation and the Evolution of Large Animal Body Plans. *American Zoologist* 38 (4): 609–20.

Carini, Alessia, Tomas Koudelka, Andreas Tholey, Esther Appel, Stanislav N Gorb, Frank Melzner, and Kirti Ramesh. 2019. Proteomic Investigation of the Blue Mussel Larval

Shell Organic Matrix. *Journal of Structural Biology* 208 (3): 107385. https://doi.org/https://doi.org/10.1016/j.jsb.2019.09.002.

Carrasco-Rando, Marta, Antonio S Tutor, Silvia Prieto-Sánchez, Esther González-Pérez, Natalia Barrios, Annalisa Letizia, Paloma Martín, Sonsoles Campuzano, and Mar Ruiz-Gómez. 2011. *Drosophila* Araucan and Caupolican Integrate Intrinsic and Signalling Inputs for the Acquisition by Muscle Progenitors of the Lateral Transverse Fate. *PLoS Genetics* 7 (7): e1002186–e1002186. https://doi.org/10.1371/journal.pgen.1002186.

Chienwichai, Peerut, Sumate Ampawong, Poom Adisakwattana, Tipparat Thiangtrongjit, Yanin Limpanont, Phiraphol Chusongsang, Yupa Chusongsang, and Onrapak Reamtong. 2020. Effect of Praziquantel on *Schistosoma mekongi* Proteome and Phosphoproteome. *Pathogens (Basel, Switzerland)* 9 (6): 417. https://doi.org/10.3390/pathogens9060417.

Choi, Harry M T, Maayan Schwarzkopf, Mark E Fornace, Aneesh Acharya, Georgios Artavanis, Johannes Stegmaier, Alexandre Cunha, and Niles A Pierce. 2018. Third-Generation *in situ* Hybridization Chain Reaction: Multiplexed, Quantitative, Sensitive, Versatile, Robust. *Development* 145 (12).

Croll, R P, and Amanda J G Dickinson. 2004. Form and Function of the Larval Nervous System in Molluscs. *Invertebrate Reproduction & Development* 46 (2–3): 173–87.

Croll, R P. 2009. Developing Nervous Systems in Molluscs: Navigating the Twists and Turns of a Complex Life Cycle. *Brain, Behavior and Evolution* 74 (3): 164–76. https://doi.org/10.1159/000258664.

Dassow, George von, and Svetlana A Maslakova. 2017. The Trochoblasts in the Pilidium Larva Break an Ancient Spiralian Constraint to Enable Continuous Larval Growth and Maximally Indirect Development. *EvoDevo* 8 (1): 19. https://doi.org/10.1186/s13227-017-0079-5.

Davidson, Eric H, Kevin J Peterson, and R Andrew Cameron. 1995. Origin of Bilaterian Body Plans: Evolution of Developmental Regulatory Mechanisms. *Science* 270 (5240): 1319–25.

Davidson, Eric H, R Andrew Cameron, and Andrew Ransick. 1998. Specification of Cell Fate in the Sea Urchin Embryo: Summary and Some Proposed Mechanisms. *Development* 125 (17): 3269–90.

Denes, Alexandru S, Gáspár Jékely, Patrick R H Steinmetz, Florian Raible, Heidi Snyman, Benjamin Prud'homme, David E K Ferrier, Guillaume Balavoine, and Detlev Arendt. 2007. Molecular Architecture of Annelid Nerve Cord Supports Common Origin of Nervous System Centralization in Bilateria. *Cell* 129 (2): 277–88. https://doi.org/https://doi.org/10.1016/j.cell.2007.02.040.

Dobin, Alexander, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. 2013. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29 (1): 15–21.

Drost, Hajk-Georg, Alexander Gabel, Jialin Liu, Marcel Quint, and Ivo Grosse. 2018. MyTAI: Evolutionary Transcriptomics with R. *Bioinformatics* 34 (9): 1589–90. https://doi.org/10.1093/bioinformatics/btx835.

Dunn, Ewan F, Vanessa N Moy, Lynne M Angerer, Robert C Angerer, Robert L Morris, and Kevin J Peterson. 2007. Molecular Paleoecology: Using Gene Regulatory Analysis to Address the Origins of Complex Life Cycles in the Late Precambrian. *Evolution & Development* 9 (1): 10–24.

Dyachuk, V, A Wanninger, and Elena Voronezhskaya. 2012. Innervation of Bivalve Larval Catch Muscles by Serotonergic and FMRFamidergic Neurons. *Acta Biologica Hungarica* 63 (Supplement-2): 221–29.

Dyachuk, Vyacheslav A. 2016. Hematopoiesis in Bivalvia Larvae: Cellular Origin, Differentiation of Hemocytes, and Neoplasia. *Developmental & Comparative Immunology* 65: 253–57.

Eisenhoffer, George T, Hara Kang, and Alejandro Sánchez Alvarado. 2008. Molecular Analysis of Stem Cells and Their Descendants during Cell Turnover and Regeneration in the Planarian *Schmidtea mediterranea*. *Cell Stem Cell* 3 (3): 327–39.

Emms, David M, and Steven Kelly. 2019. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biology* 20 (1): 1–14.

Farias, Leonardo P, Fernanda C Cardoso, Patricia A Miyasato, Bogar O Montoya, Cibele A Tararam, Henrique K Roffato, Toshie Kawano, et al. 2010. *Schistosoma mansoni* Stomatin like Protein-2 Is Located in the Tegument and Induces Partial Protection against Challenge Infection. *PLoS Neglected Tropical Diseases* 4 (2): e597–e597. https://doi.org/10.1371/journal.pntd.0000597.

Figueiredo, Barbara Castro Pimentel, Natan Raimundo Gonçalves De Assis, Suellen Batistoni De Morais, Vicente Paulo Martins, Natasha Delaqua Ricci, Rodrigo Marques Bicalho, Carina Da Silva Pinheiro, and Sergio Costa Oliveira. 2014. Immunological Characterization of a Chimeric Form of *Schistosoma mansoni* Aquaporin in the Murine Model. *Parasitology* 141 (10): 1277–88. https://doi.org/10.1017/S0031182014000468.

Forsthoefel, David J, Nicholas I Cejda, Umair W Khan, and Phillip A Newmark. 2020. Cell-Type Diversity and Regionalized Gene Expression in the Planarian Intestine. *Elife* 9: e52613.

Gąsiorowski, Ludwik, Carmen Andrikou, Ralf Janssen, Paul Bump, Graham E Budd, Christopher J Lowe, and Andreas Hejnol. 2021. Molecular Evidence for a Single Origin

of Ultrafiltration-Based Excretory Organs. *Current Biology* 31 (16): 3629-3638.e2. https://doi.org/https://doi.org/10.1016/j.cub.2021.05.057.

Gawad, Charles, Winston Koh, and Stephen R Quake. 2016. Single-Cell Genome Sequencing: Current State of the Science. *Nature Reviews Genetics* 17 (3): 175–88. https://doi.org/10.1038/nrg.2015.16.

Giribet, Gonzalo. 2008. Assembling the Lophotrochozoan (= Spiralian) Tree of Life. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1496): 1513–22.

Grande, Cristina, and Nipam H. Patel. 2009. Nodal Signalling Is Involved in Left-Right Asymmetry in Snails. *Nature* 457 (7232): 1007–11. https://doi.org/10.1038/nature07603.

Haeckel, Ernst. 1874. Memoirs: The Gastraea-Theory, the Phylogenetic Classification of the Animal Kingdom and the Homology of the Germ-Lamellæ. *Journal of Cell Science* 2 (54): 142–65.

Haszprunar, Gerhard, Luitfried v Salvini-Plawen, and Reinhard M Rieger. 1995. Larval Planktotrophy—a Primitive Trait in the Bilateria? *Acta Zoologica* 76 (2): 141–54.

Hatschek, Berthold. 1878. *Studien Über Die Entwicklungsgeschichte Der Anneliden: Ein Beitrag Zur Morphologie Der Bilaterien*. A. Hölder.

He, Yan, Aude Jouaux, Susan E Ford, Christophe Lelong, Pascal Sourdaine, Michel Mathieu, and Ximing Guo. 2015. Transcriptome Analysis Reveals Strong and Complex Antiviral Response in a Mollusc. *Fish & Shellfish Immunology* 46 (1): 131–44.

Hejnol, Andreas, Mark Q Martindale, and Jonathan Q Henry. 2007. High-Resolution Fate Map of the Snail *Crepidula fornicata*: The Origins of Ciliary Bands, Nervous System, and Muscular Elements. *Developmental Biology* 305 (1): 63–76.

Hejnol, Andreas. 2010. A Twist in Time—The Evolution of Spiral Cleavage in the Light of Animal Phylogeny. *Integrative and Comparative Biology* 50 (5): 695–706. https://doi.org/10.1093/icb/icq103.

Henry, Jonathan J, and Mark Q Martindale. 1999. Conservation and Innovation in Spiralian Development. *Hydrobiologia* 402: 255–65.

Henry, Jonathan Q, Andreas Hejnol, Kimberly J Perry, and Mark Q Martindale. 2007. Homology of Ciliary Bands in Spiralian Trochophores. *Integrative and Comparative Biology* 47 (6): 865–71.

Hoff, Katharina J, Simone Lange, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32 (5): 767–69.

Hooper, Scott L, and Jeffrey B Thuma. 2005. Invertebrate Muscles: Muscle Specific Genes and Proteins. *Physiological Reviews* 85 (3): 1001–60.

Huan, Pin, Gang Liu, Hongxia Wang, and Baozhong Liu. 2013. Identification of a Tyrosinase Gene Potentially Involved in Early Larval Shell Biogenesis of the Pacific Oyster *Crassostrea gigas*. *Development Genes and Evolution* 223 (6): 389–94.

Jagersten, Gosta. 1972. Evolution of the Metazoan Life Cycle. *Acad. Press.*

Jenner, Ronald A. 2000. Evolution of Animal Body Plans: The Role of Metazoan Phylogeny at the Interface between Pattern and Process. *Evolution & Development* 2 (4): 208–21.

Kakoi, Shota, Koryu Kin, Katsumi Miyazaki, and Hiroshi Wada. 2008. Early Development of the Japanese Spiny Oyster (*Saccostrea kegaki*): Characterization of Some Genetic Markers. *Zoological Science* 25 (5): 455–64. https://doi.org/10.2108/zsj.25.455.

Kerner, Pierre, Aissam Ikmi, Dario Coen, and Michel Vervoort. 2009. Evolutionary History of the Iroquois/Irx Genes in Metazoans. *BMC Evolutionary Biology* 9 (1): 1–14.

Kimball, Casey, Kaleigh Powers, John Dustin, Vanessa Poirier, and Jason Pellettieri. 2020. The Exon Junction Complex Is Required for Stem and Progenitor Cell Maintenance in Planarians. *Developmental Biology* 457 (1): 119–27.

Klein, Allon M, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner. 2015. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* 161 (5): 1187–1201. https://doi.org/10.1016/j.cell.2015.04.044.

Kocot, Kevin M. 2016. On 20 Years of Lophotrochozoa. *Organisms Diversity & Evolution* 16 (2): 329–43.

Kong, Yawei, Gu Jing, Zhenguang Yan, Changzhong Li, Ningping Gong, Fangjie Zhu, Dongxian Li, Yaorun Zhang, Guilan Zheng, and Hongzhong Wang. 2009. Cloning and Characterization of Prisilkin-39, a Novel Matrix Protein Serving a Dual Role in the Prismatic Layer Formation from the Oyster *Pinctada fucata*. *Journal of Biological Chemistry* 284 (16): 10841–54.

Kurita, Yoshihisa, Naoki Hashimoto, and Hiroshi Wada. 2016. Evolution of the Molluscan Body Plan: The Case of the Anterior Adductor Muscle of Bivalves. *Biological Journal of the Linnean Society* 119 (2): 420–29. https://doi.org/10.1111/bij.12812.

Lengerer, Birgit, Robert Pjeta, Julia Wunderer, Marcelo Rodrigues, Roberto Arbore, Lukas Schärer, Eugene Berezikov, et al. 2014. Biological Adhesion of the Flatworm *Macrostomum lignano* Relies on a Duo-Gland System and Is Mediated by a Cell Type-Specific Intermediate Filament Protein. *Frontiers in Zoology* 11 (1): 12. https://doi.org/10.1186/1742-9994-11-12.

Lengerer, Birgit, Julia Wunderer, Robert Pjeta, Giada Carta, Damian Kao, Aziz Aboobaker, Christian Beisel, Eugene Berezikov, Willi Salvenmoser, and Peter Ladurner. 2018. Organ Specific Gene Expression in the Regenerating Tail of *Macrostomum lignano*. *Developmental Biology* 433 (2): 448–60. https://doi.org/https://doi.org/10.1016/j.ydbio.2017.07.021.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16): 2078–79.

Liu, Gang, Pin Huan, and Baozhong Liu. 2020. Identification of Three Cell Populations from the Shell Gland of a Bivalve Mollusc. *Development Genes and Evolution* 230 (1): 39–45. https://doi.org/10.1007/s00427-020-00646-9.

Lomsadze, Alexandre, Paul D Burns, and Mark Borodovsky. 2014. Integration of Mapped RNA-Seq Reads into Automatic Training of Eukaryotic Gene Finding Algorithm. *Nucleic Acids Research* 42 (15): e119–e119.

Love, Alan C, Abigail E Lee, Mary E Andrews, and Rudolf A Raff. 2008. Co-option and Dissociation in Larval Origins and Evolution: The Sea Urchin Larval Gut. *Evolution & Development* 10 (1): 74–88.

Lv, Z., Qiu, L., Liu, Z., Wang, W., Chen, H., Jia, Y., Jia, Z., Jiang, S., Wang, L. and Song, L., 2018. Molecular characterization of a cathepsin L1 highly expressed in phagocytes of pacific oyster Crassostrea gigas. Developmental & Comparative Immunology, 89, pp.152-162.

Lyons, Deirdre C., and Jonathan Q. Henry. 2014. Ins and Outs of Spiralian Gastrulation. *International Journal of Developmental Biology* 58 (6–8): 413–28. https://doi.org/10.1387/ijdb.140151dl.

Malagoli, Davide, Mauro Mandrioli, Fabio Tascedda, and Enzo Ottaviani. 2017. Circulating Phagocytes: The Ancient and Conserved Interface between Immune and Neuroendocrine Function. *Biological Reviews* 92 (1): 369–77.

Marlétaz, Ferdinand, Katja T C A Peijnenburg, Taichiro Goto, Noriyuki Satoh, and Daniel S Rokhsar. 2019. A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Current Biology* 29 (2): 312–18.

Marlow, Heather, Maria Antonietta Tosches, Raju Tomer, Patrick R Steinmetz, Antonella Lauri, Tomas Larsson, and Detlev Arendt. 2014. Larval Body Patterning and Apical Organs Are Conserved in Animal Evolution. *BMC Biology* 12 (1): 7. https://doi.org/10.1186/1741-7007-12-7.

Martín-Durán, José María, and Bernhard Egger. 2012. Developmental Diversity in Free-Living Flatworms. *EvoDevo* 3 (1): 7. https://doi.org/10.1186/2041-9139-3-7.

Maslakova, S A, M Q Martindale, and Jon L Norenburg. 2004. Vestigial Prototroch in a Basal Nemertean, *Carinoma tremaphoros* (Nemertea; Palaeonemertea). *Evolution & Development* 6 (4): 219–26.

Maslakova, Svetlana A. 2010. The Invention of the Pilidium Larva in an Otherwise Perfectly Good Spiralian Phylum Nemertea. *Integrative and Comparative Biology* 50 (5): 734–43. https://doi.org/10.1093/icb/icq096.

Meyer, Néva P, and Elaine C Seaver. 2009. Neurogenesis in an Annelid: Characterization of Brain Neural Precursors in the Polychaete *Capitella sp. I*. *Developmental Biology* 335 (1): 237–52.

Miyamoto, Hiroshi, Fumiko Miyoshi, and Jun Kohno. 2005. The Carbonic Anhydrase Domain Protein Nacrein Is Expressed in the Epithelial Cells of the Mantle and Acts as a Negative Regulator in Calcification in the Mollusc *Pinctada fucata*. *Zoological Science* 22 (3): 311–15.

Natsidis, Paschalis, Paschalia Kapli, Philipp H Schiffer, and Maximilian J Telford. 2021. Systematic Errors in Orthology Inference and Their Effects on Evolutionary Analyses. *Iscience* 24 (2): 102110.

Nielsen, C. 1995. Animal Evolution. Interrelationships of the Living Phyla. Oxford University Press, Oxford.

Nielsen, Claus. 2002. The Phylogenetic Position of Entoprocta, Ectoprocta, Phoronida, and Brachiopoda. *Integrative and Comparative Biology* 42 (3): 685–91.

Nielsen, Claus. 2004. Trochophora Larvae: Cell-lineages, Ciliary Bands, and Body Regions. 1. Annelida and Mollusca. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 302 (1): 35–68.

Nielsen, Claus. 2005. Trochophora Larvae: Cell-lineages, Ciliary Bands and Body Regions. 2. Other Groups and General Discussion. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304 (5): 401–47.

Nielsen, Claus. 2012. *Animal Evolution: Interrelationships of the Living Phyla*. Oxford University Press on Demand.

Nuria, R, Antonio Villalba, and Asunción Cao. 2020. Differences in Proteomic Profile between Two Haemocyte Types, Granulocytes and Hyalinocytes, of the Flat Oyster *Ostrea edulis*. *Fish & Shellfish Immunology* 100: 456–66.

Orii, Hidefumi, Takashige Sakurai, and Kenji Watanabe. 2005. Distribution of the Stem Cells (Neoblasts) in the Planarian *Dugesia japonica*. *Development Genes and Evolution* 215 (3): 143–57.

Paganos, Periklis, Danila Voronov, Jacob Musser, Detlev Arendt, and Maria Ina Arnone. 2021. Single Cell RNA Sequencing of the *Strongylocentrotus purpuratus* Larva Reveals the Blueprint of Major Cell Types and Nervous System of a Non-Chordate Deuterostome. *BioRxiv*.

Page, Louise R. 2009. Molluscan Larvae: Pelagic Juveniles or Slowly Metamorphosing Larvae? *The Biological Bulletin* 216 (3): 216–25.

Peterson, Kevin J, R Andrew Cameron, and Eric H Davidson. 1997. Set-aside Cells in Maximal Indirect Development: Evolutionary and Developmental Significance. *BioEssays* 19 (7): 623–31.

Peterson, Kevin J., and Douglas J. Eernisse. 2001. Animal Phylogeny and the Ancestry of Bilaterians: Inferences from Morphology and 18S RDNA Gene Sequences. *Evolution and Development* 3 (3): 170–205. https://doi.org/10.1046/j.1525-142X.2001.003003170.x.

Peterson, Kevin J, Mark A McPeek, and David A D Evans. 2005. Tempo and Mode of Early Animal Evolution: Inferences from Rocks, Hox, and Molecular Clocks. *Paleobiology* 31 (S2): 36–55.

