Tracking cell fates during limb regeneration

Bahadir Çağri Çevrim

Numéro National de Thèse : 2021LYSEN006

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N°340
Biologie Moléculaire, Intégrative et Cellulaire

Discipline : **Sciences de la vie et de la santé**

Soutenue publiquement le 22/02/2020, par :
Bahadır Çağrı ÇEVRİM

Tracking cell fates during limb regeneration

Suivi du destin des cellules au cours de la régénération des membres

---

Devant le jury composé de :

Prof Enrique AMAYA	Professeur, Mancherster University	Rapporteur
Prof Mansi SRIVASTAVA	Professeure, Harvard University	Rapporteure
Dr Guillaume BALAVOINE	DR CNRS, Institut Jacques Monod	Examinateur
Prof Angelika STOLLEWERK	Professeure, Queen Mary University of London	Examinatrice
Dr Michalis AVEROF	DR CNRS, Institut de Génomique Fonctionnelle de Lyon	Directeur de thèse

Table of Contents

Abstract

The ability to regenerate lost body parts is shared by many animals. Although scientists have been interested in regeneration for a long time, our knowledge of this process remains very limited. With the exception of zebrafish, established genetic model organisms (flies, nematodes, and mammals) have very limited regenerative abilities. During my thesis project, I used the emerging model organism Parhyale hawaiensis and focused on their peripheral sensory organs (sensilla), to explore two fundamental questions of complex organ regeneration: first, how faithful is regeneration in restoring these organs? And second, which are the cellular progenitors of the sensilla during limb regeneration, and what is their developmental potential?

The comparison of the sensilla of regenerated and non-amputated legs of Parhyale revealed that the diversity, distribution, and function of the sensilla are faithfully restored during regeneration. Next, I focused on identifying the progenitors of sensilla cells during regeneration. I improved an existing live imaging and cell tracking technique of limb regeneration and extended it to record the entire course of regeneration. To identify the sensilla cells in these recordings, I tried to generate cell-type-specific reporter animals, first by inserting a reporter into putative marker genes using CRISPR, then by generating cis-regulatory element reporters. As neither of these approaches was fruitful, I developed a method to identify cell fates in the regenerating limb via antibody stainings. This is a promising technique to identify the progenitors of different cell types in regenerating limbs.

Résumé

La capacité de régénérer les parties du corps perdues est partagée par de nombreux animaux. Même si les scientifiques s'intéressent à ce processus depuis très longtemps, nos connaissances sur la régénération, même pour certains des sujets très basiques, sont très limitées. À l'exception du poisson zèbre, les organismes modèles génétiques établis (mouches, nématodes, et mammifères) ont des capacités de régénération très limitées. Au cours de mon projet de thèse, j'ai utilisé l'organisme modèle émergent Parhyale hawaiensis et me suis concentré sur leurs organes sensoriels périphériques (sensilles), pour explorer deux questions fondamentales de la régénération d'organes complexes: premièrement, à quel point la régénération est-elle fidèle à la restauration de ces organes? Et deuxièmement, quels sont les progéniteurs cellulaires de les sensilles lors de la régénération des membres et quel est leur potentiel de développement?

La comparaison de les sensilles des jambes régénérées et non amputées de Parhyale a révélé que la diversité, la distribution et la fonction de les sensilles sont fidèlement restaurées lors de la régénération. Ensuite, je me suis concentré sur l'identification des progéniteurs des cellules de les sensilles pendant la régénération. J'ai amélioré une technique existante d'imagerie en direct et de suivi cellulaire de la régénération des membres, et je l'ai étendue pour enregistrer tout le cours de la régénération. Pour identifier les cellules de les sensilles dans ces enregistrements, j'ai essayé de générer des animaux rapporteurs spécifiques au type de cellule, d'abord en insérant un rapporteur dans des gènes marqueurs putatifs à l'aide de CRISPR, puis en générant des rapporteurs d'élément de régulation cis. Comme aucune de ces approches n'était fructueuse, j'ai développé une méthode pour identifier le destin des cellules dans le membre en régénération via des immunomarquages. Il s'agit d'une technique prometteuse pour identifier les progéniteurs de différents types de cellules dans les membres en régénération.

"If there were no regeneration there could be no life. If everything regenerated there would be no death. All organisms exist between these two extremes. Other things being equal, they tend toward the latter end of the spectrum, never quite achieving immortality because this would be incompatible with reproduction."

Richard J. Goss, 1969, p. 1

1.Introduction

In the broadest sense, regeneration is the replacement of a body part, as part of a natural process of tissue turnover or upon damage. Regeneration can be seen in different levels of life; cells regenerate some of their organelles or specialized parts, like the unicellular algae *Acetabularia* regenerating its stalk (Novák and Bentrup, 1972). The cell can be renewed to regenerate the function of some tissues, as in the physiological renewal of epithelia during an individual's lifetime. In some animals, regeneration ability extends beyond these mechanisms; they can regenerate body parts with multiple cell types and complex architectures, upon loss or injury. In this study, I focus on the latter case and use the term regeneration to refer to post-embryonic complex organ regeneration.

The ability to regenerate is widely but unevenly distributed among the animal kingdom (figure 1.1). There are species with extensive and very poor regenerative abilities in almost every major animal phylum, but the extent of regenerative capacity varies among species; for some, it is limited to specific internal organs like liver regeneration in humans, some can regenerate external appendages like tail regeneration in lizards, and some animals can regenerate large portions of their trunk as can be seen in the whole-body regeneration of planarians (Bely and Nyberg, 2010), (Tiozzo and Copley, 2015). For some animals there are also temporal limitations, for example, *Xenopus* tadpoles can regenerate some body parts including limbs and internal organs whereas adults can not (reviewed in Yun, 2015).

Regeneration is an old field of scientific research, the first studies dating back to the 1740s; the initial work of Abraham Trembley on Hydra, was followed by Charles Bonnet on annelids, Lazzaro Spallanzani on amphibians, and Peter Simon Pallas on planarians, shortly after (Carlson, 2007). Although regeneration has been studied for such a long time, one of the early questions is still abiding: how did regeneration evolve? Most of the early studies on regeneration were reporting the presence or

absence, and the extent of regeneration in diverse species. The resulting picture shows a very patchy distribution of regeneration in the tree of life, which does not provide a clear picture on the evolutionary origins of regeneration. The wide distribution of regenerative capacities in basal lineages of Metazoa, e.g. in sponges and cnidarians, advocates for the presence of regenerative abilities in early metazoans. On the other hand, there are closely related species with very different regenerative abilities in every major animal phylum, suggesting that this ability has been lost and/or re-gained several times independently (Goss, 1992; Tiozzo and Copley, 2015). Mapping the presence of regeneration traits may not be sufficient to infer the evolutionary history of regeneration. Comparing the similarities and differences between the regenerative mechanisms of diverse species, such as the cellular and molecular processes involved, would help further to identify whether they share a common origin or not.

Since the beginning of the molecular biology revolution, scientists have started to untangle how biological molecules are coordinating life. Since then, our understanding of many developmental mechanisms, such as embryonic axis formation, and segmentation, has deepened. Detailed information about these mechanisms is coming from the workhorses of molecular biology: *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Mus musculus*. Extensive tools and methods attracted more scientists to work on these model organisms, contributing further to the resources available to work with these animals. Studies on these animals expanded our understanding of the molecular dimension of life. Yet, our understanding of mechanisms of regeneration remained limited, mainly because these animals have very limited regeneration abilities. Many technological advances in recent decades, e.g. affordable sequencing technologies, easily adaptable genome editing techniques like CRISPR, and advanced imaging techniques, accelerated research in non-conventional model organisms (Gladfelter, 2015). Additional animal models were established for regeneration studies, helping to unravel the mechanisms by which a damaged organ is regenerated.