Philippe, Hervé, Henner Brinkmann, Richard R Copley, Leonid L Moroz, Hiroaki Nakano, Albert J Poustka, Andreas Wallberg, Kevin J Peterson, and Maximilian J Telford. 2011. Acoelomorph Flatworms Are Deuterostomes Related to *Xenoturbella*. *Nature* 470 (7333): 255–58. https://doi.org/10.1038/nature09676.

Philippe, Hervé, Albert J Poustka, Marta Chiodin, Katharina J Hoff, Christophe Dessimoz, Bartlomiej Tomiczek, Philipp H Schiffer, et al. 2019. Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria. *Current Biology* 29 (11): 1818-1826.e6. https://doi.org/10.1016/j.cub.2019.04.009.

Plass, Mireya, Jordi Solana, F Alexander Wolf, Salah Ayoub, Aristotelis Misios, Petar Glažar, Benedikt Obermayer, Fabian J Theis, Christine Kocks, and Nikolaus Rajewsky. 2018. Cell Type Atlas and Lineage Tree of a Whole Complex Animal by Single-Cell Transcriptomics. *Science* 360 (6391).

Raff, Rudolf A. 2008. Origins of the Other Metazoan Body Plans: The Evolution of Larval Forms. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1496): 1473–79.

Rawlinson, Kate A. 2010. Embryonic and Post-Embryonic Development of the Polyclad Flatworm *Maritigrella crozieri*; Implications for the Evolution of Spiralian Life History Traits. *Frontiers in Zoology* 7 (1): 12. https://doi.org/10.1186/1742-9994-7-12.

Rawlinson, Kate A. 2014. The Diversity, Development and Evolution of Polyclad Flatworm Larvae. *EvoDevo* 5 (1): 9. https://doi.org/10.1186/2041-9139-5-9.

Rawlinson, Kate A, Francois Lapraz, Edward R Ballister, Mark Terasaki, Jessica Rodgers, Richard J McDowell, Johannes Girstmair, Katharine E Criswell, Miklos Boldogkoi, and Fraser Simpson. 2019. Extraocular, Rod-like Photoreceptors in a Flatworm Express Xenopsin Photopigment. *ELife* 8: e45465.

Rieger, R.M., 1994. The biphasic life cycle—a central theme of metazoan evolution. *American Zoologist*, *34*(4), pp.484-491.

Rouse, G.W., 1999. Trochophore concepts: ciliary bands and the evolution of larvae in spiralian Metazoa. *Biological Journal of the Linnean Society*, 66(4), pp.411-464.

Ruppert, Edward E. 1978. A Review of Metamorphosis of Turbellarian Larvae. *Settlement and Metamorphosis of Marine Invertebrate Larvae* 65: 82.

Salvini-Plawen, Luitfried von, and Ernst Mayr. 1977. *On the Evolution of Photorecptors and Eyes*. Plenum Press.

Santagata, Scott. 2015A. Phoronida. In *Evolutionary Developmental Biology of Invertebrates 2*, 231–45. Springer.

Santagata, Scott. 2015B. Ectoprocta. In *Evolutionary Developmental Biology of Invertebrates 2,* 247–62. Springer.

Scimone, M Lucila, Mansi Srivastava, George W Bell, and Peter W Reddien. 2011. A Regulatory Program for Excretory System Regeneration in Planarians. *Development* 138 (20): 4387–98.

Shapiro, Ehud, Tamir Biezuner, and Sten Linnarsson. 2013. Single-Cell Sequencing-Based Technologies Will Revolutionize Whole-Organism Science. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg3542.

Sivka, Urška, Nataša Toplak, Simon Koren, and Jernej Jakše. 2018. De Novo Transcriptome of the Pallial Gland of the Date Mussel (*Lithophaga lithophaga*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 26: 1–9. https://doi.org/https://doi.org/10.1016/j.cbd.2018.02.001.

Sly, Belinda J, Margaret S Snoke, and Rudolf A Raff. 2003. Who Came First--Larvae or Adults? Origins of Bilaterian Metazoan Larvae. *International Journal of Developmental Biology* 47 (7–8): 623–32.

Song, Xiaorui, Hao Wang, Lusheng Xin, Jiachao Xu, Zhihao Jia, Lingling Wang, and Linsheng Song. 2016. The Immunological Capacity in the Larvae of Pacific Oyster *Crassostrea gigas*. *Fish & Shellfish Immunology* 49: 461–69.

Stanke, Mario, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. 2006. AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Research* 34 (suppl_2): W435–39.

Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. Using Native and Syntenically Mapped CDNA Alignments to Improve de Novo Gene Finding. *Bioinformatics* 24 (5): 637–44.

Strathmann, R.R., Jahn, T.L. and Fonseca, J.R., 1972. Suspension feeding by marine invertebrate larvae: clearance of particles by ciliated bands of a rotifer, pluteus, and trochophore. The Biological Bulletin, 142(3), pp.505-519.

Taguchi, Shunsuke, Kunifumi Tagawa, Tom Humphreys, and Nori Satoh. 2002. Group B Sox Genes That Contribute to Specification of the Vertebrate Brain Are Expressed in the Apical Organ and Ciliary Bands of Hemichordate Larvae. *Zoological Science* 19 (1): 57–66. https://doi.org/10.2108/zsj.19.57.

Tanay, Amos, and Arnau Sebé-Pedrós. 2021. Evolutionary Cell Type Mapping with Single-Cell Genomics. *Trends in Genetics*. https://doi.org/https://doi.org/10.1016/j.tig.2021.04.008.

Tarashansky, Alexander J, Jacob M Musser, Margarita Khariton, Pengyang Li, Detlev Arendt, Stephen R Quake, and Bo Wang. 2021. Mapping Single-Cell Atlases throughout Metazoa Unravels Cell Type Evolution. *Elife* 10: e66747.

Teshirogi, W, S Ishida, and K Jatani. 1981. On the Early Development of Some Japanese Polyclads. *Rep Fukara Mar Biol Lab* 9: 2–31.

Thiruvalluvan, Manish, Paul G Barghouth, Assaf Tsur, Limor Broday, and Néstor J Oviedo. 2018. SUMOylation Controls Stem Cell Proliferation and Regional Cell Death through Hedgehog Signaling in Planarians. *Cellular and Molecular Life Sciences* 75 (7): 1285–1301.

Tirapé, Ana, Coralie Bacque, Raphaël Brizard, Franck Vandenbulcke, and Viviane Boulo. 2007. Expression of Immune-Related Genes in the Oyster *Crassostrea gigas* during Ontogenesis. *Developmental & Comparative Immunology* 31 (9): 859–73.

Wang, Jing, Lingling Zhang, Shanshan Lian, Zhenkui Qin, Xuan Zhu, Xiaoting Dai, Zekun Huang, Caihuan Ke, Zunchun Zhou, and Jiankai Wei. 2020. Evolutionary Transcriptomics of Metazoan Biphasic Life Cycle Supports a Single Intercalation Origin of Metazoan Larvae. *Nature Ecology & Evolution*, 1–12.

Wang, X, and M Yu. 2008. Embryonic Development of the Polyclad *Notoplana humilis*. *CHINESE JOURNAL OF ZOOLOGY-PEKING*- 43 (2): 96.

Wanninger, Andreas. 2015. Entoprocta. In *Evolutionary Developmental Biology of Invertebrates 2*, 89–101. Springer.

Weiss, Ingrid Maria, Noreen Tuross, L I A Addadi, and Steve Weiner. 2002. Mollusc Larval Shell Formation: Amorphous Calcium Carbonate Is a Precursor Phase for Aragonite. *Journal of Experimental Zoology* 293 (5): 478–91.

Wilson, Edmund Beecher. 1892. *The Cell-Lineage of Nereis*. Ginn & Company.

Wollesen, Tim, Carmel McDougall, and Detlev Arendt. 2019. Remnants of Ancestral Larval Eyes in an Eyeless Mollusk? Molecular Characterization of Photoreceptors in the Scaphopod Antalis Entalis. *EvoDevo* 10 (1): 25. https://doi.org/10.1186/s13227-019-0140-7.

Wu, Longjun, Kailey E Ferger, and J David Lambert. 2019. Gene Expression Does Not Support the Developmental Hourglass Model in Three Animals with Spiralian Development. *Molecular Biology and Evolution* 36 (7): 1373–83.

Wu, Longjun, Laurel S Hiebert, Marleen Klann, Yale Passamaneck, Benjamin R Bastin, Stephan Q Schneider, Mark Q Martindale, Elaine C Seaver, Svetlana A Maslakova, and J David Lambert. 2020. Genes with Spiralian-Specific Protein Motifs Are Expressed in Spiralian Ciliary Bands. *Nature Communications* 11 (1): 1–11.

Wurzinger-Mayer, Andrea, J Reuben Shipway, Alen Kristof, Thomas Schwaha, Simon M Cragg, and Andreas Wanninger. 2014. Developmental Dynamics of Myogenesis in the Shipworm *Lyrodus pedicellatus* (Mollusca: Bivalvia). *Frontiers in Zoology* 11 (1): 90. https://doi.org/10.1186/s12983-014-0090-9.

Xu, Fei, Tomislav Domazet-Lošo, Dingding Fan, Thomas L. Dunwell, Li Li, Xiaodong Fang, and Guofan Zhang. 2016. High Expression of New Genes in Trochophore Enlightening the Ontogeny and Evolution of Trochozoans. *Scientific Reports* 6 (September): 1–10. https://doi.org/10.1038/srep34664.

Young, Craig M, Mary A Sewell, and Mary E Rice. 2002. *Atlas of Marine Invertebrate Larvae*. Vol. 6. Academic press San Diego.

Yurchenko, Olga V, Olga I Skiteva, Elena E Voronezhskaya, and Vyacheslav A Dyachuk. 2018. Nervous System Development in the Pacific Oyster, *Crassostrea gigas* (Mollusca: Bivalvia). *Frontiers in Zoology* 15 (1): 10.

Yurchenko, Olga V, Anna V Savelieva, Natalia K Kolotuchina, Elena E Voronezhskaya, and Vyacheslav A Dyachuk. 2019. Peripheral Sensory Neurons Govern Development of the Nervous System in Bivalve Larvae. *EvoDevo* 10 (1): 1–9.

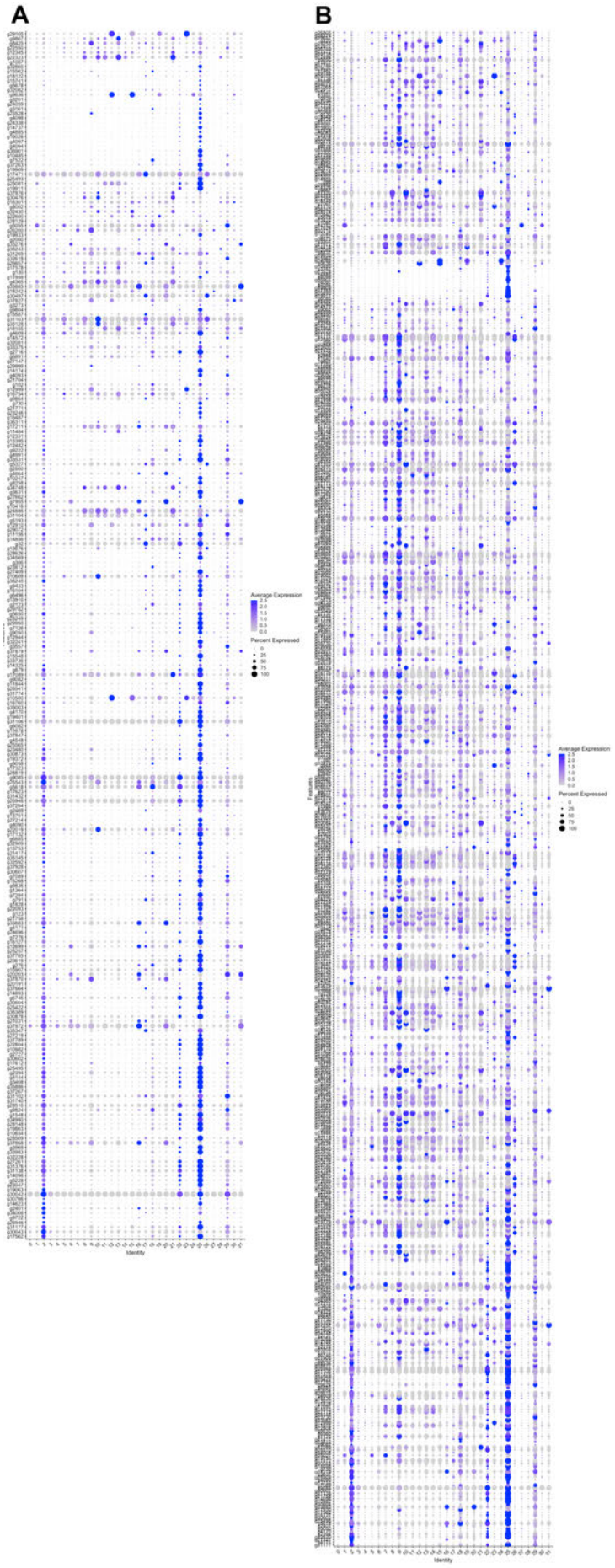
Zayas, Ricardo M, Francesc Cebrià, Tingxia Guo, Junjie Feng, and Phillip A Newmark. 2010. The Use of Lectins as Markers for Differentiated Secretory Cells in Planarians. *Developmental Dynamics* 239 (11): 2888–97.

Zhang, Guofan, Xiaodong Fang, Ximing Guo, L I Li, Ruibang Luo, Fei Xu, Pengcheng Yang, Linlin Zhang, Xiaotong Wang, and Haigang Qi. 2012. The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation. *Nature* 490 (7418): 49–54.

Zhao, Ran, Takeshi Takeuchi, Yi-Jyun Luo, Akito Ishikawa, Tatsushi Kobayashi, Ryo Koyanagi, Alejandro Villar-Briones, Lixy Yamada, Hitoshi Sawada, and Shunsuke Iwanaga. 2018. Dual Gene Repertoires for Larval and Adult Shells Reveal Molecules Essential for Molluscan Shell Formation. *Molecular Biology and Evolution* 35 (11): 2751–61.

Appendix 1: Genes co-expression in ciliary clusters

Genes co-expressed between the oyster ciliary cells and cluster Pc11 (ciliary lobes and apical tuft) are very specific to ciliary clusters (cluster Cg 2, Cg 22 and Cg 25 in figure A). However, genes co-expressed between the oyster ciliary cells and cluster Pc10 (ciliary lobes) are also expressed in neoblast clusters (cluster Cg 7 and Cg 9 in figure B). Result shown here are co-expressed ciliary genes generated by SAMAP in oyster scRNA seq.


A
B
Average Expression
Percent Expressed
Identity
Features

Appendix 2: Sequences of gene primers used for chromogenic *in situ*

Cgi_ = *C. gigas*. Pcr = *P. crozieri*. _fd = forward primer _rv = reverse primer

Gene name	Gene ID	Sequence	Probe size (bp)
Cgi_Tektin_fd	g25422.t1	CTTGACACACTATGCAAACGC	1035
Cgi_tektin_rv	g25422.t1	AGTCCCTTGAGTGAGTCCTC	
Cgi_b-tubulin_fd	g31762.t1	AACAACTGGGCTAAGGGACA	943
Cgi_b-tubulin_rv	g31762.t1	ACTCAGCCTCAGTGAACTCC	
Cgi_Tyrosinase_fd	g35736.t1	TGCGGAAAGGACAACTCAAC	1459
Cgi_Tyrosinase_rv	g35736.t1	GTAGCTCCTGGTGATGTGGT	
Cgi_Prisilkin_fd	g11820.t1	GGCGCCATTCACTCTTTGG	555
Cgi_Prisilkin_rv	g11820.t1	ACAAGATTGGTGTAAGAGCCA	
Cgi_Engrailed_fd	g30277	GGAGGTTGCACACGAACAAA	590
Cgi_Engrailed_rv	g30277	TGTTATTGCCTCCCGTGGAT	
Cgi_SoxE_fd	g27288	ATGAGCGACACTGACGAAGG	1335
Cgi_SoxE_rv	g27288	CGAGCGATGGTTGGACAAAT	
Cgi_Dopamine-beta-hydroxylase_fd	g5709	GCCGATTCTGACGATGATGG	933
Cgi_Dopamine-beta-hydroxylase_rv	g5709	TGTGGTCCGTAACTCATCCC	
Cgi_Mucin_fd	g10584	CAGGTCTTGACGCCATGATG	1109
Cgi_Mucin_rv	g10584	TGTAGATCCTATGGCTGCGG	
Cgi_Hs3s5-1_fd	g22525	TTTGGCTGTTTGGAATGGGA	702
Cgi_Hs3s5-1_rv	g22525	ACTCTTGACTGAAGCCCGAG	
Cgi_FRMFamide-receptor_fd	g33239	GGATCTGGGTCTACGTGTCC	711
Cgi_FRMFamide-receptor_rv	g33239	ACATGCTCGTTACACTCGGA	
Cgi_Pax6_fd	g27529	CGACACCGCCAGTAATTCAG	1161
Cgi_Pax6_rv	g27529	GGATCGTAACTTCGTGGTGC	
Cgi_Synaptotagmin_fd	g22701	AGAGGTGTAGAGGTGGTTGC	1342
Cgi_Synaptotagmin_rv	g22701	CCAGCATGTCACTCCAATGG	
Cgi_7B2_fd	g1151	TGTCTGTTCTGGCCATTTCG	651
Cgi_7B2_rv	g1151	CCTGCTCTGCCTCTCTCTTA	
Cgi_NeuronalAcetylcholineReceptor_fd	g32208	CTCAGAGTACCAGGAGCGAG	1506
Cgi_NeuronalAcetylcholineReceptor_rv	g32208	GTTCTCCGCCTGCATTTCAT	
Cgi_Somatosin_Receptor_fd	g25548	AGGGTACAAGGATGATCGGT	902
Cgi_Somatosin_Receptor_rv	g25548	TGGCTCGCATGAACTGAAAT	
Cgi_BMSP_fd	g10452	TGCCACCAATGAAAGGAAGC	1097
Cgi_BMSP_rv	g10452	ATCCCATGCCCTCGACATAG	
Cgi_Marker_31_3_fd	g7838	CGCTCTGTGTGTTCTACTGA	527
Cgi_Marker_31_3_rv	g7838	TCTCCTCTTGTTCCGTTGTG	
Cgi_myosin-9-like_fd	g14190	TGGATGAGCTGGAACTTCGT	1039
Cgi_myosin-9-like_rv	g14190	AGGTACATGGGTGGGTCTTG	