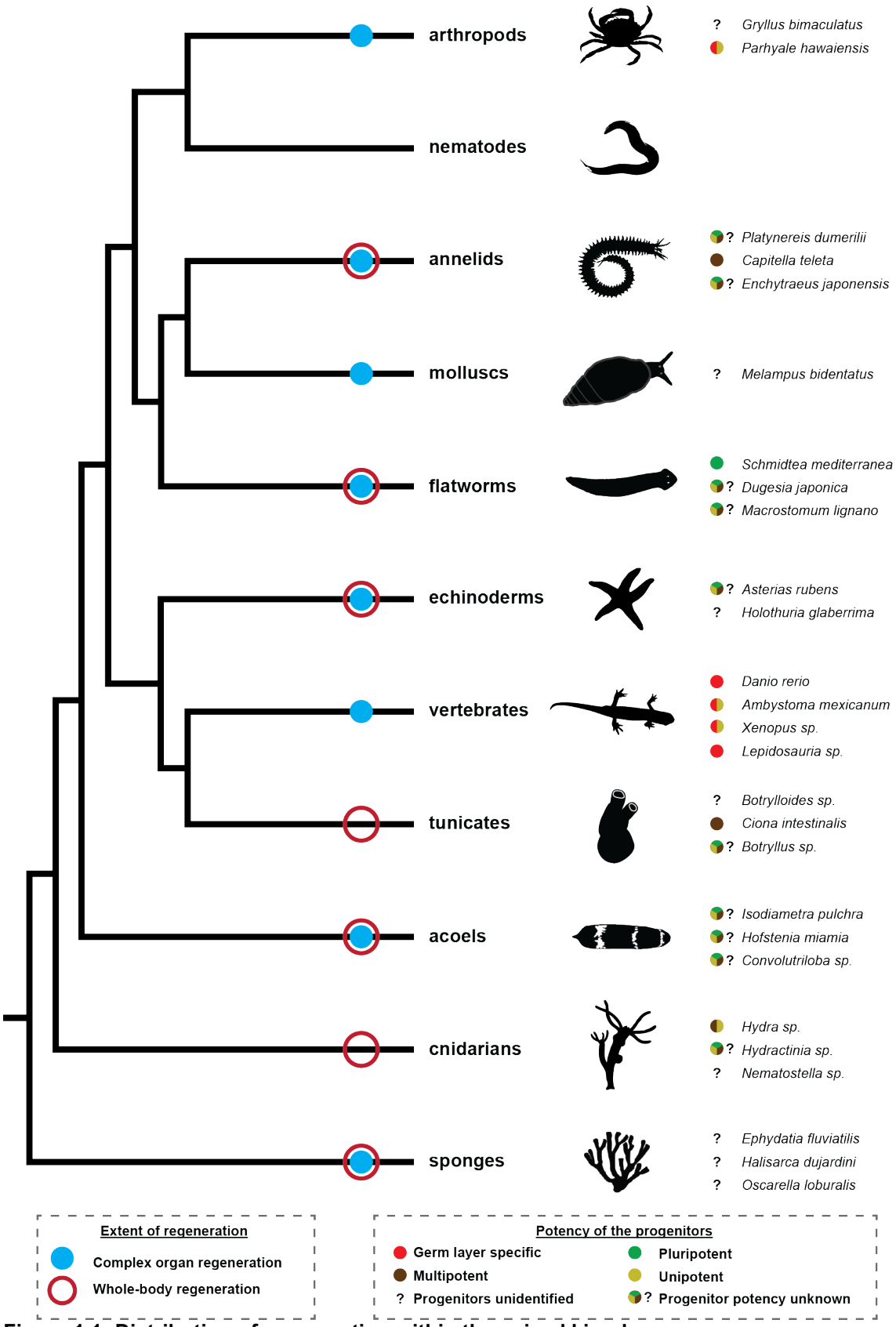


Figure 1.1: Distribution of regeneration within the animal kingdom.
The presence of whole-body regeneration and complex organ regeneration is marked for each phylum of the animal kingdom. Some commonly used model organisms for regeneration studies and the potency of their cellular progenitors are noted. Species marked with two colors have different progenitor populations with different potencies.

A fundamental topic of developmental biology is the regulation of cell potency. Embryonic development starts with a single omnipotent cell, the zygote, and as development proceeds, the potency of most cells becomes restricted. Identifying the potency of the cells that participate in regeneration is an important step for understanding the underlying mechanisms. A wounded tissue may contain different types of progenitor cells for regeneration. First, the regenerated tissue may arise from resident stem cells, self-renewing cells that can generate one or multiple different cell types. Second, terminally differentiated cells may become converted into dividing progenitor cells through a process of dedifferentiation. Finally, a differentiated cell may change fate and transdifferentiate into another cell type, which may not necessarily require cell divisions (Jopling et al., 2011). Regeneration could involve several of these processes at the same time. It is important to determine the identity and the potency of the progenitors of regeneration, which would eventually allow us to study the molecular mechanisms controlling them.

Unipotent and multipotent progenitors of cnidarian regeneration

Cnidarians have remarkable regenerative abilities and are the oldest models for regeneration studies (Holstein et al., 2003). The widely studied model *Hydra* has its body consisting of two epithelial layers: an outer layer of ectodermal cells and an inner layer of endodermal cells. During regeneration, cells in each layer can divide and contribute to the regeneration of the layer they belong to (T. C. G. Bosch, 2007). In addition to these unipotent epithelial progenitors, there are also multipotent interstitial cells (i-cells) that can generate other cell types, e.g. gametes and neurons, but not the epithelial cells (T. Bosch and David, 1987), reviewed in (David, 2012). In another cnidarian model with a similar body plan, *Hydractinia*, i-cells are the sole progenitors of regeneration (Gahan et al., 2016; Frank et al., 2020). In both species, it is unclear whether all the i-cells have the same potential, or whether there are different subpopulations of i-cells with different potency. Therefore, even though *Hydractinia* i-cells are pluripotent as a population, the presence of individually pluripotent progenitors is yet to be confirmed.

Pluripotent progenitors mediate planarian whole-body regeneration

The adult stem cells of planarians, called neoblasts, are the only dividing cells in the body of planarians and the cellular progenitors of whole-body rege

neration (Baguna et al., 1989). At least some of the neoblasts of *Schmidtea mediterranea*, called clonogenic neoblasts, are pluripotent (Wagner et al., 2011). Even though some neoblasts are capable of producing every cell type in the body, there is

strong evidence for subpopulations of neoblasts being fated to regenerate certain cell types, like eyes, muscles, or protonephridia cells. These specialized-neoblasts are expressing the molecular markers of the tissues they are fated to regenerate (reviewed in (Reddien, 2013)). The details of how the fates of neoblasts become restricted are unknown, but the fate of each neoblast is strongly correlated with its localization in the body (reviewed in (Reddien, 2018)). In two other animal clades in which whole-body regeneration is present, tunicates (Kassmer et al., 2019) and acoels (Gehrke and Srivastava, 2016), neoblast-like cells have been described. These cells are known to be pluripotent as a population but the potencies of the individual cells are unknown. Multipotent progenitors of whole body generation in diverse species share a common molecular signature, the germline multipotency program genes such as *piwi*, suggesting that these cells might have a single evolutionary origin and were present in early bilaterians (Gehrke and Srivastava, 2016; Kassmer et al., 2020; Lai and Aboobaker, 2018). But we need a more detailed description of the developmental potential of progenitors of regeneration in different species, as well as the underlying molecular machinery to have a unified view.

Emerging experimental models, like the tunicate *Botrylloides diegensis* and acoel *Hofstenia miamia*, are promising to illuminate details of whole body regeneration in diverse taxa, and along with information gathered from more established planarian and cnidarian models would allow us to make comparisons and eventually gain a better understanding of the evolution of whole-body regeneration.

De-differentiation and lineage-restricted progenitors of regeneration in vertebrates

Some amphibians can regenerate their tails, and *Xenopus laevis* (frog) tadpoles are one of the well-studied models. The fate of the progenitors of the regenerated tadpole tail is restricted to lineage of each cell type (Gargioli and Slack, 2004): Pax7-expressing satellite cells are the progenitors of regenerated muscles (Chen et al., 2006; Gargioli and Slack, 2004), and Sox2-expressing cells in the blastema are the progenitors of the spinal cord (Gaete et al., 2012). Muscle regeneration does not require muscle de-differentiation. The origin of the progenitors of other cell types is not known yet.

Initial studies on another amphibian, the salamander *Ambystoma mexicanum* (axolotl), suggested that the progenitors are changing their fate during tail regeneration (Echeverri and Tanaka, 2002). However, these results have been challenged in a later study (McHedlishvili et al., 2007). The experimental methods used (*in situ*

electroporation and tissue grafting) to label and track different cell types in these studies could not directly address the potency of the axolotl tail regeneration progenitors. Later studies showed that multipotent neural progenitors can regenerate different glia and neuron cell types of central and peripheral nervous systems during axolotl tail regeneration (McHedlishvili et al., 2012), and at least some of these progenitors are generated by dedifferentiation of adult neural stem cells into neuroepithelial-like cells (Rodrigo Albors et al., 2015).

In a study on lizard *Lepidodactylus lugubris*, Londono et al. isolated cartilage cells from the tail, cultured them, labeled them with DiI, transplanted these cells into tails of naive animals, and finally amputated the tails and followed the fate of these cells. They also labeled muscle cells genetically and followed their fate during tail regeneration. The results suggest that cartilage and muscle cells can de-differentiate into bipotent progenitors that can give rise to both of these two cell types during tail regeneration (Londono et al., 2017). These results suggest that the tail regeneration mechanisms between amniotes and amphibians are different in terms of cellular progenitors.

Unlike frogs and lizards, some salamanders can also regenerate their legs. Studies using chimeric animals – generated by grafting a transgenic GFP-expressing tissues to the legs of wild type animals – uncovered the fate restrictions of different pools of progenitors during axolotl and *Cynops pyrrhogaster* (newt) leg regeneration: progenitors derived from muscle tissue contribute only to the muscles of the regenerated legs, and ones derived from the cartilage only contribute to the cartilage. The progenitors coming from dermis/epidermis have a more flexible lineage and contribute to both dermal tissues and cartilage (Kragl et al., 2009; H. V. Tanaka et al., 2019). In axolotls, mature muscle cells do not de-differentiate during regeneration (Sandoval-Guzmán et al., 2014); satellite cells are the source of the regenerated leg muscles (Fei et al., 2017; Sandoval-Guzmán et al., 2014). In adult newts *Notophthalmus viridescens* and *Cynops pyrrhogaster*, de-differentiated muscle cells are the only muscle progenitors during leg regeneration (Sandoval-Guzmán et al., 2014; Tanaka et al., 2019). Interestingly, satellite cells are the progenitors of muscles during larval newt (*C. pyrrhogaster*) leg regeneration (Tanaka et al., 2019), but not in adults.

Adult mammals have little regenerative ability, limited to internal organs like the pancreas and liver, and murine digit tips in adult stages. But since *M. musculus* (mouse) is a well-established model organism, the mechanisms of these limited regeneration events are investigated extensively. During pancreas regeneration,

mature β-cells can de-differentiate into a unipotent progenitor state (Talchai et al., 2012). If all β-cells are depleted, α-cell and δ-cell can trans-differentiate into β-cells (Chera et al., 2014; Thorel et al., 2010). Moreover, some pancreatic cell types, e.g. duct cells, are shown to be pluripotent (Gao et al., 2003) and also trans-differentiate into β-cells (Lysy et al., 2013), reviewed in (Domínguez-Bendala et al., 2019). Mature liver cells are highly differentiated but they can re-enter the cell cycle and restore the liver mass when the liver is damaged. Bipotent hepatic progenitor cells and trans-differentiation of diverse cell types into hepatic progenitors are also contributing to liver regeneration (reviewed in Zhao et al., 2009). Overall, there is high cell plasticity during mammalian internal organ regeneration.

In mouse digit tip regeneration progenitors are lineage-restricted. By using multiple lineaging methods (cell transplantation, fate mapping of embryonic germ layers, and clonal analysis of specific cell populations using the Cre/lox system), progenitors of regenerated mouse digit tips were shown to be restricted to the germ layers they originated from. Moreover, progenitors of dorsal and ventral ectoderm, tendons, bone, and blood vessels are shown to be restricted within their lineages (Rinkevich et al., 2011).

Freshwater teleost *Danio rerio* (zebrafish) can regenerate many tissues and organs including the retina, spinal cord, brain, hair cells, heart, caudal fin, kidney, and liver (reviewed in Sehring and Weidinger, 2019). This extensive regenerative capacity combined with several molecular and genetic tools makes zebrafish one of the most powerful models for regeneration studies (reviewed in Marques et al., 2019). There are multiple studies on regeneration progenitors of different organs. By transposon-mediated mosaic analysis, the progenitors of the regenerated fins are shown to be highly restricted within their lineages. Progenitors derived from 9 different lineages, namely epidermis, iridophore, xanthophore/melanocyte, intraray glia, lateral line, osteoblast, dermal fibroblast, vascular endothelium, and resident blood lineages, are contributing to regeneration of their cognate cell types during fin regeneration (Tu and Johnson, 2011). A study has independently confirmed these results by tracking the fate of randomly labeled clones via the Cre-lox system (Stewart and Stankunas, 2012). Tracing single osteoblasts during regeneration showed that these de-differentiate and contribute to the regenerated fin (Knopf et al., 2011; Sousa et al., 2011). *De novo* generated osteoblasts, derived from a specific osteoblast progenitor cell type, were also identified in the regenerated fins (Singh et al., 2012; Ando et al., 2017). This dual progenitor pool of osteoblasts is also present in *Oryzias latipes* (medaka) fin regeneration (Dasyani et al., 2019). Apart from the fins, the potency of progenitors has

also been investigated in the regeneration of other organs. Adult hair cell regeneration relies on unipotent progenitors (Pinto-Teixeira et al., 2015), whereas multipotent progenitors are sustaining the regeneration of cardiac myocytes, smooth muscles and endothelial cells (Beltrami et al., 2003; Vicinanza et al., 2017), and of kidney cells, all of which are derived from the ventrolateral mesoderm (Diep et al., 2011). Likewise, regeneration of the retina of the adult relies on multipotent stem-cell-like progenitor cells, which are derived from de-differentiation of Müller glial cells (Powell et al., 2016).

Overall, the progenitors of vertebrate limb regeneration are not pluripotent like the ones of planarians. They are bound to produce cell types that belong to their lineage of origin, and this mechanism is valid for several species. Yet, many of the tissues with identified progenitors consist of several cell types, e.g. bone tissue consists of osteoblasts, osteocytes, and bone lining cells (Florencio-Silva et al., 2015), and it is not always clear whether these cells have common or distinct progenitors. There could be multipotent progenitors producing several closely related cell types, or there could be a different progenitor for each different cell type. Additionally, whether the progenitor cells are produced by de-differentiation of some of the terminally differentiated cells or from adult stem-cell-like progenitors (like the satellite-cells for muscle) is yet to be discovered.

Interestingly, in both mouse and zebrafish, the potencies of the progenitors of internal organs and limbs are different, indicating that the regeneration mechanisms in these two body parts have different evolutionary histories.

Arthropod leg regeneration relies on lineage-restricted progenitors

There are four major groups in the arthropod phylum, chelicerates, myriapods, hexapods, and crustaceans, and all include species with good regenerative abilities (Maruzzo and Bortolin, 2013). Arthropods have similar regenerative capacities with vertebrates: they cannot regenerate their whole-body, but some post-anal structures of the trunk (e.g. (Clare et al., 1990; Mees et al., 1995)), eyes (e.g. (Ventura et al., 2018)), and limbs (reviewed in (Suzuki et al., 2019)) are reported to regenerate in some species. To my knowledge, there is no report of internal organ regeneration except in the larval imaginal discs of Drosophila melanogaster (Schubiger, 1971).

D. melanogaster imaginal discs are simple tissues composed of a continuous epithelial sheet, folded in two layers of cells with different fates (Aldaz and Escudero, 2010). Upon tissue damage, the cells of the imaginal disc proliferate and give rise to cells with distinct cell fates (reviewed in (Fox et al., 2020)).

The most extensively studied complex organ regeneration model in arthropods is limb regeneration. A few different arthropods have been used to study leg regeneration, the most classic models being *Gryllus bimaculatus* (cricket) and *Blattella germanica* (cockroach). These studies have contributed to the understanding of different aspects of regeneration, such as positional identity and hormonal control (Das, 2015). But lineage tracing techniques are lacking for these animals therefore the cellular progenitors of the limb regeneration are not discovered yet (Suzuki et al., 2019). Recently, the genetically tractable crustacean *Parhyale hawaiensis* was established as an experimental system for studying limb regeneration (Alwes et al., 2016; Konstantinides and Averof, 2014). Genome and transcriptome assemblies, and established genetic approaches such as transposon- or CRISPR-mediated transgenesis make Parhyale an attractive model for regeneration studies (reviewed in (Grillo et al., 2016)).