Cgi_irx-2_fd	g21502	TCCTTCCAGCACGTTAACCT	808
Cgi_irx-2_rv	g21502	GTTAGACTGAGGCTGTGGGT	
Cgi_Marker_hemocytes_3_fd	g20442	GGGCTATTTCCTGAGTCCGA	1296
Cgi_Marker_hemocytes_3_rv	g20442	CTTCAGGGCTTCGAACAAGG	
Cgi_SDE-2_fd	g16728	GTCTCCAAGTTTCTGTGCCG	945
Cgi_SDE-2_rv	g16728	TGGCTGTCATCGAGTTCTGT	
Cgi_troponinT_fd	g5835	AGCCGAGTTTGAAGAGCAGA	700
Cgi_troponinT_rv	g5835	GTAGGATCGGTGGTCTGTGT	
Cgi_APOBEC 1_fd	g22246	AGGGAGCAACAGAGGGTATG	851
Cgi_APOBEC 1_rv	g22246	GATTTCCAGCACCACGAGAC	
Cgi_mab21-like-2_fd	g16585	ATATGTTAGCGGCCCAGTCC	878
Cgi_mab21-like-2_rv	g16585	GTCCCATTCCATTTCCCGTG	
Cgi_collagen-alpha-1-VII_fd	g16652	CGCCGTCACATCCATGAAAT	842
Cgi_collagen-alpha-1-VII_rv	g16652	CGGTCATGGCTAAAGTTCCG	
Cgi_GNQQNxp_fd	g16262	GTTCCTGCTCTCCCTGACTT	726
Cgi_GNQQNxp_rv	g16262	CCTGACTCCCTGATGCTCAA	
Cgi_cluster_7_fd	g24584	ATGAAGCAGTACCCACCCG	402
Cgi_cluster_7_rv	g24584	CGTGAAATCCGGCAATGACA	
Pcr_troponin_fd	g14959	AATGAAGAAGCGACGTGAGC	847
Pcr_troponin_rv	g14959	CGTGTAAGGACCGGCAATTT	
Pcr_Vasa_fd	g3163	AGATGCGCCTCCTTTATCGA	1016
Pcr_Vasa_rv	g3163	TACCAGCATAACCGGCATCT	
Pcr_pero_opsin_fd	g25197	TTGCACGCGAATATTCAGGG	712
Pcr_pero_opsin_rv	g25197	GTAGCGTAACCAGACCAGGA	
Pcr_Ropsin_fd	g18600	TCCCTGTCCTTTTCGCCAAA	523
Pcr_Ropsin_rv	g18600	TATTACAACGGCCCCCAACC	
Pcr_5HT_fd	g13280	GACGCATTCAAGACCTCGAT	1008
Pcr_5HT_rv	g13280	TAGACTCCACACTTCCGACG	
Pcr_mucin_fd	g20284	CCCAAAGCACAGCATCCTTT	1461
Pcr_mucin_rv	g20284	CACATCCTTCGACGCACATT	
Pcr_jumonjii_C_fd	g21369	GCGTGAAGTCAGCAAAGGAA	1035
Pcr_jumonjii_C_rv	g21369	TGTTTAGTACGATGGCAACCTG	
Pcr_FRMFamide_receptor_fd	g491	GACAGAAGAACCCACGGACT	1039
Pcr_FRMFamide_receptor_rv	g491	ATAGTTCCCAGTTCTCCGGC	
Pcr_metalloproteinaseNAS13like_fd	g3137	ATGAGGTTCTTGCCCGATCA	1042
Pcr_metalloproteinaseNAS13like_rv	g3137	CGCCAACTTGAGCCCAATAG	
Pcr_GuanineNucleotideBinding_fd	g24247	GAGCAATGGACGCACTGAAA	809
Pcr_GuanineNucleotideBinding_rv	g24247	GGCAAGGCTGTAATCTGTCA	

Appendix 3: HCR probes for *P. crozieri*

Pool name	Gene id	Sequence
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCTGGCCTTTCGGGGTCATTTCCACG
B3_Mc_rops_23_Dla0	g18600	TTATGCCCGCTCAAAAGCTTGGTTGttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttACGCTTTCTTTGTTGACGTTGTTGT
B3_Mc_rops_23_Dla0	g18600	GAGATGTTCTTGGTTTCGTTTCCCTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCGCTGACCGCACCTCTACCTGCTGG
B3_Mc_rops_23_Dla0	g18600	AGTCCATGTTGGAGACGGAGCTGGTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttTTCTTTGGCTTTGGGTGGACAGCAG
B3_Mc_rops_23_Dla0	g18600	GCTGGCAGTCGTGCTTGTGGTCTCAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttTGAAGGCGGAATTTTGGATGGGTCA
B3_Mc_rops_23_Dla0	g18600	ATGAGCCATGGGAATTTCTTCTCCAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttATGATTTGGCGAAAAGGACAGGGAT
B3_Mc_rops_23_Dla0	g18600	CATAGACGATGGGATTGTAGACAGCttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttTCGTCCTACCAAAGACATCAAAGCG
B3_Mc_rops_23_Dla0	g18600	TGATGTGAAAGGTGTCAGATGATCTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttAAGAGGACAAAGATGGTGACTGAGG
B3_Mc_rops_23_Dla0	g18600	AAAGCGTATGGGAACCAGGAAAGGAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCACTGCTCTTGAATTGAGATTTCAA
B3_Mc_rops_23_Dla0	g18600	TAGCGGCTTGAATATCAGATTTCTTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttGTGCTGTCGCACAGCTTTCACGATT
B3_Mc_rops_23_Dla0	g18600	CTTAGCCATTGCCATCATTTCAAGTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttGGAAGAGTAAAGCCAAAGATGTACA
B3_Mc_rops_23_Dla0	g18600	TAATAGCAGAAGATGATGGTAAGAAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttTCTGAGAGAGATAGTCAAAAGTACA
B3_Mc_rops_23_Dla0	g18600	CAGCATTGAAGATGATGTTTCCTTTttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttTCCGTAGCCGAAAAATGGTGGAATA
B3_Mc_rops_23_Dla0	g18600	AGTTTGGAAACCCTCAGGAACATAAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCTGCACATAACCAAGGCTCTAGTCT
B3_Mc_rops_23_Dla0	g18600	CAAAGGAGACTCCAGATCCAAGCGAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttGTACAATAACAAAGAATCTGTCTAC
B3_Mc_rops_23_Dla0	g18600	TCAATGATTTCAATTGCTCAAAAGGttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttAAAGCCAAAGAGGCCACCTATGAAG
B3_Mc_rops_23_Dla0	g18600	GATCAAGGCCATTGTGGTTATGGAGttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttGCCCACCTGGTGTTAAATGAGGACA

B3_Mc_rops_23_Dla0	g18600	TACATTTCACAGGTGAGTCTACCCCttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttAGAAAGTAAGATCAGAGATAGCTAA
B3_Mc_rops_23_Dla0	g18600	TGAGGAGTGGGAATCCATTTACAGCttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCAAAGATTTGGTTACACTGAAGATG
B3_Mc_rops_23_Dla0	g18600	GATGATAAAGAAATTTGGAGCAGTAttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCAAATCCCAACTACAGAAATAAAGA
B3_Mc_rops_23_Dla0	g18600	ATGACCAAAAGATTGCCTACAATTCttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttCCGGTGGAAAATTTCGCCAGTGGGG
B3_Mc_rops_23_Dla0	g18600	CAACAAGGTAATGATATACCTCTGGttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttATGCCATGGGGCCACTGTACCGTTA
B3_Mc_rops_23_Dla0	g18600	AACGATCGCTTCAAAGTCGGCATCCttCCACTCAACTTTAACCCG
B3_Mc_rops_23_Dla0	g18600	GTCCCTGCCTCTATATCTttGAACTTGGAGTCGAAGCTAGTTGAG
B3_Mc_rops_23_Dla0	g18600	GGTGCAAGGCTTCCATTAGCAAGCAttCCACTCAACTTTAACCCG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaACTGGCCTCTTGAGGTGGCTCTTCT
B1_Mc_trop_18_Dla0	g14959	CTACTCCGCTTCAATCTCAACTTCAtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaGCTTCTGCCGGATGGGATCTGACAG
B1_Mc_trop_18_Dla0	g14959	GGGAGGCCGTTTTCCGTAGGGGCCGtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaCACCTTGGTCAGTCATGACGATTTG
B1_Mc_trop_18_Dla0	g14959	GACCCGTGTAAGGACCGGCAATTTCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTTATATTGGGAGCCATCGAACACC
B1_Mc_trop_18_Dla0	g14959	CTGTGGCTCTATTCTGGTATACAGAtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTTACACGCTCATACTGACTGTACA
B1_Mc_trop_18_Dla0	g14959	CGTCTTTCAGTGAATGAGCGGCGATtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTTCACTAAGTGGGTCAGGTTCCGT
B1_Mc_trop_18_Dla0	g14959	CAACTTTAGGAGGAATTCCGCTAAAtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTTGTTCATTTGACGTGCTCGCTCG
B1_Mc_trop_18_Dla0	g14959	TTGAACGTTGACTTTCTTTTTGTCAtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTGAATCTATCCTCAAGATCATATT
B1_Mc_trop_18_Dla0	g14959	AATTCCATCAAATCATACTGTTGCCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaACAATTCCTTGGCCTTATCTGCCAG
B1_Mc_trop_18_Dla0	g14959	CGCCTTCTAATCTGTAAATATGATGtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTGGTTGGATTCTCTGTTCCAAGATG

B1_Mc_trop_18_Dla0	g14959	TTGCACATTGAAACCATCAATGGATtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaGATTTCGCCATTTCTTGTTGTGATT
B1_Mc_trop_18_Dla0	g14959	CGTTTTTCATTCTCAAGTTGCTCTTtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaCACTGCGTTTGGTAATGACAAAATT
B1_Mc_trop_18_Dla0	g14959	TGGAACCGCCATTTAGTTCCTGACCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTCGTTCTTCTTCTTTATCTCTCGC
B1_Mc_trop_18_Dla0	g14959	TTTGCCCACCAAATTTTTTTGCTCCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaTTGATTCGCTGTCTCTCTTCCTCTT
B1_Mc_trop_18_Dla0	g14959	TCACGTTTTTTCTGTTCAGCCTCTCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaGAAGCTCAGCCATTCGCCTTTCTTC
B1_Mc_trop_18_Dla0	g14959	CCCTTCTCCTTTGCTCTTCTACAGCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaCTTCTCACGAAGTCGTTGAATCTCC
B1_Mc_trop_18_Dla0	g14959	TTCGCGTTCCTTTTTTCGTTTTTCCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaATCTCTTCATATTCACGCCATTCTT
B1_Mc_trop_18_Dla0	g14959	TCCTCTTTCTTCCGTTCTTCAGTTCtaGAAGAGTCTTCCTTTACG
B1_Mc_trop_18_Dla0	g14959	GAGGAGGGCAGCAAACGGaaCCATCTCCATTCGAAGCTCGGCTTC
B1_Mc_trop_18_Dla0	g14959	CATCGCGCTCACGTCGCTTCTTCATtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGTAACTTCAGTGCCATTGTTACCC
B1_McVasa_33_Dla0	g3163	CTACCATCCATCATCATCATTTGTCtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCCCGACATGAAGTCATTATCCTCAT
B1_McVasa_33_Dla0	g3163	CTAGGTTCAGTTTCATAATTCTGACtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaACCTCGGTACGTTGCGTCGCATGTC
B1_McVasa_33_Dla0	g3163	TTTCGCCTCTGCCGAAGTTATGTCTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGCATAACCGGCATCTTTCTGAAGG
B1_McVasa_33_Dla0	g3163	TGACTGTCCTTGTCTTGAGCCAGTAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGAGTGTCAACGAGGAAGTTCGCGA
B1_McVasa_33_Dla0	g3163	CCAGGCACAGCCTGTCCAGCCTTTTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTGATTGCTTTTCCTTTGTTGCCGCA
B1_McVasa_33_Dla0	g3163	CGGCATCTGATGAGGGATCGTAGAAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCGAGTCGATATCAGAAGGAATGTTG
B1_McVasa_33_Dla0	g3163	ACCAGTCCGTCCAATCCTATGGACGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTCAAGACCTCTAGCTGCTACAGAAG
B1_McVasa_33_Dla0	g3163	TTAACGACATAATCGACTTCTGGGAtaGAAGAGTCTTCCTTTACG

B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTGAAGGTACGCAGGGCTTCCTCACG
B1_McVasa_33_Dla0	g3163	CAACAAGGAGCATGTGTTTTCCTCTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTGTTGCTGAGAAGCCCAACTCACAC
B1_McVasa_33_Dla0	g3163	CGGCTGCTCTCGGTCACCATGTATAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTTTTGTTCGACGAAAACCAGTACTT
B1_McVasa_33_Dla0	g3163	TCCATCCCCAAGCAATCAGCCTTTTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCACTCATCAGTATATCACGTAAACA
B1_McVasa_33_Dla0	g3163	TCGTCATTCGTCCACTGCTATCGCGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaGTCAGGAGATTCAACCAAAATTATT
B1_McVasa_33_Dla0	g3163	ATCGCGGTTTTTTACTTCACCATCTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaACAATGCCAACAGCAAGGAAAAGAT
B1_McVasa_33_Dla0	g3163	TGTTCAACATCTGCACATGCAGCGCtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaGAACTTCTTTTGGAAAAGTTGCACT
B1_McVasa_33_Dla0	g3163	CTTTGCAGTATTTAGCAGCCATCTGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGTCATTCTACTGAATATCTCTTCG
B1_McVasa_33_Dla0	g3163	CATGAGTACTTGCTTAGGATCATCCtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCGATCCGCCTCGTCCAAAACTACAT
B1_McVasa_33_Dla0	g3163	TCAGGGGCAAAGCCCATATCTAGCAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTCTCCAAGAAATCCTTCAATCTCCC
B1_McVasa_33_Dla0	g3163	TTAGGAAATTGAAGGATATCCATCCtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaGTCCGAACGTATATGATTTTTCAGA
B1_McVasa_33_Dla0	g3163	AGTAGCAGATAATATGTGCGCTCCAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCGACCATTTATGCCAGTATTTTTAG
B1_McVasa_33_Dla0	g3163	GCGGTTTGAATTTTTCCATACAACTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGACCAATTCCCTCGTCGGACCGAC
B1_McVasa_33_Dla0	g3163	ATTTCTCACTCCACTGATAGGTCTGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaATTCGATAAAGGAGGCGCATCTTCT
B1_McVasa_33_Dla0	g3163	GAGAGCTCTAGGAAATACTGTGATGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGAAATGCAGCTGTTTTGCCCGAAC
B1_McVasa_33_Dla0	g3163	TTCAACAAGTAATCTATTATGGGCAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCAGCTTGAACTATTGGTATGGAATA

B1_McVasa_33_Dla0	g3163	TCATCGAACATGCCATCAAATCTCGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaACGTGGAATATTGGCAAGAAGTTGA
B1_McVasa_33_Dla0	g3163	CTGCACCGGAGTTGGCTTCGTGAATtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaATAAACTTAGGCTTCTCGCGCCCTG
B1_McVasa_33_Dla0	g3163	TGTAAATTGGCTTCCTTAAAAGAAGtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaCGAAATTTATCCCCGCCTTTACGAT
B1_McVasa_33_Dla0	g3163	CTTCGACATCGATCCTGTCATAATTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaGATGTAAGTCGAAGGAGGAGGACCA
B1_McVasa_33_Dla0	g3163	ATCATCGTCGTCATTATAAGCTTCAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTCTGTACAATCACGAGCTAAGTGAC
B1_McVasa_33_Dla0	g3163	CCTGAAGGTCCATCACCACTTGGTTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTGTCTTCTTGAGGGCAGTCTCTGGC
B1_McVasa_33_Dla0	g3163	TTTGATTACATTTGTGACACGTTCTtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaAGTGCGGTCACCACCGCCGCCGTCA
B1_McVasa_33_Dla0	g3163	ATGTCCCTCTTGATTACATCGCCGAtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaGGTTGATTACATTTCCGACAGTTTC
B1_McVasa_33_Dla0	g3163	TCGGGACACTCCCGCGCGAAATGTCtaGAAGAGTCTTCCTTTACG
B1_McVasa_33_Dla0	g3163	GAGGAGGGCAGCAAACGGaaTTTCGGGACACTCCCGTGCGAAATG
B1_McVasa_33_Dla0	g3163	CACCACCTCCAAAACCCCCACCGTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCACGCGTGTCAAAGTCTCCACGATT
B1_Mc_g7263_33_Dla0	g27905	CTTTCTGATGCAAACATTGATTCCTtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCACTGTGTGCGATGAACGGTCCTCT
B1_Mc_g7263_33_Dla0	g27905	ATATGAACAACCCTGTTATTCTTGAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTGACGATTCGTCGGCAACGATTTCT
B1_Mc_g7263_33_Dla0	g27905	AACGACAAGGTTCTTTCACTATACGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaAATTTTGCGTACATTTCTACGGCGA
B1_Mc_g7263_33_Dla0	g27905	GGTGTCCAAAACCCATTTTAAGTAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCATCGACAAATTCTCTTTGCGATCT
B1_Mc_g7263_33_Dla0	g27905	CTTTCTCTGCTGATGCGATCTTTAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGTCGGTGTGTTGTGTATATCACACG
B1_Mc_g7263_33_Dla0	g27905	GTGAGCGACGACACTTGCACCCAATtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGACACGTACTTTGTCTCGAGGACAA

B1_Mc_g7263_33_Dla0	g27905	ATATCTGGATTTTCGTGGACATCCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaAGAATACATCGAGGTCCTGTTAATC
B1_Mc_g7263_33_Dla0	g27905	ACGGCGACCTTCTTTTGTCGTGAAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCCAGGCGTTTGTTACACGTCTTGGT
B1_Mc_g7263_33_Dla0	g27905	ATAACGATTTTGTGGTAATTTTCAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTCGACGTTTCAGCCGTATACGCAAT
B1_Mc_g7263_33_Dla0	g27905	TCTTCCCTTTGGACAACAGCATTTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTTGATTGTCCTGATGAAGGAGCACT
B1_Mc_g7263_33_Dla0	g27905	CATCGACAGCTGCGCATGTGATAGAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTTCTGGTCTTTTTGCATTTTAGGAA
B1_Mc_g7263_33_Dla0	g27905	AGCGGTTGCATTTTGTGTCATGCACtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGCGATTGTTTCGCTTATCTAGGCGA
B1_Mc_g7263_33_Dla0	g27905	TTTGTTAATGGAAGTGCTTGAGATAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTGATTCACGCAAATCTGCTTCACAG
B1_Mc_g7263_33_Dla0	g27905	CTCTTCGTTATTATTCTCAAAAGTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTGGTAATCTTTCTCCGCACACATTT
B1_Mc_g7263_33_Dla0	g27905	TCTTTGGACAACAGCAGCGTTTTCTtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCGCTTTTGAGATGCACGACTTTTTG
B1_Mc_g7263_33_Dla0	g27905	CCCTTTGCGTCGAATCCTGAAATATtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGTTCTGCAGTAAACTGGACTGTACT
B1_Mc_g7263_33_Dla0	g27905	CATTTCCACTTCAGGACTTGTTTTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCAACCAATCGTCTACAGACTCTCTT
B1_Mc_g7263_33_Dla0	g27905	TCTTGACGACTTTGATGCATCGTCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCTTCGTGTTATGACACTTTGGACAG
B1_Mc_g7263_33_Dla0	g27905	CCGATATCGGTGTCCATAGCAGCGTtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGTTTTGCATCGACACTGTCGATGAA
B1_Mc_g7263_33_Dla0	g27905	CAGCGTCTCATTCGATGGATTCGAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCGTACATTGTTGCCTTGTTGCATCG
B1_Mc_g7263_33_Dla0	g27905	TGACTCTCTTCACCAAGGTATCTCGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTGTGAGAACAGTCCATTTCTTCGCC
B1_Mc_g7263_33_Dla0	g27905	ACGATGAACACGAACTCTTCTGCTAtaGAAGAGTCTTCCTTTACG