Following limb amputation in *Parhyale*, cells located in the distal leg stump proliferate and regenerate a new leg (Konstantinides and Averof, 2014). Fate mapping using genetic mosaics revealed a lineage restriction with respect to germ layers: ectodermal and mesodermal cells only derive from ectodermal and mesodermal progenitors, respectively (Konstantinides and Averof, 2014), in a manner similar to the vertebrate leg, tail, and fin progenitors, described previously. Cell transplantation experiments revealed the presence of satellite-like cells in Parhyale, which can act as progenitors of muscle cells during regeneration (Konstantinides and Averof, 2014). Before this study, satellite cells had only been identified in chordates (Somorjai et al., 2012). Their presence in *Parhyale* suggests this cell type may have been present in the last common ancestor of vertebrates and arthropods (Baghdadi and Tajbakhsh, 2018; Konstantinides and Averof, 2014). Later, satellite cells were also identified in *D. melanogaster* (Boukhatmi and Bray, 2018; Chaturvedi et al., 2017), supporting the hypothesis that satellite cells may be shared among diverse bilaterians.

In *Parhyale*, lineage tracking has been established using a live imaging approach (Alwes et al., 2016). In this technique, transgenic animals with a ubiquitous nuclear reporter are mounted using surgical glue, immobilizing their amputated leg on a coverslip, and the regenerating leg is subsequently imaged with a confocal microscope. The resolution of the recordings is high enough to track individual cells through time and space; thus, cell behaviors and lineages can be directly observed during regeneration. Lineaging epidermal cells during leg regeneration revealed that there are no specialized epidermal progenitors for the epidermis. Instead, most

epidermal cells in the stump are contributing to the regenerated leg epidermis by generating a small clone of daughter cells (Alwes et al., 2016).

The objective of my PhD research has been to identify the progenitors of terminally differentiated cell types in regenerated limbs using the live imaging technique. As a pilot, to develop a generalized approach that could apply to any cell type, I decided to focus on sensory organ cells.

Arthropod peripheral sensory organs

Arthropods constitute the largest phylum in the animal kingdom, containing over a million species with diverse morphologies and adapted to diverse habitats. A common feature shared by all arthropod species is the rigid exoskeleton covering their body, the cuticle, and the peripheral sensory organs, called sensilla, that decorate the cuticle surface (Hallberg and Hansson, 1999). There are two types of sensilla, based on their localization and function: external and internal sensilla. External sensilla (also referred to as sensory bristles) are used to sense chemical and mechanical cues from the environment. Internal sensilla, on the other hand, are specialized to perceive physical signals, like stretch and pressure, from the internal tissues. The two most common types of internal sensilla are multidendritic neurons, which are present in all arthropod groups, and chordotonal organs, which are only found in insects and crustaceans (reviewed in Hartenstein, 2005). The structures of external sensilla and chordotonal organs have been described in various arthropod species (Clarac, 1990; Field and Matheson, 1998; Whitear, 1960), but the majority of studies on the development of these organs come from insects, mainly *D. melanogaster* (Hartenstein, 2005). According to developmental data from insects and structural data from both insects and crustaceans, external sensilla and chordotonal organs are considered to be homologous, and all are built with a common plan (Hallberg and Hansson, 1999).

Arthropod peripheral organs consist of multiple cell types organized in a stereotypical manner. External sensilla posses an external shaft of various shapes and sizes, depending on species and the function of the sensillum; this is built by a type of outer accessory cell named the trichogen (shaft) cell. Another type of outer accessory cell, the tormogen (socket) cell, makes the joint that attaches the shaft to the neighboring epidermis. The outer accessory cells of the sensillum surround two types of inner cells: a sensory neuron, and a thecogen (sheath) cell that ensheaths the dendrite of the neuron. Part of the neuronal cell body and the axon could be ensheathed by glia, which is absent in some types of sensilla (figure 1.2a). Chordotonal organs share several structural features with the external sensilla

(Moulins, 1976). The sensory neurons of chordotonal organs are organized in units called scolopidia, which are ensheathed basally by a ligament cell (homologous to the glia of external sensilla (E. C. Lai and Orgogozo, 2004)) and apically by a scolopale cell (the homologue of the sheath cell). A cap cell, located apically, attaches the scolopidia to the attachment cell, which anchors the whole sensory organ structure to the cuticle. The latter two cells are homologues of the socket and shaft cells, respectively, of the external sensilla (reviewed in Hartenstein, 2005, figure 1.2b).

The development of *D. melanogaster* sensilla is well studied and there is a good understanding of both cellular and molecular dynamics of this process. First, small clusters of cells in the developing epidermis and imaginal discs start to express so-called proneural genes, and these clusters are defined as proneural clusters. The proneural genes are all bHLH transcription factors: *achaete* (*ac*), *scute* (*sc*), and *lethal of scute* (*l'cs*) are expressed in the proneural clusters that give rise to mechanosensory sensilla and most of the multidendritic neurons, *atonal* (*ato*) is expressed in the proneural clusters of chemosensory sensilla and chordotonal organs, and *amos* is expressed in the proneural clusters of chemosensory sensilla and some multidendritic neurons (Goulding et al., 2000; Jarman et al., 1993; 1995; Ruiz-Gómez and Ghysen, 1993). Proneural genes activate expression of a set of genes and signaling pathways, including *Notch*/*Delta*, which leads the singling out the sensory organ precursor (SOP) fate in a single cell, or a small number of cells; the others cells of the proneural cluster adopt an epidermal fate (Hartenstein and Posakony, 1990; Lyman and Yedvobnick, 1995). The SOP cell acquires a different morphology from the surrounding epidermal cells, they become more spherical in shape and their nuclei are located more basally (Hartenstein et al., 1994). There are several molecular markers expressed in all SOPs, e.g. *neutralized* (*neur*), *senseless* (*sens*), and *asense* (*ase*); *cut* is expressed exclusively in the SOPs of external sensilla (reviewed in Hartenstein, 2005). SOPs, also called pI cells, go through a series of stereotypical cell divisions to generate all sensilla cells. The first division of SOP generates two cells pIIa and pIIb. pIIa is the precursor of the outer accessory cells and it divides once more to generate the shaft and the socket cell. pIIb is the precursor of the neural lineage of the sensilla; division of pIIb generates px and pIIIb cells, and pIIIb divides to generate the sensory neurons and the sheath cell. The fate of px cells depends on the type of sensilla they belong to (Lai and Orgogozo, 2004). Chordotonal organs have a similar cell lineage, with the terminal cells acquiring different identities (as explained previously). In most external sensilla px goes through apoptosis, but in some sensilla, it differentiates into glia. In multiply-innervated sensilla, px goes through an additional round of division and the

daughter cells differentiate into additional neurons (Lai and Orgogozo, 2004). In the multidendritic neuron lineages, pIIa and pIIIb cells go through apoptosis and the px cell differentiate into a multidendritic neuron. In non-innervated bristle lineages, the pIIb cell goes through apoptosis and only the outer accessory cells differentiate (Hartenstein, 2005).

The cellular architecture of insect and crustacean sensilla resemble each other ((Hartenstein, 2005), see figure 6.3 in chapter 6), yet there are some differences, for example, crustacean external sensilla may contain multiple shaft cells and neurons (Hallberg and Hansson, 1999). Further, there are some indications that some aspects of sensilla development are shared among different arthropod groups. During limb development in the spider *C. salei*, the groups of epithelial cells specified as SOPs express spider *achaete/scute homolog* (*ash*) genes. Some of these cells proliferate and later delaminate from the epithelia and migrate basally. The delaminating cells express a marker of the neural lineage of *D. melanogaster* sensilla, *Prospero*, and are thought to differentiate into neurons and glia (Stollewerk and Seyfarth, 2008). Notch signaling is important both in the SOPs as well as the surrounding epithelial cells, as in *D. melanogaster* (Gold et al., 2009). In the myriapod *Glomeris marginata* SOPs express *ato* and *ash* genes (Pioro and Stollewerk, 2006). Likewise, *ato*, *ash*, and *pros* orthologs are also expressed in developing sensory organs in the crustacean *Daphnia magna* (Klann and Stollewerk, 2017). These data suggest that the developmental mechanisms used to specify the sensilla are at least partly conserved within the arthropods. There are also data suggesting that different sensilla morphologies are generated as a result of the variation in the expression pattern and dosage of proneural genes, rather than involvement of novel genetic factors (Klann and Stollewerk, 2017). These data further support the idea that a common molecular program controls the development of diverse sensilla in different arthropods.