B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCTGAAAATGAAGCTAACTGTGACTC
B1_Mc_g7263_33_Dla0	g27905	CTAACCAAACAAGTCTTTGTTCGGCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGACTTTCTTTGATCGGAGCTCGGGA
B1_Mc_g7263_33_Dla0	g27905	GGATGTTTCCTGTATGTCGAATGCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaACGACGAATACAAGATCCCTTGATT
B1_Mc_g7263_33_Dla0	g27905	ACAGCATCTCTCTCTTTTCACAGATtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCGTTTACAAGTTTTGCGGTTGCAGT
B1_Mc_g7263_33_Dla0	g27905	CCTGCCCAAATAGTATAAATGATATtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaAGCAAATTTCCTTCTTTATACGAAG
B1_Mc_g7263_33_Dla0	g27905	GTTTAACCACTCTTCGTCGGGGACAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaACGGAAGGACAAGATCTTGATTGTA
B1_Mc_g7263_33_Dla0	g27905	ACGAGGCTTGCATTTGCAATGTTTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaCGACGGACTATCCTTGGACATTTAA
B1_Mc_g7263_33_Dla0	g27905	TATCTAGTGCGTCGATTACATGGTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaTGACGCAGCGATTATGACGAAGATG
B1_Mc_g7263_33_Dla0	g27905	TCCGCTTCAATTTTCTGCTTTTCAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaATGTGCTTGACATTCCCTCTCAACC
B1_Mc_g7263_33_Dla0	g27905	GGTTGTTGTGATTCGATATTTGAAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaGAAGAAACTCTGCGCTTCAGAATGC
B1_Mc_g7263_33_Dla0	g27905	CGGTTTGGTCGGCAGCAGCAGACCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g7263_33_Dla0	g27905	GAGGAGGGCAGCAAACGGaaATGTGATCTTCTGACGACAATTTTT
B1_Mc_g7263_33_Dla0	g27905	GACATGCCGAACGGTGATTGCTTAAtaGAAGAGTCTTCCTTTACG
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaAAATACCGTCCTGCTTGTTGTCGTC
B2_Mc_Macif1_28_Dla0	g2413	TTAAAATTGGCCGCTTGAGGATGACaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaGAAGCTGATGATCCGCTCATGCTGG
B2_Mc_Macif1_28_Dla0	g2413	ATCTGTGAGCCACCTGAACATTCAAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaAGCTTGTGGATGAGCTTCCTCCCAT
B2_Mc_Macif1_28_Dla0	g2413	CTTCAACAACTTTGGTTGAGCTTGAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTACGCCACCAACAATAACACCTCCA
B2_Mc_Macif1_28_Dla0	g2413	AATAATTTCTCCTCCGGAGCAAGTGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCGAGCCTCTTCACCTTCCAGGAGTT

B2_Mc_Macif1_28_Dla0	g2413	GATGATGAGTACGAGATTCCGGACAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTAAGCTTGGCATCAAGGAGGGTCTG
B2_Mc_Macif1_28_Dla0	g2413	TGTACGCCACGATTTCAAGCTCAAGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaACGAGTTTCAGTAAGGGAACGAGTA
B2_Mc_Macif1_28_Dla0	g2413	TTCTTCCAAGAGCGAGGCCAATTCAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCGCATTTCCTCAATTTCTCTAGAGA
B2_Mc_Macif1_28_Dla0	g2413	GTCGAAATTTCAATTTCTTGATCACaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTTCTTGTCTCCAATTCAGGAACAGT
B2_Mc_Macif1_28_Dla0	g2413	TTTGGATTTGGGCCTCCAATTGAGAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaACGAGTGTTGTCTTCCTTGATGTGT
B2_Mc_Macif1_28_Dla0	g2413	TCTCAGTTTGGTCAATTGTGTCTTGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaGTCTTGTACTCGGTAACTTTGACGC
B2_Mc_Macif1_28_Dla0	g2413	ATTTCCATGGTAGTGGTGGTGTTGGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTCTCCAATTTAGTATCATACTCATT
B2_Mc_Macif1_28_Dla0	g2413	ACATGGTTTCCAACTCTGTCTTGACaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTGCTTTCCAGAAATCGCGGTTTTCG
B2_Mc_Macif1_28_Dla0	g2413	GATATCATGAATAGCCTGGGCCAATaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaAACTCTTTGAGTTCCGCTTCATGAA
B2_Mc_Macif1_28_Dla0	g2413	GAACTGTCACGATAAGCCAAACTGGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCTGTTTGGCACTCATTCTCGGCATT
B2_Mc_Macif1_28_Dla0	g2413	GCTTGTGGAATTCAAGCTCTTCTCTaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCGTTCGCAATCTGTCGATTTCCACA
B2_Mc_Macif1_28_Dla0	g2413	GTGAATGAGAGTTTCATGATCGAGGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCCGCGGTCATCATCGAAACGTGAGA
B2_Mc_Macif1_28_Dla0	g2413	AGTCGGTTAACTTCGCTTCTGTCTCaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCGTAGTCAGAAATCAGATGATTCAG
B2_Mc_Macif1_28_Dla0	g2413	TTTTGCGGAGGATGACAATCTCACTaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTTGTTCGCGAGCCTCGTCAATGAGA
B2_Mc_Macif1_28_Dla0	g2413	TTCAATGGTCTCTCTATCAATTGTCaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaGAAGTGATTTTGACCTCAAGTTCGG
B2_Mc_Macif1_28_Dla0	g2413	TGAATCTCCGAAACACGATCTTCCAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTTCTCAGTGCCTTCAATTCGCCTTC
B2_Mc_Macif1_28_Dla0	g2413	TGCGACGTTCGCATTCATCCAATTTaaATCATCCAGTAAACCGCC

B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaAGTCCATCGTTCTCTGAGCGTCGTC
B2_Mc_Macif1_28_Dla0	g2413	TGTAACTTTCACTTTAGCCGATTCTaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTCCAAAAGGCGCACTTTCTCAATGT
B2_Mc_Macif1_28_Dla0	g2413	TCAACAGCAAGCTTTCGGTTTTGGGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaAATCCGTCTTTTCTCTTGTTCGGCC
B2_Mc_Macif1_28_Dla0	g2413	CAGAAAACCGTTGGTTGAGGTTCTGaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCATGGCTGCGGTAGCCCCTGCGCTC
B2_Mc_Macif1_28_Dla0	g2413	CACGATGTTATTGACGCCTGTAGCAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCCAGAGCCTCCACCGACAGCGAATG
B2_Mc_Macif1_28_Dla0	g2413	GAGAAAGAGCGGTAATCTTCACGAAaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaCTGCACCATATCCATGGGTTCGAGC
B2_Mc_Macif1_28_Dla0	g2413	CTGAGGCCGATGAGGCACTGGCTCCaaATCATCCAGTAAACCGCC
B2_Mc_Macif1_28_Dla0	g2413	CCTCGTAAATCCTCATCAaaTGAGGAGGAAGAAGCCATTCTTGCG
B2_Mc_Macif1_28_Dla0	g2413	AACAATTTGTCGACCAGAGGACGAAaaATCATCCAGTAAACCGCC
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaAGCAACTCCACATTGATTGTTCTTG
B1_Mc_g16674_18_Dla0	g26945	TCAAACTAGGGGATAACTGGCTTGTtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCCCCAAGATTCTCCCCAACTGTTCT
B1_Mc_g16674_18_Dla0	g26945	CTGGACATGTGGATATATCCATTCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaTTCCATAGCCAACAGCCAAGACTCC
B1_Mc_g16674_18_Dla0	g26945	TTAGCCAGTAATCTTTGCCGTCCTCtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaTTCATTGTAAACGCCGCTTCGATAC
B1_Mc_g16674_18_Dla0	g26945	ATCCAGTTTGGTGCGACTGCACCGGtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCCAACAGAAATAGGACCAATAGTTG
B1_Mc_g16674_18_Dla0	g26945	TGGAAACTAGCATGACTGGCGTCAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCTTCTTGGATGTCTACGTATCCTGT
B1_Mc_g16674_18_Dla0	g26945	CTGCGTTCTTCAAATCATCTTCGCTtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaGCGACAGAAGAGACCGGTCTTTGCT
B1_Mc_g16674_18_Dla0	g26945	AGTGGCTCCAACATTTTCTTTTTCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCCTTTGTTGGCTTTAATGTATCGGA
B1_Mc_g16674_18_Dla0	g26945	TAAGGATAAGATTGTTCAGTATCGAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCATTGTTTCCATAGTGTTTGCTGCA
B1_Mc_g16674_18_Dla0	g26945	CGTTGTCCATGAGGCCACCTTTGCAtaGAAGAGTCTTCCTTTACG

B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCAGAACTCCTGTTTTCTTTTTTGTT
B1_Mc_g16674_18_Dla0	g26945	GACCAAGTTTTGTTCTGACAGGGAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaGCCCAACAAGATCCACATTGTTTCT
B1_Mc_g16674_18_Dla0	g26945	CCTTCCAATGAGCCAGTTGCACTGAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCACGCCAATCGACAGTTTCTGGAAG
B1_Mc_g16674_18_Dla0	g26945	TTTTTACATCAGTGACGTATCCAAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaTGCCTTAGGTCCAGATTTGGTCTGG
B1_Mc_g16674_18_Dla0	g26945	GACATTCTTGGCAGGGAGGAACGTAtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaTATTCATCGCCAGTTAAGTCTCCGT
B1_Mc_g16674_18_Dla0	g26945	TATCGGTAGCCAAGCAACATGTCCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaCATTCATGTCAGCTTCAAGATTGTG
B1_Mc_g16674_18_Dla0	g26945	CATTCATGCCAAGACGGTAAGTATGtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaAATCATTCGTCGTAAAGTTTCATGT
B1_Mc_g16674_18_Dla0	g26945	TTCAATGATTTTCAAATTAGTCTCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaGTCTTGAAGAGCTCCCAATGATTGT
B1_Mc_g16674_18_Dla0	g26945	GGACTTTCATATTCCTTCTTATAAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g16674_18_Dla0	g26945	GAGGAGGGCAGCAAACGGaaGTGTTCCGGCTGAACAGCTCAAAAC
B1_Mc_g16674_18_Dla0	g26945	ATTCCTCATTCACATTTCCGAATCGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGATTGTGGATAATACTCTCAGTCTG
B1_Mc_g30984_20_Dla0	g29271	CTATTTGCGTCTCTGGTTAAGGCTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGCGGCCATCAATCGTCTCATGCTCT
B1_Mc_g30984_20_Dla0	g29271	CGCTCGTCATCGCGGTCTTTCTCTTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCTCCCTCATGTTCTACCAGACCGTA
B1_Mc_g30984_20_Dla0	g29271	GTTCCAGTGGTACGAAGAGTCCAGGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaAGAGACAGGATCAGTCTCCTGTTCC
B1_Mc_g30984_20_Dla0	g29271	TACTGGTCCGTATTCTGACTCATATtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaTCAGGCACTCGTTTTGCAACGGCGT
B1_Mc_g30984_20_Dla0	g29271	TCTACCTCCTCAGGAGCTTCCTCAAtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGTTCGACTTCCAGAGGATCATATTC
B1_Mc_g30984_20_Dla0	g29271	GACTGTAGTCGTCGAGCATTCCTCGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaATCAACAACTTCTTCAGGAATCTCG
B1_Mc_g30984_20_Dla0	g29271	TGGGACTTCATCCTCAGCATAACTTtaGAAGAGTCTTCCTTTACG

B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaTCTAACCTAGCCAATTGATCAGGAT
B1_Mc_g30984_20_Dla0	g29271	CCATCATTGGAGCGAGATAGCATGTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGGACGGAATTCATGGACATATGGGA
B1_Mc_g30984_20_Dla0	g29271	TGTCTTCTATATCACGTTTCTTCCTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaAATGCGATTGAGACGAGTGCCGGTC
B1_Mc_g30984_20_Dla0	g29271	GGATGCCATCTTTGACATATCACGGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCGTGATCCAGACTCATTCTCCAACA
B1_Mc_g30984_20_Dla0	g29271	TTTGTTGTTGCTGCACGTTTTCGACtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGATACATGTACCGCTCATCATCATA
B1_Mc_g30984_20_Dla0	g29271	ATGGACCAGGTCCGAATTTACTCAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaTTGAGGCCGCTTAGGTTTGTCTCCC
B1_Mc_g30984_20_Dla0	g29271	CTGGCGATCAGAATCTTCATTAATCtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCTGCTGCGCTTCTTTTCTTCTTCAA
B1_Mc_g30984_20_Dla0	g29271	TTGATGTCCTTGTTCTCTTTTAGGTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGACTGATGAGAGTGTGAAGGTCTTT
B1_Mc_g30984_20_Dla0	g29271	GGGGAGCATTTGGTTCCGAATCGCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaAGGCTTAGAGGCCATCATTGCAGCC
B1_Mc_g30984_20_Dla0	g29271	AGGCTCAGACTCCTGAGTTTTTGTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGCCGCGGCGTGGCTGTCTGCGACCT
B1_Mc_g30984_20_Dla0	g29271	TCTTTTACAGGGTCGTCCTCCATTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGGGGAACAGGTCCTGATTTTTTCAC
B1_Mc_g30984_20_Dla0	g29271	CCTGCTTTTCCATTAATTTAGAATTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCATGACGGTTTCATCTGCCCGCTTC
B1_Mc_g30984_20_Dla0	g29271	TTTGTGAGAATTCTGTTGATTCCCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGTGGAAAATGCTGCTAAGGCGACAA
B1_Mc_g30984_20_Dla0	g29271	CGGTTCACTGAATCGGCAGGCAGCGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGATACATGTACCGCTCATCATCATA
B1_Mc_g30984_20_Dla0	g29271	ATGGACCAGGTCCGAATTTACTCAGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaTTGAGGCCGCTTAGGTTTGTCTCCC
B1_Mc_g30984_20_Dla0	g29271	CTGGCGATCAGAATCTTCATTAATCtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCTGCTGCGCTTCTTTTCTTCTTCAA
B1_Mc_g30984_20_Dla0	g29271	TTGATGTCCTTGTTCTCTTTTAGGTtaGAAGAGTCTTCCTTTACG

B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGACTGATGAGAGTGTGAAGGTCTTT
B1_Mc_g30984_20_Dla0	g29271	GGGGAGCATTTGGTTCCGAATCGCCtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaAGGCTTAGAGGCCATCATTGCAGCC
B1_Mc_g30984_20_Dla0	g29271	AGGCTCAGACTCCTGAGTTTTTGTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGCCGCGGCGTGGCTGTCTGCGACCT
B1_Mc_g30984_20_Dla0	g29271	TCTTTTACAGGGTCGTCCTCCATTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGGGGAACAGGTCCTGATTTTTTCAC
B1_Mc_g30984_20_Dla0	g29271	CCTGCTTTTCCATTAATTTAGAATTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCATGACGGTTTCATCTGCCCGCTTC
B1_Mc_g30984_20_Dla0	g29271	TTTGTGAGAATTCTGTTGATTCCCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGTGGAAAATGCTGCTAAGGCGACAA
B1_Mc_g30984_20_Dla0	g29271	CGGTTCACTGAATCGGCAGGCAGCGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGCCGCGGCGTGGCTGTCTGCGACCT
B1_Mc_g30984_20_Dla0	g29271	TCTTTTACAGGGTCGTCCTCCATTGtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGGGGAACAGGTCCTGATTTTTTCAC
B1_Mc_g30984_20_Dla0	g29271	CCTGCTTTTCCATTAATTTAGAATTtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaCATGACGGTTTCATCTGCCCGCTTC
B1_Mc_g30984_20_Dla0	g29271	TTTGTGAGAATTCTGTTGATTCCCAtaGAAGAGTCTTCCTTTACG
B1_Mc_g30984_20_Dla0	g29271	GAGGAGGGCAGCAAACGGaaGTGGAAAATGCTGCTAAGGCGACAA
B1_Mc_g30984_20_Dla0	g29271	CGGTTCACTGAATCGGCAGGCAGCGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaTGGCTCTTGTTTCGTGAAGGAAGTT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TCACAGACTATGATAGTTGACTTCAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGACGGACTATTTGAGGTCAGCGTAG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	AAAGGAAGTGAGCAATAAGGATTGTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaTGACGTAATAACTTGGAGCTGCCGG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GTTGACCACTCATATCGTACTGTTGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaAGGCTGTGTTGCATATCCTGAGTAA
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GTAACTTGGATACATCTGCCCACTTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaCCCAGCCCTGTTGCAGGACTCCAGA
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAAGGCATCATGTAGCTTCCGCTGAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGGGGATTGACACTACTGACCATTGC
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	CTGCTGCAGCACCACCTGAGGAGACtaGAAGAGTCTTCCTTTACG

B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGGTCACTAGCTGGGAGGCAGCGGTT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TGCAGGAACAACGACTTGCTGTGCGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaTCGGCGCTATTGATTTGAACTGGAG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GTCATGACTCCGGAGTTGTCAATGGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGAGTGGGATGACCTTCCGATGCGGT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TTTGACCGATAACGCCAGTGTCGTTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaACTGGACGGACCGTTATTGACATTC
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GCTGACGGAACTTTGAACAGAAGGAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaTTAGCAGCGCGATCCATTGATCGTT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TGATCACACCTGCCTTCATTGCGCAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGCACTCTCGACTCTGGCAAACTTAT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TAGCTCGGCGATTCTTAAACCAAACtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGTCTGGATATTGAGTTTTCGTGAAA
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	CCCCGCCACCTGTTCCCTCATAAAGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGTAAATGTCGTTCTTTCGCGTCTTT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TCTTCCAAAATATCTAATTGTGTTCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaTGACAGCAGTGGCGACGGGGTGGTA
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TCCTTGGCATGCCAGCCATAGGAGCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaCATGGTGCATGGAGGAACCATGTTC
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GGCGGTCGTTGGAGGTAGGCTTGACtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaCCCAAAGCCGAAATTCCAGGTAGCG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	CCACATGGCAAGACGGTCGTTGAAGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaGATAAGCTCCATACATGCTTGAGTT
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TATTACCGGCATGAGTGCTGTAGGTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaAGTACGATGCGGCGTAATTCTGTTC
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GTAGGGATCCATCAACAGAGAAGACtaGAAGAGTCTTCCTTTACG
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	GAGGAGGGCAGCAAACGGaaCTTGAAATATGACCAGAGCGCGCAC
B1_Pc_ng1930_Otx_1_20_Dla0	g1930	TGAGAATGGCTCGATGTCAATGCTTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTACTATGACACAGATGATGAACATA
B1_G26528_24_Dla0	g10760	TTAGAAGACAATGAATAACACAATGtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaCATCCCCATGTGTTGAGTTGCACAT