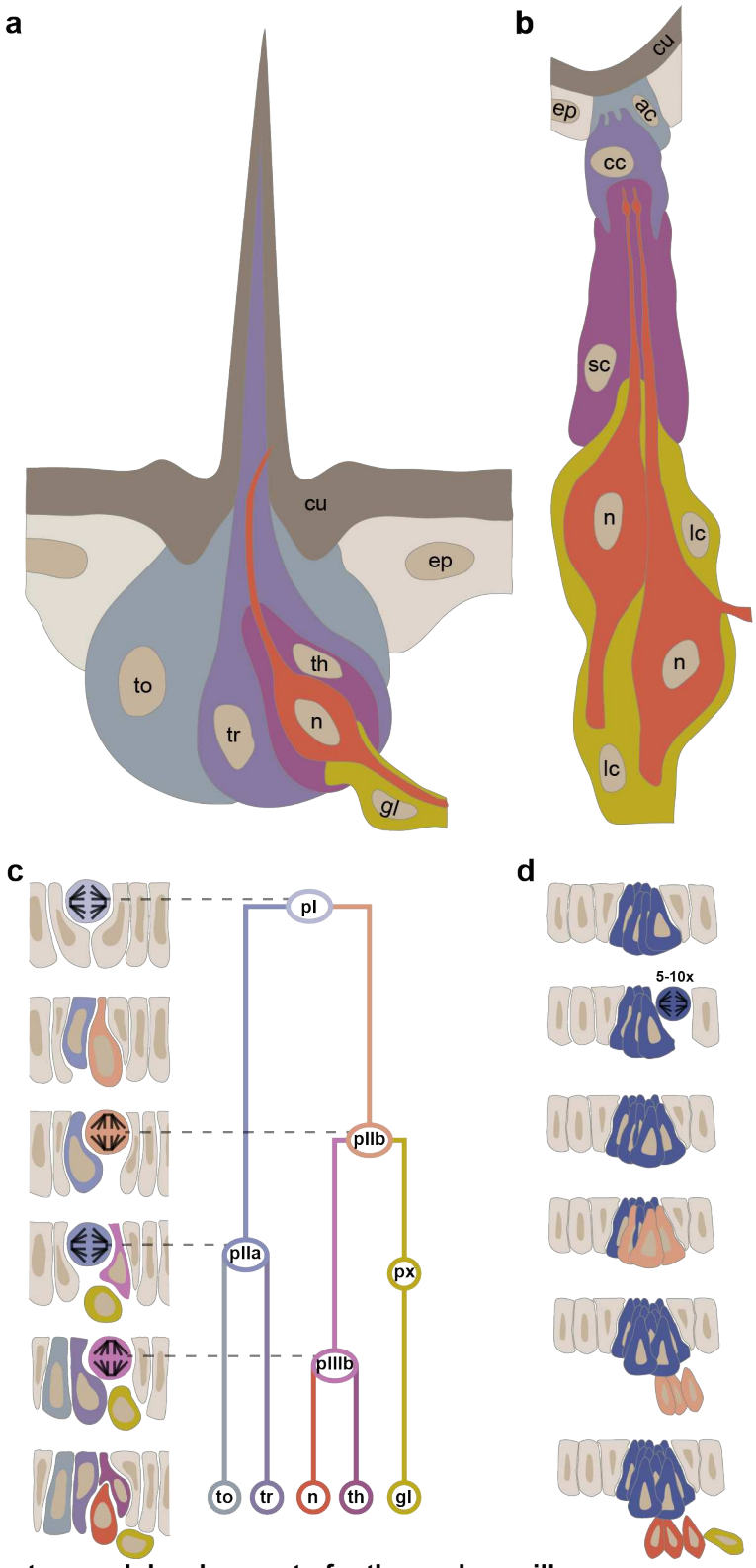


Figure 1.2: Structure and development of arthropod sensilla
Schematic representation of an insect (a) external sensillum and (b) chordotonal organ. (c) In D. melanogaster, all the cells that make up an external sensillum or a chordotonal organ are generated from a single precursor cell (called SOP or pI cell) following a stereotypic pattern of cell divisions. (d) In the spider Cupiennius salei, a group of epidermal cells acquire the SOP fate. Some of these cells proliferate, and later, a part of this SOP group delaminates and acquires neuronal or glial fates. ac: attachment cell, cc: cap cell cu: cuticle, ep: epidermal cell, gl: glia, n: neuron, th: thecogen, tr: tricogen, to: tormogen (panels a, b, c are modified from (Hartenstein, 2005), and d is modified from (Stollewerk and Seyfarth, 2008)).

Research objectives

The goal of my thesis work is to identify the progenitors of sensory organs during *Parhyale* limb regeneration, and to determine the degree of commitment of these progenitors. To achieve this, I aim to live image the regenerating limbs continuously from amputation until the end of regeneration, to identify the nuclei of the sensory organ cells in the last frames of the recordings, and to trace back the lineage of these nuclei. By back-tracing the cell lineage of the identified cells, I will be able to accurately identify their progenitors at every stage of regeneration. To determine the developmental potential of each progenitor, I will trace the cell lineages that derive from individual progenitor cells and identify whether these progenitors contribute to a single or multiple (different) cell types in the regenerated limbs.

Sensilla is a good model to address these questions for multiple reasons. First, they consist of multiple cell types. These cells are located within, or close to, the epidermis, which renders the accessible for imaging. Their cellular behaviors, such as the delamination of the internal cells of the sensilla from the epidermal layer, as well as the known molecular markers of sensilla cells could be used to identify these cells in new species. During *Parhyale* development and/or limb regeneration, sensilla cells could be derived from a single SOP, as in *Drosophila*, or a group of epidermal cells may acquire SOP fate simultaneously and give rise to different components of a sensillum, as in spiders. *Parhyale* sensilla could also be produced by one of these mechanisms during limb regeneration. In either case, it would be interesting to identify the origins of the SOPs and determine whether these can derive from any epidermal cell lineage or from a predetermined set of precursors. If multiple epidermal cells acquire the SOP fate to generate a sensillum, as in spiders, the question is whether these epidermal cells are derived from a single or multiple progenitors. The answers to these questions would help to identify whether the SOP fate is determined by cell lineage, or whether it depends on the position of the cells. In this context, my thesis work consists of three main parts.

First, I examine the distribution and diversity of sensory organs on *Parhyale* legs, and assess the degree to which these are faithfully restored after regeneration. Regenerated *Parhyale* legs are reported to have normal morphology in terms of their segmental morphology, but a detailed examination is lacking. I present a detailed comparison of regenerated and unamputated legs using scanning electron microscopy, to identify the distribution of setae before and after regeneration (chapter 2).

Next, I present the optimization and adaptation of long term live imaging of limb regeneration, to be able to track the progenitors of terminally differentiated cells

throughout the entire course of regeneration. Previous live recordings performed in the lab typically covered a time course of 2-3 days of leg regeneration. My aim is to cover the entire course, which can take up to two weeks, and to improve the live imaging method in order to avoid damage due to light exposure over long periods (chapter 3).

Finally, I focus on the identification of the sensory organ cells in the regenerated legs. For this purpose, I present two genetic approaches, CRISPR-mediated knock-ins and *cis*-regulatory element reporters (chapters 4 and 5, respectively) taking advantage of previously identified markers for sensory organs. Lastly, I present an antibody staining approach to identify the fate of cells in the regenerating legs after the course of live imaging (chapter 6).

2. Regeneration restores the diversity and the pattern of sensory organs in the legs of *Parhyale*

Abstract

Many animals have the ability to regenerate parts of their body that have been injured or amputated, including organs with complex architectures and multiple differentiated cell types. The regenerated organs resemble the injured organs that they replace, but it is usually not known whether they represent perfect or imprecise replicas of the original structures. Here we address this question in the regenerated limbs of the crustacean Parhyale hawaiensis, a genetically tractable system where it is possible to elucidate the cellular origins of regenerated tissues. We focus on the array of external sensory organs that decorate the distal part of Parhyale legs. These sensory organs represent complex markers that allow us to track the accuracy of regeneration in fine detail. We describe eight types of external sensory organs present on Parhyale legs, distributed in stereotypic positions or patterns on the surface of the leg. We find that regenerated legs carry the full diversity of external sensory organs, distributed precisely in their original locations. The numbers of regenerated sensory organs are indistinguishable from those found in size-matched un-amputated limbs. We conclude that regeneration faithfully restores the array of external sensory organs in the legs of Parhyale.