B1_G26528_24_Dla0	g10760	CGATCAATCAATATCACCACTTCATtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTGGTTGTTGACTGACTTTCATCGTT
B1_G26528_24_Dla0	g10760	CTTGATGTTCTCTGTCAACATGTTCtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaATTATCTTCCTCCACTAATCTTTGC
B1_G26528_24_Dla0	g10760	GCCGACAATATCAGTAAAATTGTCAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaATATCAGAGATCAATCGCAAAATGA
B1_G26528_24_Dla0	g10760	TTTTCAATATCTTTATACCTCGAGAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaAGCTTTCTCCTGCACAACTGAATGG
B1_G26528_24_Dla0	g10760	TTGAAATTGTATATAGACTTCCACAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTATCAGACTCCAAAACAAAACAGCA
B1_G26528_24_Dla0	g10760	TGATGACCTCCACTGCAATCCCCAAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTTTAGTGTGAGCGGCATTGGAGTTG
B1_G26528_24_Dla0	g10760	TGCAGCAAAAGAAAAGCTGTGTTCAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaACGCTATCAAAAGTCCGAACATCCC
B1_G26528_24_Dla0	g10760	ATTCTGATGTGAGAACAACGCTTTCtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaCATCAATAATTGAGTAGGTATCATT
B1_G26528_24_Dla0	g10760	GAATATGCGATATCCTCTATCCGTAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaGGAGTTATATCTTGAAGTTGACTTC
B1_G26528_24_Dla0	g10760	GTCAAAGCTACGGCTATGATCAAATtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaGACTGGCTCTTCCATTGATTGTTTT
B1_G26528_24_Dla0	g10760	TCATGAGGAATTCCAAGCGAATGAGtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaCATGACATTGTGTGCTCCGTAAAAC
B1_G26528_24_Dla0	g10760	AGAGATGTTGGAGGATTTCCACTGTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaGAAAAATCATATTTTTTTGCGGAGA
B1_G26528_24_Dla0	g10760	TTTTCATCCTTGCATGGCTTCAATTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaACATTGATTGGGTTGAAAACTTTAA
B1_G26528_24_Dla0	g10760	TTAGGTAAACTCTGCTTACGCATTTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaATATGAGGAAGAAAGTCCAGTAGGA
B1_G26528_24_Dla0	g10760	TAATCGGCATTTATATCTGGCGTTTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTTCACGAACATGATTTTGTCATATC
B1_G26528_24_Dla0	g10760	CATTGAGTACTAGTACGGTACACATtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaGATCCTCTATACCTGACTTTGGTAT

B1_G26528_24_Dla0	g10760	CTATTACCAAAGTGCTCGCAGGATCtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTGCTATTACACGAGGGGTAGTATTC
B1_G26528_24_Dla0	g10760	AAGTTTTGCATCTTTTATCCTGTCAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaATGCCGATGTAGATTTCCAGGGAAA
B1_G26528_24_Dla0	g10760	CTTCCGATTTCCCACTCCTCCCAAGtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaCTCTTGTGATCAGTGGAGTTATGTA
B1_G26528_24_Dla0	g10760	TGCCTGAACTCTCAGAGAACCCCGTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaCGCATGTGATTCAAAGCTCCTATAT
B1_G26528_24_Dla0	g10760	AGTTCTGTTTTTGTCTTTCGATTTAtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaTCCATCTGATAGTATGGAAGACTTT
B1_G26528_24_Dla0	g10760	TTGAAGATGTTCGTGAGTCTCTCTTtaGAAGAGTCTTCCTTTACG
B1_G26528_24_Dla0	g10760	GAGGAGGGCAGCAAACGGaaGAATAAAAATAGAATCTAAAATTAA
B1_G26528_24_Dla0	g10760	CAGCGGCCTTGTCAGCACCCGCATGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAATTGTGGCACCACTCGATAGTTCA
B1_Pc_g44657_28_Dla0	g33590	TCATTTTTTATCTTCGCCTTCATCGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTTTGATTTACTGTCATCCGCATCTT
B1_Pc_g44657_28_Dla0	g33590	ATCTTTTCAATGAGTTCCAACAATTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCAGAGCATCTAGTATTCGGTGAGAT
B1_Pc_g44657_28_Dla0	g33590	TTTTCTTTGTAGTTGATGAGCCCTTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTGACTCTTTCAGATTGCCAAAATCG
B1_Pc_g44657_28_Dla0	g33590	TTTCTGTCCAATGTGTGCCACAACTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCTGTCCACTAACATGTCTTCGAGCG
B1_Pc_g44657_28_Dla0	g33590	TTGTCATATTTATATCCTAAAGCTCtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaACTGCCTACAAGCACCAGGCTTAGA
B1_Pc_g44657_28_Dla0	g33590	CAGTTGGCATGTAGGAGACAAAAGTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTGGCAACATGCGGTAGTACCAGCGT
B1_Pc_g44657_28_Dla0	g33590	CCAGAAAGCAGGAAGTTCGTGATTTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAAAATATAAAACGGATCGTTAGGTG
B1_Pc_g44657_28_Dla0	g33590	ATCAGCCGATCTATGTTAGCGTAATtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCCATACCCTTTATCTGTCCGCTCAT
B1_Pc_g44657_28_Dla0	g33590	TATTGGAATCATGGAATGTTCCGCGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCTTGCCACTCGTAAGCCAATATGAC

B1_Pc_g44657_28_Dla0	g33590	TTCTCGGGTCTTATTGCAAAGACATtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAAAGGAACTCCAAGGACCCGTCGGA
B1_Pc_g44657_28_Dla0	g33590	CCCTGACATTTATTCTCCGAGTAGAtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTCCAACGACGCGTCAAGATGTCAAG
B1_Pc_g44657_28_Dla0	g33590	CTTTGGTTGTCGGAAATTTATTGTTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAGTTCTGTATCGCATCGAGCAATGA
B1_Pc_g44657_28_Dla0	g33590	CAACTCTTTCCCGCAGTTGGCACAAtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTGCAAAAAGTAAGCTCCAGACTCAA
B1_Pc_g44657_28_Dla0	g33590	CCCCAACTGGCATATGGATTATCATtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAATCGTTGGTGCAGATGTCACACTT
B1_Pc_g44657_28_Dla0	g33590	TGAGATACTGTGTCTCTTTACCAACtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAGTGAAGTTTGCGATATTCAATTTT
B1_Pc_g44657_28_Dla0	g33590	CGCATCAGTCCAATCCCAATATGGAtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCAATACAACAAGAACATGCGGTGGA
B1_Pc_g44657_28_Dla0	g33590	AGAGCCACTTTTCGTAGTCTTTCTTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAATTGATCCCGTTCTTGCCACTGTA
B1_Pc_g44657_28_Dla0	g33590	AGGGGTTAATTGGTTGACCATTTGCtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAATGGAAAATCTGGCGAACCAATTC
B1_Pc_g44657_28_Dla0	g33590	CTTGCACTTGGCTGCATTATTGGTGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTTGCGCCAAGGGGCACCCATAGACC
B1_Pc_g44657_28_Dla0	g33590	AACATCATTTCTTCATGGTGAGACAtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaGGGGTGATTTATTCCTTTTTGACGC
B1_Pc_g44657_28_Dla0	g33590	TTGTAGTCGATACAGTAGGAACAAGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaAGTTTTGGTCATCTTTAATGCATTT
B1_Pc_g44657_28_Dla0	g33590	TAGCACTTTCCAAAAGTTAGTTTTTtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTTTGAACCAGCGTAGCCGAACCAAC
B1_Pc_g44657_28_Dla0	g33590	CGTACAAACCGAGTCCGATATTGACtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaCCACGCACTCACAAGCATATTTGAA
B1_Pc_g44657_28_Dla0	g33590	CACTACAGTCTTGACCTGAATACTGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaGTGGAGAAACCAATATGCGTAGTTT
B1_Pc_g44657_28_Dla0	g33590	GCGTGATGGCCAACTCATCCGAAAAtaGAAGAGTCTTCCTTTACG

B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaGGTCGACAGATACCACGTGAAGGAA
B1_Pc_g44657_28_Dla0	g33590	TGGACAGGCTCTCGGTACACTTGCAtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTTGGACAGCAAGTAGCATCTGGTTT
B1_Pc_g44657_28_Dla0	g33590	CACATGCTGCTTTATGACCAGGTGGtaGAAGAGTCTTCCTTTACG
B1_Pc_g44657_28_Dla0	g33590	GAGGAGGGCAGCAAACGGaaTATGGCATTTGTGATGCCAGTTAGG
B1_Pc_g44657_28_Dla0	g33590	AACATTTCCTGTGCATGAATGATGGtaGAAGAGTCTTCCTTTACG
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCGTGTATGGCGTGGTCAGTTCAATG
B2_Pc_g44721_33_Dla0	g1340	TCATGGTCCGTGACTGCCATTTAGCaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTGGTCCAGTGAGTTGTTGTCTCCCG
B2_Pc_g44721_33_Dla0	g1340	TCGATTTCATCGCCCTTGAAAGGTTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaATTTGTTGGGAATGAGCTTGCTGTG
B2_Pc_g44721_33_Dla0	g1340	TTTCCGAATACTCGTCAGGAGACTGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTCCAATAAACAGCCAAATGCCAACT
B2_Pc_g44721_33_Dla0	g1340	CATACCGTGCAGGAAATAAATAAAGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaATCAGCGCTATGTTGAAAAACACTC
B2_Pc_g44721_33_Dla0	g1340	GCCCATGTCAAAACGGGCAGATTAAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCCTCAAATCTGTCTGATTCTTTCTC
B2_Pc_g44721_33_Dla0	g1340	TCAAGGGAACGATTGGTACGCATGGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGAAGGCTCCACAGCCAATCAGACCC
B2_Pc_g44721_33_Dla0	g1340	TGATATATAAGGTTTATGCATGCATaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCAATATTCTCCTCCGACAATTAATC
B2_Pc_g44721_33_Dla0	g1340	AACATTGGTATGATAAGCGCAAGAAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCAATACTGAAACCCACATACGCGAT
B2_Pc_g44721_33_Dla0	g1340	CAGTATGTATCGTGGTCACAACAAAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCTTGTGATAAAGAAGTCTAAGACGC
B2_Pc_g44721_33_Dla0	g1340	AGCAATGAGGACATTTCGACCTGGAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTCCGAATCCTTCATTAGATATAAAT
B2_Pc_g44721_33_Dla0	g1340	AATGGTCCTCTGAGGTAGCCTACTTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCACACCCCACGAGGACATAAGAAAT
B2_Pc_g44721_33_Dla0	g1340	CTTTCACGTATCGCAAAATAATCAGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTGAGGCGTAGTTGATGAACACGGCG

B2_Pc_g44721_33_Dla0	g1340	CGTGCCGATACTAAGAAAAGAAACCaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaAGAGGAATAAGAGGAACTTTGACAA
B2_Pc_g44721_33_Dla0	g1340	ACAGTGGTGATAGCTCCGAAGAAGTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGAAGTCCGTCACTGGCCATAGAATA
B2_Pc_g44721_33_Dla0	g1340	GATGCACCCAGCCCATCCACGAGTGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaAATGACATTGGTACCAGCTGCGACT
B2_Pc_g44721_33_Dla0	g1340	AATTCGAGGAATTGCGTATGTACAGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCAAGTGATATTTGGATGGAAAAATA
B2_Pc_g44721_33_Dla0	g1340	GCTCCTATACTGACAAGTATCCGAAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaAATACCAAGGCTTATAACCTACGAA
B2_Pc_g44721_33_Dla0	g1340	CTGCCAAGCCTTGTTGAGCCTCTACaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCACAATACTGACTGCACTCAAAATA
B2_Pc_g44721_33_Dla0	g1340	AATCGAGTGCACGACACAAACAACGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGCTTCCTCTCCAGCGACAGAAATGC
B2_Pc_g44721_33_Dla0	g1340	ACTGGAATGGTGCGTCGGGGATTCTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTCCCAGACCCTGATAATAGCCCCCC
B2_Pc_g44721_33_Dla0	g1340	CAAAACCGATAAAGGCATTAAATGCaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaTTTGTCTCCTGTCAAAAGTTCCTTA
B2_Pc_g44721_33_Dla0	g1340	ACCCATTGGAAGGAAATACCCGTGTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaATGCCAATAATAACTAAAGTTAAGA
B2_Pc_g44721_33_Dla0	g1340	AAGTGAGCATACATATATCCAGTGGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaACCTATTTCCAAAAGAAAGAATGGC
B2_Pc_g44721_33_Dla0	g1340	AAAGAATGATATTGACATTGGATGTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGAAATTTGGATATCGGTTTAACATT
B2_Pc_g44721_33_Dla0	g1340	TGCAAAAATGCAAATACCCATGGCAaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaACGTATTCTAGTATACCATTGCTGA
B2_Pc_g44721_33_Dla0	g1340	AGACTGTCTGCTGGCGCTAATGGTCaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaAGGAAGCCAACACTGCTGCGGAGCA
B2_Pc_g44721_33_Dla0	g1340	TGCCATTGGACATGATGTCTATGTTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCCATCCAACGTTGAAAGCTGCGAAC
B2_Pc_g44721_33_Dla0	g1340	TGACAAAGCATACTCCAACAGCATCaaATCATCCAGTAAACCGCC

B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGCTGACCCTGCTCGTGGAATATGGG
B2_Pc_g44721_33_Dla0	g1340	CCTGCTATGACGTATGTATATGCGTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGGAGAGCAACGAAGCCTGCCAACAT
B2_Pc_g44721_33_Dla0	g1340	ATAACTCTGAATAACACATGCCGTTaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaCTTTTTTACAACCACTGCAATAACT
B2_Pc_g44721_33_Dla0	g1340	GCAAATGAAGGCCGATGGACCACTGaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaATGCCGTACATGAGAATGTCAAACT
B2_Pc_g44721_33_Dla0	g1340	TAGATAGAATTACCCAACATCGCTCaaATCATCCAGTAAACCGCC
B2_Pc_g44721_33_Dla0	g1340	CCTCGTAAATCCTCATCAaaGTTGACCGGCATGATCAATAGGCTT
B2_Pc_g44721_33_Dla0	g1340	TCAAATTCCGGTTGAGTGGTGATTCaaATCATCCAGTAAACCGCC
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaATTATGTGATGGTTGATGAAGAGGT
B1_PC_g38703_33_Dla0	g31956	TTATGATGCGTCTTGTGATGTCGTTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTATGTTACCATTGTCCCTTTGTGCT
B1_PC_g38703_33_Dla0	g31956	AAATTTGTTCGTCGACCAGTGACCCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCATGCTGCTTGTATTTTTGAAGAAT
B1_PC_g38703_33_Dla0	g31956	TGTCAACCTCTCCTCGACATTCAAGtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaATTTTCTTCACTGAAGCCCTTCATC
B1_PC_g38703_33_Dla0	g31956	AGCCGTTGAGCTGCTGATATGAAACtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaAGACAATATCGAGGCATAAGCAGTC
B1_PC_g38703_33_Dla0	g31956	GATGCTGTATTCATTGCATCTCCAAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGCGCTCCAACTTTCAGTGTCAAAGG
B1_PC_g38703_33_Dla0	g31956	TGACTGATGCCACAATGGAGCCACTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaAATGTTTTGCAACAAATGAAGACCC
B1_PC_g38703_33_Dla0	g31956	CCCAGATTTCGGCACGGAGACCGTAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCCATTCGCTTCTGGCAATCCCGACG
B1_PC_g38703_33_Dla0	g31956	TTTGCAATTTCCAATGCATGACGATtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCTTTGTAAACGTCGTAGATTTCTAA
B1_PC_g38703_33_Dla0	g31956	TTACATAACTGTCGGCTATCGTTTCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCAGCAGTTCAATGATTTGCAGCGGT
B1_PC_g38703_33_Dla0	g31956	GTCATCGACAACATTATATACATTAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGGAATTTCGGTAAATAAAATTGTGA

B1_PC_g38703_33_Dla0	g31956	CTTTCACTAACAATTGCCTCAAAAGtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTACCTGCCTTCAAGTAAACAGCTAC
B1_PC_g38703_33_Dla0	g31956	TATCGTAAGACGTTGGAGGTACAGCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaATCTGCCCGAGCCGTCTCCTCCTGA
B1_PC_g38703_33_Dla0	g31956	CTTGGGTATCATGCGGAACAGAAGCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGCCAAATGACTGACATTTTCAAAAG
B1_PC_g38703_33_Dla0	g31956	GCTCGTCTCTTTTCAACGATTGTGTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTGCTGCCACCAACGAAGAACGGGAA
B1_PC_g38703_33_Dla0	g31956	TTGAAATGATTCTCATGTACCATGCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCCATCGGTATTCCCGATCAGCAGCA
B1_PC_g38703_33_Dla0	g31956	GGCAAGGAGAACGCTAAAGGCTACAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaAGACTGAGCTGATACGCAGAATAAC
B1_PC_g38703_33_Dla0	g31956	ACAACAAATTTCTGAGCTTCAACTAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaATTCGTGCATTTCATTGGTGCGATT
B1_PC_g38703_33_Dla0	g31956	TGCATCTTTCCATTTGTAGCAAACTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGTGGACGTGCAAAGCTACGTCTAAG
B1_PC_g38703_33_Dla0	g31956	GGCACTCTTGTTAACACAATGAGCTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTACATTCTCTCCTCCTTTTTCAGAA
B1_PC_g38703_33_Dla0	g31956	CCTCTTAAGCGTGAAATAGGGTAAGtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaAATACTTCACCAGTTTTTCCTCTCG
B1_PC_g38703_33_Dla0	g31956	TTCGGACGTTGGTGTAAATGCTCTTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGGTGACCGTAAGGTCACAACGAGCC
B1_PC_g38703_33_Dla0	g31956	TCTGGCTTTATTCAAAAACAGCTCCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCTCAGACTGTCTACCAAGTAAAACA
B1_PC_g38703_33_Dla0	g31956	AAATTACTAACAGACACTCGCTCAAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCATCGACGGATAGACGAAGATCTCG
B1_PC_g38703_33_Dla0	g31956	TGATGACAGCATCTATCCGGCCCCAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCAAAGAGGTATAAAGATCAAGTTTG
B1_PC_g38703_33_Dla0	g31956	TCGATGGGTCTCTACAAGTCGTTGGtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCGAAGGGCATCTATCGCGTGAATGA
B1_PC_g38703_33_Dla0	g31956	ATCAATTGTGTTGATCTGGTGTTTGtaGAAGAGTCTTCCTTTACG

B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGGTTGAAACGGTCTGCAACAGCATG
B1_PC_g38703_33_Dla0	g31956	ACACGTTGTTAGCTTTCATGAACACtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTGAAAGTCTTGCGTATGCCAGCCGA
B1_PC_g38703_33_Dla0	g31956	GGACTTATTGGTAACCTTTATCTGCtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaTCTCTTTCACGAAGTAAGTCAAATA
B1_PC_g38703_33_Dla0	g31956	TTTGGAGTCTCATAAGCCATGGCATtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGGTATGATGCATGATCCAATTTTTC
B1_PC_g38703_33_Dla0	g31956	ATTTGAATAATGATTGGACACTTGTtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaGAAGATTAATACGCAGGCAAATACT
B1_PC_g38703_33_Dla0	g31956	ATAATTAACCGTATGCTGTGATATAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCGCAGCATATGCACTTGTTTGTTCC
B1_PC_g38703_33_Dla0	g31956	TTCGGCAAGTTGACGATCCCAATCAtaGAAGAGTCTTCCTTTACG
B1_PC_g38703_33_Dla0	g31956	GAGGAGGGCAGCAAACGGaaCCAGTGCTTCATTGCCTTTACTGCT
B1_PC_g38703_33_Dla0	g31956	TTTGATCCTGCTGGGTTTTGATGAAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAATTTCATCACTCGACGGGAATTTG
B1_Pc_ng4824_TRH_30_Dla0	g4824	CTATAATACTTTACAAAAATTCGATtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaACACGTTTATTGTGGAAAGCATCGG
B1_Pc_ng4824_TRH_30_Dla0	g4824	TTCGCGTCAATTCTAATTAGGCAAGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTATTTCGCAAAGAATAAGGAGTCGC
B1_Pc_ng4824_TRH_30_Dla0	g4824	CAAGACCATTCTCAGGAAAAGTTTGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTGCAATGAATTTCTTCTTTCTTTTC
B1_Pc_ng4824_TRH_30_Dla0	g4824	CTTTCCGAGAAATTCTCCAAATCGTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGAACGATCTGGTGCCATTCGAGAGT
B1_Pc_ng4824_TRH_30_Dla0	g4824	AGGAATTCACCGCCGCTATAGTACAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTTTTCCAAGAAATTCTCCAGTTCT
B1_Pc_ng4824_TRH_30_Dla0	g4824	TGGAAGATATCAATGATGATTGATCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGGGACTTTTTGCTCCAAATTTGCGA
B1_Pc_ng4824_TRH_30_Dla0	g4824	CCCCAAAAATTCTCCGTATCTCTTTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaCGCTTTCCAATAAATTCAGAAACCC
B1_Pc_ng4824_TRH_30_Dla0	g4824	TCTTCTATCAGCTGAGATAAAAGCGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAATGCAAAAATCGATGAACACCTAT

B1_Pc_ng4824_TRH_30_Dla0	g4824	TTTCGATGTCATCAACAGGTTGATCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTGTTTGCGTCCCTGTTTCAACCAC
B1_Pc_ng4824_TRH_30_Dla0	g4824	ATCTTTTGTCAGTTGCCGTATATCTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAATGCCGTAATATTCCCAGAACACT
B1_Pc_ng4824_TRH_30_Dla0	g4824	ATGCATCCACGCTGCTGAAAAATGGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaCCAATAGAGGGGAAAAGTCTTTCGG
B1_Pc_ng4824_TRH_30_Dla0	g4824	CGTTGAAATTTTTGAAATGCGTTCTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTTATTTTGCCGTTGATACTCCATC
B1_Pc_ng4824_TRH_30_Dla0	g4824	TGCTAGAGAGACACGATGGCTCCCTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTCCACAACACATTCTTTATTTCTA
B1_Pc_ng4824_TRH_30_Dla0	g4824	CGACCATGGACATAGAATGGACCCGtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTGAAAAATGGATGCTTTCTCGGAGT
B1_Pc_ng4824_TRH_30_Dla0	g4824	TGTCTTTGAAGTGGTCCAGCAAATAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAAGATTCCATAGTAACCTTGTTTGC
B1_Pc_ng4824_TRH_30_Dla0	g4824	GATTACAGAATTCATGGTCAAATAAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGAGTGAAGTCCGTTTATATAATTTT
B1_Pc_ng4824_TRH_30_Dla0	g4824	CTTCTCTCTTGATCAAAGACAGCCCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaCCATATATATTCCACCGTAGTGGTA
B1_Pc_ng4824_TRH_30_Dla0	g4824	AAGGTTTACGACATTCCACATCCATtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaCTGTATTCCTTTGTATTGTTGCAAA
B1_Pc_ng4824_TRH_30_Dla0	g4824	AATTAAAATCCTGAAAAAGTCTGCTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAGCAGTTCTTCTGATTTTTTATCTG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GACTTGTATTCAGGATAGTGATGATtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTCCGTCTGCATTTATCAAGCAAATG
B1_Pc_ng4824_TRH_30_Dla0	g4824	AGATGACATGTATAAAATCAAAGTTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGTAAATTTGATGGATAATATGCGGG
B1_Pc_ng4824_TRH_30_Dla0	g4824	ATACTGCACTGGAATAATAGCTGACtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaAAGAATTTTTCTGTTGTTCTCCGGC
B1_Pc_ng4824_TRH_30_Dla0	g4824	CCATATTTTGTCTTCTTGGCATTCAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTCGAGGGAGGTTTTTGCACATTTTT
B1_Pc_ng4824_TRH_30_Dla0	g4824	CTACTTGCCGACGCATTTTTGCGTAtaGAAGAGTCTTCCTTTACG

B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaCTTTGCATTCTCTTCTACAATTGTG
B1_Pc_ng4824_TRH_30_Dla0	g4824	CTTTTTCTTGATACTATTGTGTTTAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGGCGTTCCATTCGACGATGGATATG
B1_Pc_ng4824_TRH_30_Dla0	g4824	ATTTTTCTATGATTAGTGGAAACAAtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTTTTGCACTGTTCGCTATGCTAGA
B1_Pc_ng4824_TRH_30_Dla0	g4824	CCCCTTTCACCGTCGTAGAGATTTCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaGATCCTTTGAAATCGACTTTTATCC
B1_Pc_ng4824_TRH_30_Dla0	g4824	AATCCTTTGATTCGACGGCTTTGTCtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaACGTCATTATTGATTTTTGTCTCAT
B1_Pc_ng4824_TRH_30_Dla0	g4824	TATATCAAAGTATATATCCATTTCTtaGAAGAGTCTTCCTTTACG
B1_Pc_ng4824_TRH_30_Dla0	g4824	GAGGAGGGCAGCAAACGGaaTTGAGATCACCACTGCCAGCGTGAA
B1_Pc_ng4824_TRH_30_Dla0	g4824	GGAATTGTATAGACAAGTAGAATGCtaGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAACATTAATTCTCGATCAGCGGTTTCT
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TCATGCAACTTTCGGAATAACTGACTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAATGACCCAAGCGCCAGCTCCGCGCA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	ACACTGGTTCCATTTCCAGCGCCATTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAATTCTTAGACTCCAAGGATCATCATA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TTCTATAAACCATGTAAGAGTATTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAGATGAACGTGACCAGTGGTGTCAAG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TATTTGGGTGTATTTATTATTTTTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAGACATATTTCTCAAAATTGGCATGA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TGTGTGTTCCAATTGCGACCCTCAGTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAACTCCAACATTAGAGCATTTTGTCAA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	CCTTATAGCGCCTCTGATGTGATCCTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAGCAAACATCTTCGTAACCGTACCAA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	AGTTTGGTTTGAATGGATAAACACCTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAACACCCAAGGATAAAAATCTCTCTG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TCAGCTAGCACCAACTTCGAAGCAGTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAATGGACAGCAATTTACACAGGTTTGT
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	CTTTGTAGATTGCAGGGAAAGGTGATAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAACGTGCCTTGCATGTCATAATCCATA

B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	ACCAACGCGTTCGTACATCACTGATTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAATTCATGTAGTCGTAGGTTCGTCCAT
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	ATAAAGAAATCAATAACTTTCATTATAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAAACTGAGGGAGTAATTTCTATTGAC
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TACAGTCTGCGCAGTTAACATAGTCTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAAACAGCCGTAAAAGCGTGATAAGCA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGCAATGCGACTCTCAACTTCCCCTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAATCCATGTTGAGGCCATCTAAATTGT
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	CTATTTACTGCAACACTGCTCTCGATAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAATGTAGGCATCTGTTTTCGTTGGTGA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TCACTACGGTAGCAACGGACCGCTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAACAAGGTAACTCTTACGTTACGAGCT
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	AATATCCGCAACAGAAATATCTGTCTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAGCGGTGATGCCCGTCACTATTAGAG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GCATAACAGATTAGTTGTGGTAGATTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAAATATGATGGTAATTGTGAAGTTGA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	TAAGCTTTGTCCAGTCGAATTTTGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAAAGGACGAGTTTGACAGAGTGTCTGG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	CCCATAGAATCCAAATGATTGTTTATAGAAGAGTCTTCCTTTACG
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGGAGGGCAGCAAACGGAACCAAGGATTGCACATGCGATTAGAA
B1_Pc_g38095_cluster_15_marker_2_20_Dla0	g38095	GAGCAGGGACAATTGGAGGATCTAATAGAAGAGTCTTCCTTTACG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCATCACCATAATGGATGCGAATAGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	TCATGCCTTCATTGCAGCATTCTTCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTTAGCTCCACCAGTTGTAGAAGAGC
B3_g24624_cluster_14_marker_70_Dla0	g24624	ATTGCAATATTTTCAGATCTCACTGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCAGTTTGAGTAGTGCTCAGTGCCGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	AACCGCCGCTGACACTGTTTACTGCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTAGCACTTGACATGGCTTGCAGTTGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	TTGTGCTGATGTCATCATGTTCACCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCTGATCAGAGGAATGGAAGTAGATG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCGAAAATTGCTTGAAATTGTCCGTTCCACTCAACTTTAACCCG

B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTTACGAGATCAGTTCCAGGGACTCC
B3_g24624_cluster_14_marker_70_Dla0	g24624	TAGCCGCAACTTGTGCTGTCGTAAATTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTGCTGTGCTCCAACCGCTCGACAGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	GACGACGTTTCGCATTTCTGAAATCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTACTTGAGCCATTTGGAGAGTTGAGC
B3_g24624_cluster_14_marker_70_Dla0	g24624	CCCATAGCAGCCTTCACTTTCGTCCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCTCGACTCACTGAAGTGCTGCTGAG
B3_g24624_cluster_14_marker_70_Dla0	g24624	TTGTGAGAGTTCCAATAGTTGGCAATTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGAAAAACTCTTTCATGCCCATCAGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	AGACATCTGCGTTGTCGTTAGACCATTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGTGGCCGGAGAAGTGAGATATCTGT
B3_g24624_cluster_14_marker_70_Dla0	g24624	ACGTCCGATAATGTGGACAGAGTTATTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTATAATGTGCTTGTTCCCATGTTTGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	CTCCAGCCTGCAGCTGACTCTCTGTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTAATTTCAGAACTCGTCAAAGCAGCA
B3_g24624_cluster_14_marker_70_Dla0	g24624	GGCATCCAATGAGATGTTAAGTGTTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCTTGTCCATGAGGAGACTGTTGAAC
B3_g24624_cluster_14_marker_70_Dla0	g24624	ATGACTCCAATCTGACGAACAGTTTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTCGATACCTTTGAGGAATTCCAACC
B3_g24624_cluster_14_marker_70_Dla0	g24624	AAACTCCTTTGACTTTTGTCATGAGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTGTGAATTCAGAAGTTGTCATCCCA
B3_g24624_cluster_14_marker_70_Dla0	g24624	TTGTGAGCCAAAGAACTCGGTGCAGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTACACGAGTTCCCATCGCCTTAGCGT
B3_g24624_cluster_14_marker_70_Dla0	g24624	ATATCCGTATCTGTCAGGGTACTAGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTCCTACGTCTTGCTTGATAAGTTTG
B3_g24624_cluster_14_marker_70_Dla0	g24624	CACAGGTTACCGCCGTTGTTGATCGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGTCTTGTCCTGATTTTTCATTTGTT
B3_g24624_cluster_14_marker_70_Dla0	g24624	GGCAGAAAAACCTCTACTCAGAATTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCGCTCTTCTCTTCTTTTCATCATTG
B3_g24624_cluster_14_marker_70_Dla0	g24624	TCTGTGGAGTTCATGTCCGACATACTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCCATCGTTGCAACAGCCAATTCGTT
B3_g24624_cluster_14_marker_70_Dla0	g24624	CACGGGTATCAAAGTTAGTTGCCATTTCCACTCAACTTTAACCCG

B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTAAGATGCTGGGCAAAAGGTGCCAAA
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCAATCTGATTCAGTTCTGTATCTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTCAGAAATAGTTTTAATGCTTGTGA
B3_g24624_cluster_14_marker_70_Dla0	g24624	GAGATGACTTCACTGCGATTGTACCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGACAGCTGTCCCGGATAGTTGGCCC
B3_g24624_cluster_14_marker_70_Dla0	g24624	AAGCCTTCGATTCATCCGATGACATTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTCCCAATTCTGTTTGTCCAAGACCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	AGTAAGGATCGCGTTCTCACTCAGTTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTATATCGTCTAAGGATGAGACACTTG
B3_g24624_cluster_14_marker_70_Dla0	g24624	ACAAAATTTCCCATCTGACTCAGATTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGGTTTGTGTTTTTCAATGATTTCTG
B3_g24624_cluster_14_marker_70_Dla0	g24624	TATCGACTGGTGCAGAAGAACTGTCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGTCGTCTGTGCTTCCTTTGCCAGCT
B3_g24624_cluster_14_marker_70_Dla0	g24624	CCATTCCTTCATTTGTTGCCGAGTCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTGCATGTTGGTAAAATCCGTGCTTG
B3_g24624_cluster_14_marker_70_Dla0	g24624	TGGATAAGACTGGCCATTGTAGAAGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGAGTGACATCCGTTCCCGTCAGAGT
B3_g24624_cluster_14_marker_70_Dla0	g24624	TGCTGTGTCGAAGAAATTGAGCACATTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTGCATTTGTTGCCGATTGCTTTAACT
B3_g24624_cluster_14_marker_70_Dla0	g24624	CATATTTCCCAGGCTAAGATATTGCTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTCATTTAGCCACTTGTTCTGCGATTC
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTATCTGACATCTCTCCTTTCACTGTTCCACTCAACTTTAACCCG
B3_g24624_cluster_14_marker_70_Dla0	g24624	GTCCCTGCCTCTATATCTTTTATTGGTGCTGGAGCAGAAAGCCAT
B3_g24624_cluster_14_marker_70_Dla0	g24624	TAATGCTGTCAGTTGATAAATTTTGTTCCACTCAACTTTAACCCG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAATAGAAGTGAAAAGCACTAGAAGTC
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TTAAAAACACAATTTGAAGCTATTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACTTTCACCATCAAATGTACTGAGCC
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TACACAATATCTGACGATATTATAATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATGGCAAGAGATTTATATCCAAATTC
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TGCCGAAATCAACCGACGTGTCCATTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATGTGAAGGTCATAACAGCTTTGTAT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	ACATCTCTTTTCAATTGCGGCATTCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAGTTTGCATCTTCGTTTCAATACCTT

B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TTCTTCTTGTAAGTTTCCAGAATCGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACTAATTGCTCTACTGTGTGCCCTTT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TGCGGTGGGCAATGAGAGTGCCACATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACTTGGCATACTGCTCATACAAATTT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	AAAGGATTTCAACTCAAATCCTTGATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATCCTTTAAGTAATTCCTGTATTCCT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TGAAACTTGATGTATTCGCCTTGATTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACGACCCCACCTGCTGCCACGCCGGA
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GGTATTCTGATACGAAGTCCTTTGCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAACCTATCACCATGTCTGATATTGCT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGAACTGCCAAAGCCGACAATTCCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAATTTCCTCTATATCAGAGCTTGCAG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GCAAATTTCACTGCCTCCCTCTTCATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAAATCAACCGACGTGTCCATTTTGGC
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TCATCAAATGTACTGAGCCTGTGCCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACTTTTCAGTTGCGGCATTGGATGTG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	AGATTTATATCCAAATTCCTCACATTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATTGTAAGTTTCCAGAATCGAAGTTT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GTCATAACAGCTTTGTATCCCTTCTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAGGGCAATAAGAGTGCCACAAACTAA
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TCTTCGTTTCAATACCTTGGTTGCGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATTTCAGCTCAAATCCTTGATGCTTG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	CTCTACTGTGTGCCCTTTAGGAAAGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATTGATGTATTCGCCTTGATGCTCCT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	TACTGCTCATACAAATTTTGTTGAATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAACTGATACGAAGTCCTTTGCAACGAC
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	AGTAATTCCTGTATTCCTCAGGGTATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATGCCAAAGCCGACAATTCCTCACTT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	ACCTGCTGCCACGCCGGAGTCGAGATAGAAGAGTCTTCCTTTACG
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAATTCACTGCCTCCCTCTTCATCATTT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	ACCAGGTCTGATATTGCTTCCGCAATAGAAGAGTCTTCCTTTACG

B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	GAGGAGGGCAGCAAACGGAAAGTCCGCTTCTGCCATGATAGCAAT
B1_Pc_g14423_cluster_25_marker_1_21_Dla0	g14423	CTATATCAGAGCTTGCAGCCTTCTCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATAGTGATTCTGGCCATTTGAACGAT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CTGTAAACATTTCATTTGTTCTGTCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATCTGGTGTATTCTTTATTGGCTCAT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	AAAAATTCATTTTGTCCTAAATGTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGACCACTATTGAGCTGGATCTGTTC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TATATTTATCGGCCTCTCCGATAGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAATATTTATCCTTAGCATGTATTAGA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CGAAAAATTCGTTACTTTAAGGGGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATCTATACTAGGTTCGAATATAACCT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GCAGCTACAATAACATGCGGATGGATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTATTATCATTTCCGACCTTGGGGG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CCAGTGAACTGCCTTCGATTTCACCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAACAACTCATCAGGGGTTCTAGTGTA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TACAACGTTTTTCGCTACTCGTACTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAATTGTAATCGGACCGAGTTTTCTAA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	AATTCTAGCTTTTCATTTACGTAGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGTGCATTTTTCAAAATCAGTGCGCT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TGTCTGTTCCTAGAATAACAGGGAATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTGTATTTCGGCTTTACCCAATATA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TCTTATGGCTACATCCCCTATTTGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTAATTGTAGTTGGGTCCTGACCTA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	AGCTTACTACCATTCACCCCGCGTGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTAGTGTTATTGAAGATCCTGTATC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GTAAGTCGCTATGTAGTTTATCACTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTTAACTTGAAGTATTGAACTTTTC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TAATGCCTGTACAGTTTTACCATTATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTACTATTATCATGATGCGAGTGAT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TTAGAACATGCGGCTAAGACGGTCTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGTTTCTTTTGGAGATGTGCTGTTTT