Introduction

The ability to regenerate varies widely among animals. On one extreme of the spectrum are animals that are able to regenerate any part of their body perfectly from a small body fragment, such as some species of planarians and hydrozoans (Morgan, 1901). On the other extreme are animals that are incapable of regenerating (e.g. nematodes) or whose capacities to regenerate are limited to physiological turnover of some tissues (e.g. skin or blood renewal in mammals). Between these extremes are a wide variety of animals whose capacities to regenerate are limited to specific organs, particular stages in their lifecycle, or whose ability regenerate is imperfect. Well known examples of the latter category are lizards, whose regenerated tails replace bony vertebrae by cartilage (Goss, 1969), cockroaches, whose regenerated legs have differences in nerves, tracheae and musculature (Kaars et al., 1984), or fish, whose regenerated heart tissue beats asynchronously with respect to the rest of the heart (González-Rosa et al., 2014). In many cases regeneration is known to produce organs that have a normal appearance, but it is difficult to assess whether regeneration is perfect down to the fine structural and cellular details.

From an evolutionary point of view, the mechanisms of regeneration need only produce an organ or tissue that restores normal function. So long as function is restored, there is no added evolutionary benefit for regeneration to be perfect.

Here we focus on leg regeneration in arthropods, where external chemosensory and mechanosensory organs composed of small groups of cells are clearly visible on the external surface of the leg. These organs have been studied extensively in the fruit fly *Drosophila*, where each sensory organ typically consists of 4 cells – including hair, socket, neuron and sheath cells – which arise through stereotypic divisions from a single precursor cell (reviewed in Hartenstein, 2005). External sensory organs come into contact with the environment through highly specialised structures that protrude from the surface of the cuticle in the form of bristles or setae. These structures are produced by a single cell, the hair (trichogen) cell, in *Drosophila*, but could arise from several cells in other species (Garm and L. Watling, 2013; Guse, 1983). These structures can take a variety of shapes and sizes, depending on the type of sensory organ and its functions in each species. Within a given species, however, the morphology, cellular composition and spatial distribution of each type of sensory organ are usually stereotypic and well conserved (Hartenstein, 2005). Thus, sensory organs can provide markers for assessing the accuracy of regeneration with very high (almost cellular) precision.

Among arthropods, malacostracan crustaceans retain the ability to regenerate their limbs throughout their lifetime (Charmantier-Daures and Vernet, 2004; Maruzzo and Bortolin, 2013). We recently introduced the amphipod crustacean *Parhyale hawaiensis* as an experimental system for studying regeneration, taking advantage of genetic approaches (transgenesis, CRISPR-mediated gene editing) and genomic resources available in this species (Kao et al., 2016; Martin et al., 2016; Pavlopoulos and Averof, 2005). The ability to perform live imaging and to track cells during the course of regeneration (Alwes et al., 2016) makes *Parhyale* an attractive system in which we can study the mechanisms of regeneration at single-cell resolution. In this context, the external sensory organs of *Parhyale* limbs can serve as markers to investigate the fidelity of regeneration. Here, we describe the diversity and the pattern of sensory organs found in the distal part of *Parhyale* limbs. We assess the extent to which these structures are faithfully recovered and their functions restored following leg regeneration.

Results and discussion

The diversity of setae on Parhyale limbs

We have used scanning electron microscopy (SEM) to survey the surface of *Parhyale* limbs, focusing on the three most distal podomeres (the dactylus, propodus and carpus) of the T4 and T5 pereopods. These two limbs show almost identical patterns of setae (see below). The cuticle of the limbs has a polygonal pattern (Figure 2.2a), which reflects the architecture of underlying epithelial cells (Havemann et al., 2008), (Dillaman et al., 2013). On the limb surface we observed several types of setae, which we categorised as follows based on previous studies (Garm, 2004; Garm and L. Watling, 2013; Les Watling, 1989):

– Lamellate setae: setae consisting of a smooth shaft bearing a series of lamellae towards the tip (figure 2.1a,b). We identified two variants of lamellate setae: type-1 have a wider base and a shorter shaft with a terminal pore (figure 2.1a), type-2 tend to have a more slender base, a longer shaft and no pore (figure 2.1b).
– Plumose setae: setae with a long shaft, bearing two opposed rows of long setules, which give it a feathery appearance (figure 2.1c). There are no visible pores.

– Twin setae: composites of cuspidate and lamellate setae (figure 2.1d). The cuspidate-like main shaft bifurcates into a branch that resembles the tip of a typical lamellate seta with a terminal pore.

– Curved setae: simple setae with a long twisting shaft, bearing a pore approximately 2/3 along the length of the shaft (figure 2.1e).

– Hooked setae: simple setae with a long thin shaft tapering gradually towards the apex, a hooked shape, and a series of fine nicks prior to the tapering distal region. The shaft has a terminal pore (figure 2.1f).

– Cuspidate setae: relatively short and stout setae, bearing longitudinal ridges and no pore (figure 2.1g).

– Microsetae: very small setae bearing a terminal pore covered by a hood. One side of the shaft has a lamellated appearance. Microsetae are associated with characteristic dimples on the cuticular surface (figure 2.1h).

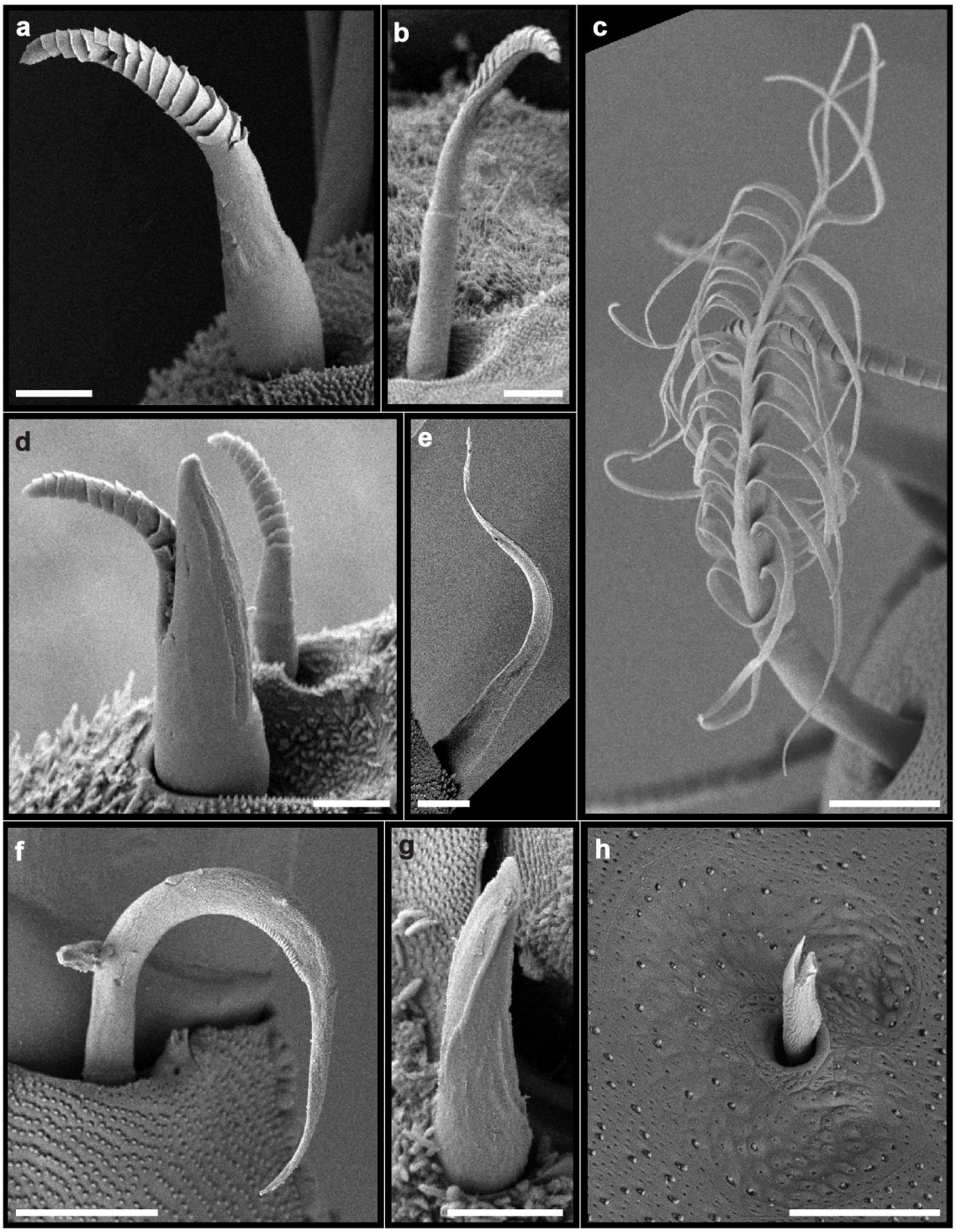


Figure 2.1: Setal types found on Parhyale limbs.
(a) Type-1 and (b) type-2 lamellate setae, (c) plumose seta, (d) twin seta, (e) curved seta, (f) hooked seta with some dirt on the shaft, (g) cuspidate seta, and (h) microseta with a specific design of the cuticle surrounding it. The scale bars are 5µm.

In addition to these sensory structures, we observe that the surface of limbs bears numerous pores, located at the junctions of epidermal cells (figure 2.2). These pores do not appear to be associated with axons (figure 2.7), we therefore conclude that they are unlikely to have a sensory function.

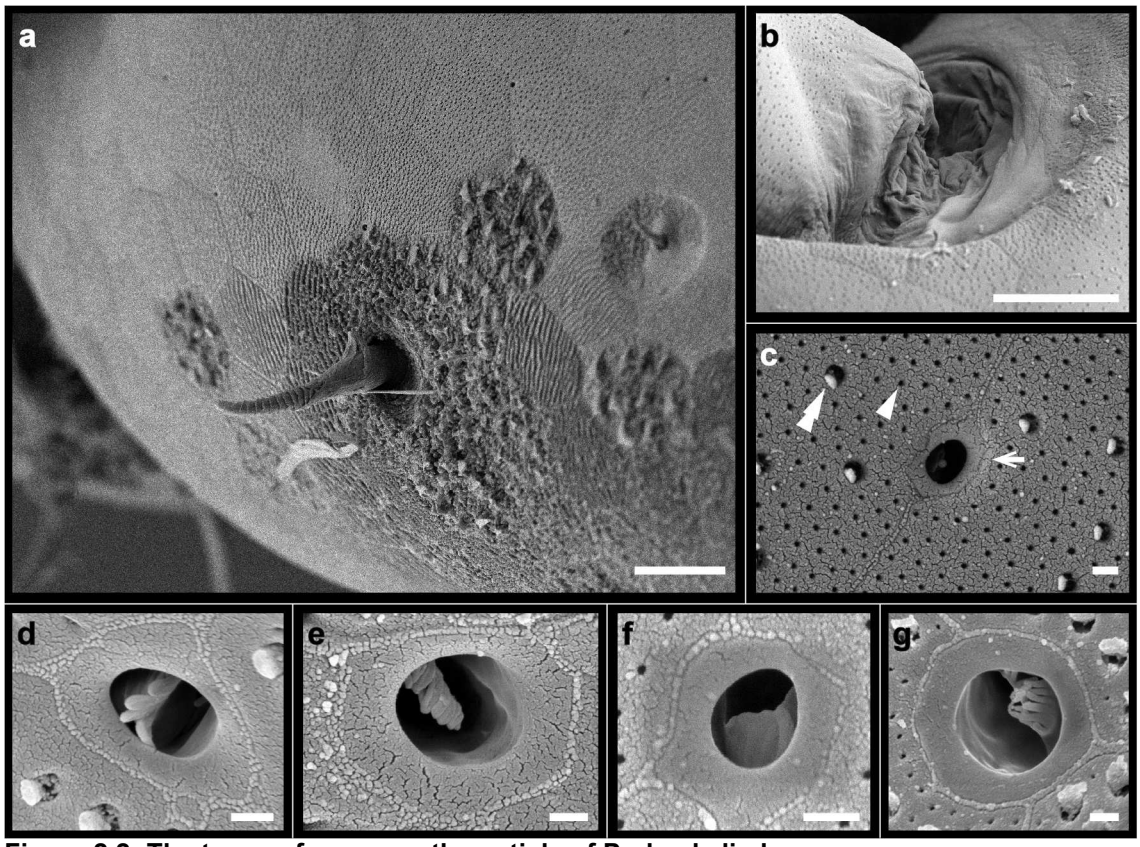


Figure 2.2: The types of pores on the cuticle of Parhyale limbs.
(a) The cuticle is organised in a polygonal pattern. The pores are covering all the surface except the ventral side where the polygons have a different architecture. (b) The arthrodial membrane covering the joint of two podomeres has no pores. (c) There are three different types of pores: small, medium, and large; labelled with arrowhead, double arrowhead and arrow respectively. (d-g) Large pores have different microstructures inside indicating different subtypes. The scale bars are 10μm for a-c, and 250nm for c-g.

The distribution of setae on Parhyale limbs

We find that each type of seta has a well-defined distribution on the distal part of Parhyale T4 and T5 limbs: either in stereotypic positions, or in arrays of several setae arranged in specific patterns (figure 2.3). The only exception to this stereotypic arrangement are the microsetae, which are well spaced on the surface of the cuticle but without an apparent conserved pattern (figure 2.3h).

The most distal podomere, the dactylus, bears three distinctive large setae placed in specific locations: a hooked seta located at the base of the terminal claw, a large curved seta on the ventral side, and a plumose seta on the dorsal side (figure 2). This distribution was invariable in all the limbs we inspected (n=22).

In the next two podomeres, the propodus and the carpus, most of the large setae are organised in four distinct groups. The first group, which we name the 'crown', consists of a row of type-2 lamellate setae, located on the dorsal margin at the distal end of each podomere (figure 2.2f-f'). The crown is present in every individual, but consists of a variable number of setae; we found that on average the propodus has 3.3

setae (s.d. 1.3, n=23) and the carpus has 3.0 setae (s.d. 0.9, n=23) on the crown (figure 2.4a and figure 2.6a).