B1_Pc_g14931_cluster_32_36_Dla0	g14931	CTGGCCAAGTATCGCTTGTATGGCATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTCATACGGTGGGGGACTGCTTCCT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	AATAATAAATTCGACGAGTTCGTGGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGGCTCTGCTTGCTCTTCTGCCATCG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	ATAGAAATTCCGCACGGCGGAGCTGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTCTCATGCGATGCCATGGTAATTT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CTGCTACGTAGTCATTATTTCTAAGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTTTCGCAATAATTTGAACAATAAC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GCATAGGTGAAATACTCTTTTAATTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATCACTGACCGCCGCGGTTGCTTCCG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GATGGACTGCTGGTTGGTAGTATGGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATGGATACCGAATTTGATTCGCTCTC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TGCAGATCACCTTGTCTTTCTCCGTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTCGTCCGTTCCATGATTCTCGATT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CCGATCGTCTTGTTTGATTTTCGATTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGCGTCATCATTTGGATTATGCAAAA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TGTAAATAATTCGGTGTGAGTTCTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACAGCATGGTCATTCATTTCGTAATT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CATTTGAATATTTTTCAATGTTGATTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAAACCGACTCTTTTGTGTCATCCGCT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GTTTGACTCGCAGTAATCTATTATATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTGTTTTCATTAAATAGATTGCGCA
B1_Pc_g14931_cluster_32_36_Dla0	g14931	ACACATTCTGTTGTTTTGTGACTTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATCATCTCGGTTCCTGCATAGATAGG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CATTGTCCTCGACCACTTCATCCTGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGTCATCTGACCTTCCATTTACATTG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TGAATGATCGGCCTCACCATCGCTTTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATCTCTGTCTCTCTGCAAATAATAGT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TTTCGTAACCATTCCTCTGCTTCGCTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAACTGTTGAAGCAACGAGTACATTCAT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CTTGTAAATAGTTGTTTTCATGTGATAGAAGAGTCTTCCTTTACG

B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATTGGCGGTCCGTTACCGTATAGTTC
B1_Pc_g14931_cluster_32_36_Dla0	g14931	TCGCTGTTCCCGTAGACGCGCCTCATAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAATGTGAAAATCGACTTCCTTCTGTAG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAAGCTCCCATCAATGGTGTCATGGTAGAAGAGTCTTCCTTTACG
B1_Pc_g14931_cluster_32_36_Dla0	g14931	GAGGAGGGCAGCAAACGGAAGCCAAGAATTATTGCGATTACTTGT
B1_Pc_g14931_cluster_32_36_Dla0	g14931	CCATGTCCTCAATCGTGGATGCAGATAGAAGAGTCTTCCTTTACG
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATGAAATTTTCTTGTTGAGAAACATT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTAGTTCATCCAAGATGTTTTGCATAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAAGTTGACTAGTATCATCCATCACTA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GGTTTCTTCATGGTATAATGATTCAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACTCCTGTATTTGGGTCTTGATCCAA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CTTCGGGACACGCAGGAACCTCGTAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAATATGCAAGACCAGCGCGAACTTGA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TCCACAAACACCACGTGTGGAATCAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAAGCTGGAGTGGTTGTGCAGTGACAG
B2_Pc_g10647_cluster_19_82_Dla0	g10647	AATTTTCCTGGTACTGTTAGGACAGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAATCCTGTTTGCACGAATTTGCAGGC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TACCTCCAGCAACACGTTGACTTTTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGAGAATTGTAATGAGACAAGTTGAA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	ACCGTCTTTATTAATATTATCCTTGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGGCTTCTGCCCTTCGAAATTAGCAG
B2_Pc_g10647_cluster_19_82_Dla0	g10647	ATCAGTCGTAGTCCAATCACTGTTGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGTGCGACTAACAAGCTATTCTTTCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CTGTGAACATACCTCGACTATAATCAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATGTGAGTGCATCAAATGATCGGACT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TCTTAATTTTTTCGCGCTCGTGTGGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACAAATTCCATATTTAGTCGTAGTCA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	AAAAGATTTGAATGCATGGTAAGTTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATATTCTCTAAATCGAACATTGAAAT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TGCGCTTTACTTTCATATGTGTCGCAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACTTTTCTGGCAGCATTATGGTATGT

B2_Pc_g10647_cluster_19_82_Dla0	g10647	TCCTTTAATTACTACTGAACGCGGGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAATCATGTTTCTGTATCTGAATCCCT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GTGAGGAAATTTCTCAGGAATTGGCAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGCATTTGACTTTTTAAGGTTACCTC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTAGTTGCATTCTTCTAATATCTGCAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGAAGCAATCTGGAGAATCTGGCATT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GAGGCCATTTATGACATACCAGGCAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATCCAATACACATCGATACACTTCAG
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTCGCAGATCCACCTCGTCCTTTCAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGCAGCAGCTTGGTCCACGTTGTTAT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	AGTAACGAGCTATTTGACTCGGATTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGAAAGTGGCTATTTCCTTGTTATCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCAGTTTGAAATGCTTTTTTCGAAAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACGTCCGTAATTGGTGGTAACAGAAA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTGTCAATCAGTTTGGCCATGTACAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACACCCTTCGAGCCTCCTGCACCGTA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GATTATTCGGATTATAACGCGCTGGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAATCTATTTGAAAAGCTAGATTGCCT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CGTCCATGATATCTTTGGTGTTTTGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATCTTCGTTTTCAAGCACAATCTTGT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	ATTGCAACTTCTACGTCATAAACATAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAAACTCATTCCCATTTTCACAGTTAC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTCTGACTTTCATATTGCACATTAAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACGCTAAGTCTGGCTCTTTCAGACCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GTATAATTTGGGTGGAGCATACAGTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAACATTCGGCTTTTTTTCTGGTTTTG
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TCCTCAGGTAGATCCATTAATTTAAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATGGATGCAGTATTGACGGTAGAGAA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	AATCTTTCAATACGATATTTTCTTTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACGATGATTTAGTTTTCTCCATAACA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GAATTTGAATGCGAGTCCACGTTTAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAACTTCTGCGAGTGGGATTGCTTTCAT

B2_Pc_g10647_cluster_19_82_Dla0	g10647	TTTGTGGATGCTGGTGTACTTAGCAAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATGACAGTAAGTTCCATTCCATAATT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	TCAACATTTTCGCGGATCTTGTACGAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAAATGTTTTCTCTTCCAAAAGGTTTA
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CAAAGTCTTGCGCTTGTACGTCGTTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAATCGTACTCCATATTCAAATCCACAT
B2_Pc_g10647_cluster_19_82_Dla0	g10647	GGTTTCTTCCACGATTTCTTCTCGTAAATCATCCAGTAAACCGCC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCGTAAATCCTCATCAAAGACGATTTATATGCATTGTTGCTTC
B2_Pc_g10647_cluster_19_82_Dla0	g10647	CCTCCATGACAGTCATAGGAAGATCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATCGACCTCGTCGTCCACCTCGCCCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TCACCAATTCCTCTGAGCACTTCCAAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCCATCACGCCCCCGTGTCCTCCCA
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CTGCCACTTCTTCCCATTCCACTGTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACTCCCATTCCACTGCTTCCCATCAT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CACTGCTTCCCATCATGCCTCCGTGAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACATCATGCCTCCGTGTCCTCCCATT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TCCGTGTCCTCCCATGCCACTGCTTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCCATTCCACTGCTTCCCATCATGC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CTGCCTCCCATCATGCCTCCGTGTCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATCATGCCTCCGTGTCCTCCCATGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CGTGTCCTCCCATTCCACTGTTTCCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACATTCCACTACTTCCCATCATGCCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	GCTTCCCATCATGCCTCCGTCTCCTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCTCCCATGCCACTGCTTCCCATAC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TGTCCTCCCATGCCACTGCTTCCCAAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCATTCCACTGCTTCCCATGGCAAC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCATTCCGCTGCTTCCCATCATACCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAGCTCATGTCGCTTCTTCTATAATCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TCCTCCCATTCCACCGCTTCCAGAAAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACTTCCCATGCCTCCCATTCCACTGC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	GAGCCTACGCGACCACCCATTCCACAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCATACCTCCCATTCCGCTGCTTCC

B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCATCATACCTCCCATTCCATTGCTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAGGCAACATGTCCTCCCATTCCACCG
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CATACCTCCCATTCCACTGCTTCCCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATAATCTCCAGAGCCTACGCGACCAA
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCAGAAATGCTCATGTCGCTTCTTCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACACTGCTTCCCATCATGCCTCCCAT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TTCCACCACTTCCCATTCCTCCCATAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACATTCCACCGCTTCCAGAAATGCTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	ACTGCTTCCCATGGCAACATGTCCTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAACGCGACCACCCATTCCACCACTTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TCGCTTCTTCTATAATCTCCAGAGCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACTCCCATTCCACCGCTTCCCATCAC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TTCCTCCCATTCCGCTGCTTCCCATAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACATTCCATTGCTTCCCATCATACCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TCCCATTCCGCTGCTTCCCATGCCTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAATTCCACCGCTTCCAGAAATGCTCA
B2_g43312_cluster11_marker_1_50_Dla0	g43312	ATTCCACTGCTTCCCATCATGCCTCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACACGACCACCCATTCCACCACTTCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CGCTTCTTCTATAATCTCCAGAGCCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATCCCATTCCACTGCTTCTCATTCCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	GCCTCCCATACCACTGCTTCCCATAAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCACTGCTTCCCATACCTCCCATTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	ATGCCACTGCTTCCCATTCCTCCCAAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATTCTCATTCCTCCCATTCCATTGCT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TGCTTCCCATGCCTCCCATTCCATTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACATGGCAACATGTCCTCCCATTCCA
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CATCATGCCTCCCATTCCACTGCTTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACTATAATCTCCAGAGCCTACGCGAC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CTTCCAGAAATGCTCATGTCGCTTCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAATTCCACTGCTTCCCATACCTCCAAT
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCATTCCACCACTTCCCATGCCTCCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAACCACTTCCCATGCCTCCCAGTCCG

B2_g43312_cluster11_marker_1_50_Dla0	g43312	GCTGCTTCCCATCACACCTCCCATTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCAATGCCTCCCATGCCTCTGCTTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CTTCCCATTCCTCCCATGCCACTGCAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACACCCATTCCACTACTTCCCATGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TGCCTCCCATGCCACCACTTCCCATAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAAAATTCCGCTGCTTCCCATGCCTCTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CATTCCACTGCTTCCCATCATGCCTAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCCATTCCACTGCTTCCCATTCCTC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTTCCATTCCGCTGCTTCCCATACAAATCATCCAGTAAACCGCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	CCTCGTAAATCCTCATCAAACCATCCCTCCCATTCCACTGCTTCC
B2_g43312_cluster11_marker_1_50_Dla0	g43312	TTCCACTGCTTCCCATTCCACTGCTAAATCATCCAGTAAACCGCC

Appendix 4: HCR probes for *C. gigas*

Pool name	**Gene id**	**Sequence**
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTTCTGCCTCTTCCATCTCCTCTGCT
B3_Cg_g31376_28_Dla0	g31376	TCAGTTATCCTCATCATCCTCCTCCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttAGGGCGTTCTCCTCCTCCACAGTTG
B3_Cg_g31376_28_Dla0	g31376	TCAGCTGCCTCTTGTTGGGCATTCCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTGGGGAATTCCTCCTGAATTGGTGC
B3_Cg_g31376_28_Dla0	g31376	CCTCAGCCTCGGTGATTTCTGGGCCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGTATTTGTGACCATAGCCAATATAG
B3_Cg_g31376_28_Dla0	g31376	CTGTGGGGGATTGTAGTTGTCCATGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGCATGAGCACCAGGCCAAAGGTTGG
B3_Cg_g31376_28_Dla0	g31376	TTCTCAAATTTTTTTCCAGCTGCAAttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTCAGCTTGGAGGACACTTTGGGAGT
B3_Cg_g31376_28_Dla0	g31376	GAAGAACTGAGATGGCGTACTGAGGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCTCGGAGAGAGGAGTTAGGAGAGGA
B3_Cg_g31376_28_Dla0	g31376	TGGAGGGAGGTTTCCAATCTCCGCAttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTCCTCACGCTCCTCTTCCTCATCCT

B3_Cg_g31376_28_Dla0	g31376	CCCACCTCAGGCTCTGGCTCATCAGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCAACCTTGACCCAGACACAGCGCCC
B3_Cg_g31376_28_Dla0	g31376	CATCCACCTCCTCAGGGGTCTCCTGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCCAGTGTTGTAAAGATCCATCACAC
B3_Cg_g31376_28_Dla0	g31376	GGGCAGAATGTGTTGTACGTGATGTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttACATTGATCACACAGTCATTGCGAC
B3_Cg_g31376_28_Dla0	g31376	TCCCTCAAAGACACAGGCTCAAAGTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCTTCAAACTGGTAGAAGCCCATTGG
B3_Cg_g31376_28_Dla0	g31376	CGGTTTCTTCTTCCTCCTCCTCATCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttAATCTGGGCTCTCAAATAGTTAATC
B3_Cg_g31376_28_Dla0	g31376	AACATGGGTACCAGCACTAATACGGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttACTGGAGCATCCAATCTTCCAGTGA
B3_Cg_g31376_28_Dla0	g31376	TTGCCTGGGAAAGGGGGATAACTGAttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTCTGGGCTGGTGTTACTGCTGGAAG
B3_Cg_g31376_28_Dla0	g31376	ACTTCTTGATGAGTCTGGCAGACTGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGACAAAGTATGTTTTCTTGTTGGTG
B3_Cg_g31376_28_Dla0	g31376	GGTCCAGGGTTTACCAGGATCATTGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGGTGGGGGTTTGAAGTCTGGTTTTG
B3_Cg_g31376_28_Dla0	g31376	GTTTTGTTTTCCTCCTTTGGGATTTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCTTCCTCATCGCCATCTTTGTCTGT
B3_Cg_g31376_28_Dla0	g31376	TGTCATCCTCCTCTTGTTCTTCTTCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTTCCTCCTCCTCTTCCTCTTCATCA
B3_Cg_g31376_28_Dla0	g31376	TTCTTCCTCCTGTTGTTCCTCGGGCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttTAATTCTGTTCAGTTCCAAACATTT
B3_Cg_g31376_28_Dla0	g31376	TCCCTGTACTCCACCTCAGCGATGTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGAGTGTCCACTAAATTTTTCAAAGC
B3_Cg_g31376_28_Dla0	g31376	CCCAGAATCTAACATGTGATAATGCttCCACTCAACTTTAACCCG

B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCAAACCAATTCCAGCTTGTTCAAAG
B3_Cg_g31376_28_Dla0	g31376	CCATACACGAATCATTTCTTCACGGttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGGCGTTTCAAGTTCTTCACTCTCTT
B3_Cg_g31376_28_Dla0	g31376	AAACAGAGTTCCATAATATTTGGAAttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttCAAAGAGTTTTTCTTGCACTTTGGC
B3_Cg_g31376_28_Dla0	g31376	GTTCATTGTCCTCACCGCCTTCTTTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGTCTCTGACTGTGTCAACATCTGAT
B3_Cg_g31376_28_Dla0	g31376	TGCAACTTCTGTAGACTGGTCAAGTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttACATCTTCAAAGATATCTACAACAT
B3_Cg_g31376_28_Dla0	g31376	AATTTTGATCGTTTTGAATCTTTGCttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGAACACGACTAAGATGGTCATATAA
B3_Cg_g31376_28_Dla0	g31376	TTGGCCTTTCATCAAGAACTTTGGTttCCACTCAACTTTAACCCG
B3_Cg_g31376_28_Dla0	g31376	GTCCCTGCCTCTATATCTttGAGATAAGATTTTGCAGAGATGAAC
B3_Cg_g31376_28_Dla0	g31376	CAACCCTGTGTCTGTACTCGCAGTGttCCACTCAACTTTAACCCG
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTAGCGGACTCATGCTACGACGGAC G
B2_Cg_g11844_33_Dla0	g11844	TTATTTCTGGAGTTTCTTTCGGAGGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCTGTAATCCCGCCCAGGACTCAACC
B2_Cg_g11844_33_Dla0	g11844	GGGGTGGATGTTAGTCCGCCATATCaaATCATCCAGTAAACCGC C
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTGCGTTCGAAGTTCATGGAATCATC
B2_Cg_g11844_33_Dla0	g11844	GGCGAGGGGCGGACTGGCTTGGATAaaATCATCCAGTAAACCG CC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCGGATCCAGGGACTGGTCCCGCAT G
B2_Cg_g11844_33_Dla0	g11844	TTTCCTGGTCTCTGTTTCCAGAAGCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCGGATGTCTCGGATTTCATCCCCAG
B2_Cg_g11844_33_Dla0	g11844	TTGGTCATGGAGTTATCCAAGACACaaATCATCCAGTAAACCGC C
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTCTTCCTCATCTCGGTGGCAAGCTG
B2_Cg_g11844_33_Dla0	g11844	GGGCGCTTCGTGAGATGTACTGCTGaaATCATCCAGTAAACCGC C
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTCCCTTGACCTTAGCTCGGTGAGCA

B2_Cg_g11844_33_Dla0	g11844	CTTCAGTGTGGCAACTGTTTCTTCCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTGAGTTTGAGCTGATCTCAGTCTCT
B2_Cg_g11844_33_Dla0	g11844	TCACGAGTTTCCAGCATTCTCTCAGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTGTGGTCAGATCAATCAACCTCTG
B2_Cg_g11844_33_Dla0	g11844	TCTCCATCTCTGTCTCGGCTCTGTGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCGCCTGGAGTTGCTGGACCTGTTTC
B2_Cg_g11844_33_Dla0	g11844	GTGCTGTTGTTCCGCCTCAGCAAGCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTCAATGTGCGATCCAGGTTACCTT
B2_Cg_g11844_33_Dla0	g11844	AGATCCGTGTTGTCGTCCTCCAGACaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCAGTACGAAGCTTTTCTCTCTCGAC
B2_Cg_g11844_33_Dla0	g11844	CCATGACCGACTTACTGGCGATCTCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGTCTTTCTCCAGTCTGGCAGAGGAG
B2_Cg_g11844_33_Dla0	g11844	GTCCAATGTCTTCTTCAGGGCAGACaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTGGCCATCTGTCGAGACAGTTCCT
B2_Cg_g11844_33_Dla0	g11844	TCCATCTCAGATTTCTCAGTCTCGAaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaAGCTGATCTTCTGAATGTTCTGGTT
B2_Cg_g11844_33_Dla0	g11844	CTCGCTCATGCAGCAGCTGTTTCTCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCAACATTGTGTGGGATTGACGGAGC
B2_Cg_g11844_33_Dla0	g11844	GTCCTGCTCTCTCTCTTGTCTCTGCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCTATCGGCAACCTCATTCTGAAGTC
B2_Cg_g11844_33_Dla0	g11844	CCCTCCAATTCAGCGCGACGGACATaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGTTTCTTAGCTTCATTCAGGGCTTG
B2_Cg_g11844_33_Dla0	g11844	GGACACGCTCCAATAAACCATGGTTaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTCGGCGATCGTGCTCACTGCTGGTG
B2_Cg_g11844_33_Dla0	g11844	GGTGGAGTCAAGCCTCTCCTGCAGGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTGGAGCTCGTTAATGATGGCCTGGT
B2_Cg_g11844_33_Dla0	g11844	CCCTTCTGGAGCTGGCTGATGCGGTaaATCATCCAGTAAACCGCC

B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCGCGCTCCTTCTGGTTGGATTCCTC
B2_Cg_g11844_33_Dla0	g11844	CGGACAGGGTCAGATTGAGGGACTGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGCGGTCAATGGTCAGTTGGAGGGAC
B2_Cg_g11844_33_Dla0	g11844	CTCCGTCTTCTGTAGGTTGGTGGACaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTGTTTAGGGCCTCAATGCGATCTT
B2_Cg_g11844_33_Dla0	g11844	GACTTGGACTCAAGGCCCTGGATTTaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTCATGACCAGCTTAAATCTCTGGAG
B2_Cg_g11844_33_Dla0	g11844	AGACCTGGTTCTCGGTCTCCTTGTCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCAGTTTTGTGCAGCGGTTCTCCGCC
B2_Cg_g11844_33_Dla0	g11844	TCCGTCAAGGGACCGTCTCCCAAGCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTCTCGTCATTGCGCATCTCATTGG
B2_Cg_g11844_33_Dla0	g11844	TCCAGGGCCATCTTCAGGGCCTTCCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGCCTCTTTTCTGTGTCTGCTGCAGC
B2_Cg_g11844_33_Dla0	g11844	TCAGCTTGGCAATCTTCTCCTCCTGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGGTCATGACTTTACGTTCTCTCTCA
B2_Cg_g11844_33_Dla0	g11844	ACTCCTCTCCAGGCTGTTGATCTTGaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaAGCGCAGTCTGAGCGCTGGCAAGGC
B2_Cg_g11844_33_Dla0	g11844	CTGCGGATGGTTTCCTCTTGCAGCAaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaCATCACCCAGGGACTTCTGGAGGAT
B2_Cg_g11844_33_Dla0	g11844	TGTCAATTCCCCTCTTGTCTTCTTCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTTCCTCCAGCTCCTTCAGCTGGATG
B2_Cg_g11844_33_Dla0	g11844	GTGTCTTCTCTCGGTCCTGTCCTTCaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaTCCCTTTCTCGCTCTGCTCCAGCAA
B2_Cg_g11844_33_Dla0	g11844	ATGCTGCGAACATTGGCCAGAGCATaaATCATCCAGTAAACCGCC
B2_Cg_g11844_33_Dla0	g11844	CCTCGTAAATCCTCATCAaaGGACTGCCTCTGGATCTACGTCGAG
B2_Cg_g11844_33_Dla0	g11844	GTTGTACGAAATCTCTGAGGGCCATaaATCATCCAGTAAACCGCC
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaAACCCTATGGCCGTGTATTTTTGTC
B1_Cg_g3234_22_Dla0	g3234	TCACTTCTCAACATCGGGTAATTGTtaGAAGAGTCTTCCTTTACG

B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaATTTCCTCGATTTCGAAAGCCGAGT
B1_Cg_g3234_22_Dla0	g3234	CCGTGAACGTCGACTTTACCCGTTTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGTTTTTTAGACTCTTTCTTAGTCTT
B1_Cg_g3234_22_Dla0	g3234	GCCGACCTTCCTCCCTTTTCTTCTTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGGCAATTTTTAAGGGAGCATTAATC
B1_Cg_g3234_22_Dla0	g3234	TAAGGCTCGAGTTTTCTTCATAAGAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaCGTTTTTCTTTCTTCTTTTCTTGGT
B1_Cg_g3234_22_Dla0	g3234	GTTAAGTTTATGAGTCCTTTCTTTTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaCATGGGCATCTACGGGTGTCACTGC
B1_Cg_g3234_22_Dla0	g3234	CATTCCAAGGTTCTGACGTAATTTCtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaACTTGGGTTATAGACATCAACATTA
B1_Cg_g3234_22_Dla0	g3234	GGGAGGATTTGGTGACGTCAGTCCAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGGGTCAATGTCATATTGATCTAATT
B1_Cg_g3234_22_Dla0	g3234	ATAGAAGCCTCGTTTTCGTCCAGATtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGCTGTTTAGCTTCGTATTTATCATT
B1_Cg_g3234_22_Dla0	g3234	CACTAAGTTCTTCGACTTCTCTGTCtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGTCTTCGCTTTCCGACCTTTCAAAG
B1_Cg_g3234_22_Dla0	g3234	TACGGTCAAAGGTTTAAAGAAAAAGtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGGTTTCTTTTTTATCTCGGGAATCA
B1_Cg_g3234_22_Dla0	g3234	TCGCTCGTAACAACGATGCCGGAAAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGTTCTCTGTCTCGTTTCTTCTTACG
B1_Cg_g3234_22_Dla0	g3234	TTCCGGTCGGCGGGAACATGAATACtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaTTCGTCTTCTTTCTCGTCATCAAAT
B1_Cg_g3234_22_Dla0	g3234	TTGTCGCTGCTCTTTGTGTATTTTTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaAACGCCCCATAGACCAAAGGAAATA
B1_Cg_g3234_22_Dla0	g3234	ATGGTTCGTGCCACTAAGAACCCCAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaCGTCACTTCCGTTATCGTACGATCG
B1_Cg_g3234_22_Dla0	g3234	TGACTATTGCTATGTAAAATGCCACtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGATAAAGATCGCTGCTATAATGGAG
B1_Cg_g3234_22_Dla0	g3234	AGATTTGTGTATAGCCACTGATATAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaCTCTTCTCCTTGATCCAACAGAGCG
B1_Cg_g3234_22_Dla0	g3234	ACGACCGACACATAGAACAGACGTTtaGAAGAGTCTTCCTTTACG

B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaAGGAAACGAGTCCAAAACTGGCCAA
B1_Cg_g3234_22_Dla0	g3234	AAAATCCAGTTGCCATAGCAACCGTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGACGTTTTTGAAATGCGTGTACTGG
B1_Cg_g3234_22_Dla0	g3234	TTTTAGTGCAAGAATGGTGCCCGCTtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaCAGGCCATGAAGATGCCATAGTTGT
B1_Cg_g3234_22_Dla0	g3234	CGACTGCTCCCATCATCATCTCGGAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGAAACCACCGCGGAATGAAGAAAGC
B1_Cg_g3234_22_Dla0	g3234	GCTTCCCCGTACCCGTCTGTGTGAAtaGAAGAGTCTTCCTTTACG
B1_Cg_g3234_22_Dla0	g3234	GAGGAGGGCAGCAAACGGaaGAAGGCAAACACTGTCTTCTGTAGG
B1_Cg_g3234_22_Dla0	g3234	CACAAACAACACATATCCCACAAAGtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaATCCGAGTCTGATTTTGACAGCTCT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TTACTGCCATCTCTGATAGTCCTCCtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTCAAGGTATTCCTTCTTCTCGGGAA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GCAAGCTGTACCAGCCTGCTGTACGtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaCGGAGTACGGGTAGAAGTAGGAACT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	AAGCACCAAAATGAGTCTCCGGCTCtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTTCCTGCTCTAGGGGTATAAACAGA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TTTGCTAGGGGGCGCCACCGGTACCtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaACGTCGTCAGGACCTACATTGTCAA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	CTGACTCCCTGATGCTCAATTTCCTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaCCTTTTCTCTCGCGAAAGCTTCCAA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TGGTTCTTAAAACACTGTCTTCGTAtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGGAAGGATAATATGCCATCGGGTAT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TTTCAACGTGAAAGATGGGCCGAATtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTTGCGGTGAGATCTCTTCATGTCGT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TCATCTGAGTATGGAAGGTCGCGTTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGCCTGCGCCTTCCGTAATAAGGCAA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TGAACATACGGGCTTTTAACGGTGAtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGTATAACTGTAAAGCTAAAGCTGGA

B1_Cg_GNXQN_g16262_20_Dla0	g16262	GTAGTTGGGAAGTTTTTGATAGGCTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaAACATATCTTCGTCGTTGTCAGTTC
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TTATACCATTCCTCTGGTAAGACCTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTTTCCTCGGAAGATTTCTCTGACGC
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GCACAGAGTTTTCTTCACCTTCCTTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTTCTTTGTCATTTTTAGCATTTTCC
B1_Cg_GNXQN_g16262_20_Dla0	g16262	ATTATTAATCTCTTCGGATGATTTCtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGGAAAATCCTCCAACAACTCATCTG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TCTTTAAGGGCTTCCACAGTCTCCGtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaCCTCTGATTCTTTAGGCTCCATGAC
B1_Cg_GNXQN_g16262_20_Dla0	g16262	TAGGTTCGCCTGATTCAGACGTAATtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGATGGCTTTTGGTAGTTCTGCAGTT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	ATCGCTTACGTCCTCAAGCTCCTTTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaGCCGTTACGCTTCCTCTTAAGTCAG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GTCTCTGCAATGTCCTCAGCCTTATtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTTATTCTTGGCTTATTTTGTTGATT
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGCAGGAACTTCCGGATCACTTTTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaTACCATTGCTGATCTCTTTTCAATA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	AAACATCATCACCTCCTGAGCGCGTtaGAAGAGTCTTCCTTTACG
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GAGGAGGGCAGCAAACGGaaCCCAAAAATGGCAACAGCGCTACGA
B1_Cg_GNXQN_g16262_20_Dla0	g16262	GTAGGGGCGGCACTGCAATATACAAtaGAAGAGTCTTCCTTTACG
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCATCGGCATGCCACGTTTATTTTGA
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	TTAGTTTGCATTTCTTCCTAAACGAaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaTGAACATTTGTGCGCTTGCCCAGTC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	GTCTCTTGCTGTGACGCAGAATCTGaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCGGCCTGTAGCCTCTTTCCCAAACG
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	ACATGGGCATGCCTCGTTTTGAGTCaaATCATCCAGTAAACCGCC

B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCATAGGCATACCCCGTTTATCCGTA
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CATGGGCATTCCACGTCCAAGTCGTaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCCAAGGTTTGACGTTGGATATTTGA
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	ACATTAGAATCGGTTAAGTAGTAACaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaGTGGATGAGGGATATGTCTTTCCTC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CGTCAAGTTGTATCAGGCGGCCAATaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCAGAGATCTCCTAAAACGGCCCGGG
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	GTCTTTCTCCTCGCCGTTGTCTTCTaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaTTGTTTTCCAGGTCACTGTCTAAGA
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCAGCAGGGCGAATAATCCTGACACaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaGCTGCAAATCTTTACCATATCTTGG
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	TTTGTAACCATTCGAGTTGTAACTGaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaTTCGCTTTCATCCGGGTACATAAAA
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	GGGAACCTGTCGTCTGCCTTCGTCTaaATCATCCAGTAAACCGCC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	CCTCGTAAATCCTCATCAaaCGACCTAACCGTAGCATTGGCATGC
B2_Cg_Myomodulin_g9633_11_Dla0	g9633	TCGTCAGTTTCACTCAGTCCATTAGaaATCATCCAGTAAACCGCC
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTACGTTTGTCTGTTGGATCTCGTCCA
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCATACAGATTTTCCGAATCGCATATTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTCTTTTGCCAAAACGCATGAATCTTT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CGCATAAATCGTTTTGCATCATCACTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTAAATCGTTTTTCTTCATCTCCACT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CTTCATCTCCACTTTTGCCAAATCTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTGTCGCCACTTTTGCCGAAACGCATA
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCCGAAACGCATAAATCGCTTTTCTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTAAACGCATAAATCGCTTTTCTTCGT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CGCTTTTCTTCGTCGCCACTTTTTCTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTTTCGTCGTCTCCACTTTTGCCGAA
B3_Cg_g7084_FRMF_23_Dla0	g7084	CACTTTTGCCGAATCGCATAAATCGTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTTTCCCAAAACGCATGAATCGCTTT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CATAAATCGTTTTTCTTCATCGCCGTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTAATCGCTTTTCTTCATCTCCACTTT
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCGTCTCCGCTTTTACCAAATCTCATTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTCCATTTCGTCTTCATCATCATATTC
B3_Cg_g7084_FRMF_23_Dla0	g7084	CGAATCGCATGAACCTTTTTTCTGCTTCCACTCAACTTTAACCCG

B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTCATGAACCGCTTACCAAATCGCATG
B3_Cg_g7084_FRMF_23_Dla0	g7084	ATCATTATCGAATCGTTTTCCAAATTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTCGAACAAATCTTTCCATAAGTGTAT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CGTTTATCGTCTCCGCTTCGACCATTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTAACGCTTATCGTCTGTTACCATGTT
B3_Cg_g7084_FRMF_23_Dla0	g7084	CCTTGGGGTCGCGACCAAATCTCATTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTGAATCTTTTTTCGTTTCCATCTTTC
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTTGTCCTCTCGTTTCCCGAACCGCTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTGGATTTCTTCCAAATCGCATGAATC
B3_Cg_g7084_FRMF_23_Dla0	g7084	AAACGCATAAATTTCTTCTCTAAATTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTGTGTTCTTCCAAATCTCATGAAACG
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCATATACGTGGGATCGTCTTCAGCTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTACGTTTGTGTTTTTCGGACGCTGAA
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTTTTCCTCAGAAACCTCGTCGGTTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTCTTCTGGAACGTCTTTGACTAATG
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCCTTTTCTGTGGAATTGCTGTCTTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTGGAGACCTGTTTCTTGTTTGAATTT
B3_Cg_g7084_FRMF_23_Dla0	g7084	GATCTGCCGATCTTCTTTTTCTCAGTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTATCCTCGACACTCCCTTGCTCCCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GACTTTGTTCAGAGCCTCGCGTAAATTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTCACTGTTCCTTCCAAACCTTATAT
B3_Cg_g7084_FRMF_23_Dla0	g7084	TTCCCGAATCTTAAGAATCTTTTGTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTATCGAAGAAATCTTTTGTCTACAGG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GATCACCACTCAGAGCTCGCTTCCCTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTTGGCATAACTCGGGGTTCCTGTAA
B3_Cg_g7084_FRMF_23_Dla0	g7084	TTGCTGACCGAATAAAATCCCAACTTTCCACTCAACTTTAACCCG
B3_Cg_g7084_FRMF_23_Dla0	g7084	GTCCCTGCCTCTATATCTTTTCAATAGTGAACCAGTTTAGAAGGA
B3_Cg_g7084_FRMF_23_Dla0	g7084	TCATCAATCAGATCGTTTGCTGATGTTCCACTCAACTTTAACCCG
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGGCGGTAGACAGCGTTGTGGATGCC
B2_Cg_g24588_19_Dla0	g24588	TTACGCCTTAAATGACATTTTGAATAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAACCGCGCCGGAGCCCAAGCACCTTGA
B2_Cg_g24588_19_Dla0	g24588	GAGCTTGAGAACGACTCCTCTCTGTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGACAGTACAGGAAGAAGTTGATAGA
B2_Cg_g24588_19_Dla0	g24588	ATTCTTTGCGGAACTTCCGCGCGCTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGGAAAAAGCTTCGTTGACAATGCGC
B2_Cg_g24588_19_Dla0	g24588	GTTGATGACGATACCAAGCTTAGCCAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAAAGCATTGGGTGACGCACTGTAGAG

B2_Cg_g24588_19_Dla0	g24588	TCAGCGATCTCCGGTTGCATCATGAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAATCAGGAGGAGGCGCGTCATCTGGTT
B2_Cg_g24588_19_Dla0	g24588	TCAGGATCAGGAAGACGAAGGTGACAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGATGCTCTTCCGTTTGGACATGCCC
B2_Cg_g24588_19_Dla0	g24588	CGCCTTCTTCGCTTTCTCCTGCCCTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAAACATGTTCAGGATAAAGATGGTCA
B2_Cg_g24588_19_Dla0	g24588	TTGGCCCTGACAACGCGAGAAATGAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGTACCCAGAATTCATAGCGCTTGCT
B2_Cg_g24588_19_Dla0	g24588	ATGGAACCAGGACTAGAAACATACAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGAAAGCGGCGTCGTCCGGTTCTCGG
B2_Cg_g24588_19_Dla0	g24588	TTCGGTCATCCCAAACTCGGTAAGAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAAAGTGGGGGGTGTTGATGAGGAAAC
B2_Cg_g24588_19_Dla0	g24588	GAGGCCGGGATGGGGTGGTAGGTCTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAACGGCCCTCTTGTTGCTGCACATTTC
B2_Cg_g24588_19_Dla0	g24588	GGAACGCAATGATACCGATCCCGACAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGTAACGATCCACAGTGACCCCGACC
B2_Cg_g24588_19_Dla0	g24588	TGCTTTTGTGAACCAGCAAACTTGTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAAAGAAGATGGGGTAGCCGATGTACG
B2_Cg_g24588_19_Dla0	g24588	ATCCAGATACTGTAGACTACGAACAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAATGACAGAGGGCACTAGCATCGTCAG
B2_Cg_g24588_19_Dla0	g24588	AAAACACACCGAACGCGTAGCTGGTAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAACCCCATGTCGGACAGCGCCTGTGCA
B2_Cg_g24588_19_Dla0	g24588	GTCCGTGATGAAGAAGAAGAACAGAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAAGACTTCCGGCTCCATATCAGGACAG
B2_Cg_g24588_19_Dla0	g24588	AGATAGACCCCGGTGGAGGTCCTCAAAATCATCCAGTAAACCGCC
B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAACGAAAATAGTACCGAAGATTGAACT
B2_Cg_g24588_19_Dla0	g24588	GGATGTTCCCGACGATTCCAACAATAAATCATCCAGTAAACCGCC

B2_Cg_g24588_19_Dla0	g24588	CCTCGTAAATCCTCATCAAACATGACGTCATCTGCTACAGTATTG
B2_Cg_g24588_19_Dla0	g24588	CACCCAGTGCATCTTCTTCGCCACCAAATCATCCAGTAAACCGCC