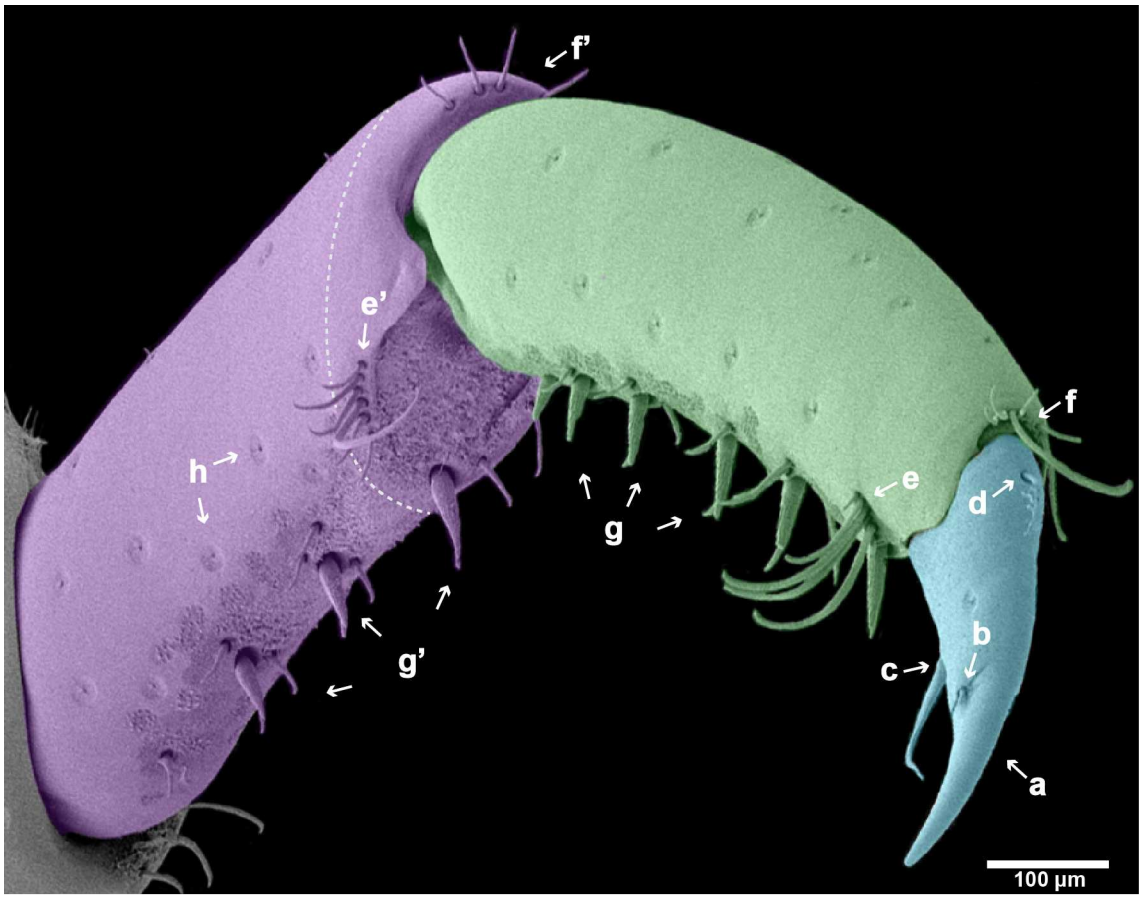


Figure 2.3: Distribution patterns of the setae.
The SEM image of the first walking leg of Parhyale (T4) focusing on three distal-most podomeres: dactylus (turquiose), propodus (green), and carpus (purple). The dactylus bears a terminal claw (a) and three distinctive setae: hooked (b), curved (c), and plumose (d) setae. The propodus and carpus has the crown groups (f, f') and the comb groups (e, e') both consist of a tight row of type-2 lamellate seta. The last group on the propodus and carpus is the row group: an array of small clusters of twin and type-1 lamellate setae on the ventral side of the podomeres (g, g'). Apart from the setal groups, there are several microsetae covering the posterior surface of the limb (h). The dashed line is marking the amputation plane.

The second and third groups, which we name 'combs', are tight rows of type-2 lamellate setae located in the distal part of the propodus and carpus. One comb is found consistently on the posterior face of the propodus and the carpus (figure 2.3e-e'). A second comb is usually found on the anterior face of the propodus (figure S2.1), but occasionally this is reduced to a single seta. On the carpus, single type-2 lamellate setae are found in place of the anterior comb. We find that each comb consists of 1 to 6 setae, with an average of 3.6 setae (s.d. 1.1, n=22) at the posterior of the propodus, 2.6 setae (s.d. 1.0, n=22) at the anterior of the propodus, and 3.0 (s.d. 0.9, n=23) at the posterior of the carpus (figure 2.4b and figure 2.6b).

The fourth group, which we name the 'ventral' setae, consists of an array of setae that are distributed with a regular spacing along the ventral side of the propodus

and carpus (figure 2.3 g-g'). The number of elements in these ventral arrays of setae correlates with the length of the podomere, suggesting that new elements are added to this pattern as the limbs grow during the lifetime of Parhyale (figure 2.5). Each element of the ventral array is made up exclusively of twin setae and type-1 lamellate setae, but there is considerable variation in the number of setae per element (figure 2.4c, figure 2.6c). In the propodus, most elements consist of a twin seta surrounded by type-1 lamellate setae; in the carpus they often consist exclusively of type-1 lamellate setae (see supp. Table S1). As we discuss later, type-1 lamellate setae may develop into twin setae, so this variation could reflect different stages in the maturation of these elements, as new ones are added during growth. We found that on average the propodus has 3.26 sets of ventral setae along its length (s.d. 0.86, n=22) and the carpus has 2.82 (s.d. 0.96, n=23) (figure 2.4b and figure 2.6b).

In addition to these groups, the propodus bears a single cuspidate seta ventrally, near the joint, on the most distal end of the podomere (n=23, supp. figure S2.1).

The distal parts of T4 and T5 legs carry almost identical patterns of setae (see suppl. Table S1), in subsequent analyses we have therefore combined the data from these limbs.

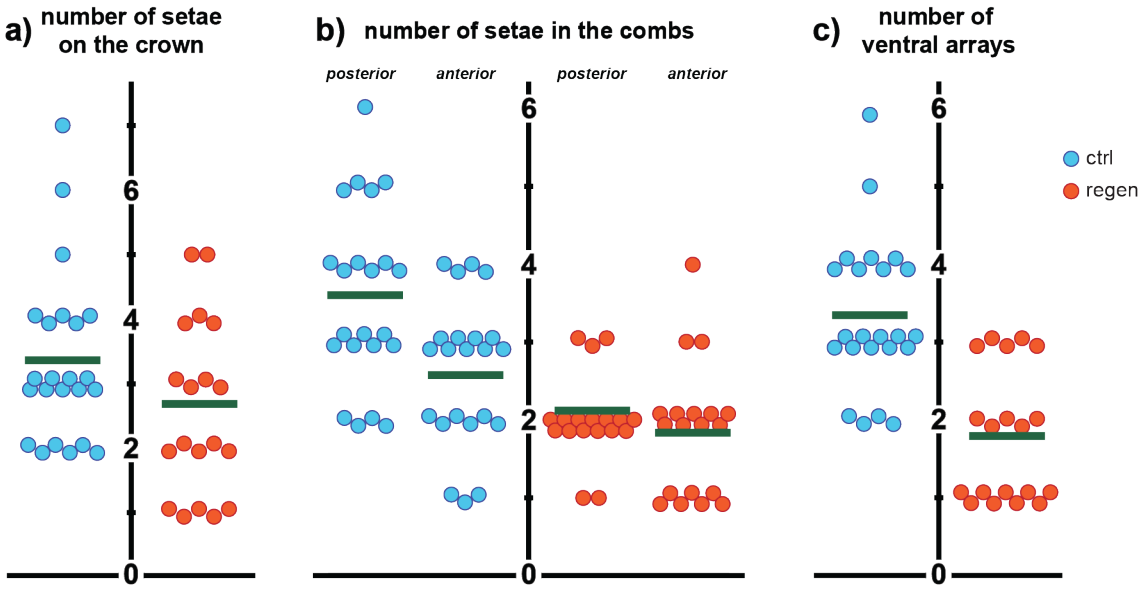


Figure 2.4: Quantification of setae on the propodus of uninjured and regenerated legs. Quantification of the number of seta in the (a) crown and (b) comb groups, and (c) number of arrays on the ventral side of the propodus in unamputated and regenerated legs. Every point represents the seta on a single propodus (blue for control, orange for regenerated legs). Green bars correspond to the mean values.

The pattern of sensory organs is fully restored in regenerated Parhyale limbs

Following the identification of these setae, we investigated to what extent these structures are faithfully regenerated following amputations at the carpus of T4 and T5 limbs (data in suppl. Table S2.1). In order to minimise genetic, environmental or age effects on the number and patterns of setae, we measured the setal patterns of control (unamputated) and regenerated limbs simultaneously in the same cohort of individuals: the observations on regenerated limbs were carried out on T4 and T5 limbs that were amputated and allowed to regenerate on one side of the animal, while the observations on unamputated limbs (described in the section above) we carried out on the contralateral limbs of the same individuals.

On the dactylus, we find that the stereotypic hooked, curved and plumose setae described earlier consistently regenerate in their original positions, with no exceptions (n=16). In the propodus and the carpus, we find that all the patterns of sensory organs that we described – the crown, the combs and the ventral setae – are also regenerated, but the number of setae in each pattern may vary depending on the size of the regenerated podomere.

Thus, the propodus of regenerated T4 and T5 limbs bears a crown (with 2.67 type-2 lamellate setae on average, s.d. 1.33, n=23), anterior comb (with 1.57 type-2 lamellate setae on average, s.d. 0.65, n=22), posterior comb (with 2.07 type-2 lamellate setae on average, s.d. 0.62, n=22), ventral setae (4.35 sets of twin and/or type-1 lamellate setae on average, s.d. 0.97, n=23), and the distally-located cuspidate seta (found in 17 out of 18 cases) (figure 2.4).

Regenerated limbs are smaller than their unamputated contralateral limbs in the first molt following amputation (Alwes et al., 2016) and gradually recover their full size during subsequent molts. We therefore reasoned that the smaller number of setae observed in a regenerated propodus might reflect the smaller size of the field in which they develop. To test this idea, we quantified the number of sets of setae found in the ventral array of the propodus in relation to the length of this podomere, on the images obtained by scanning electron microscopy. To add to that limited dataset, we imaged additional unamputated and regenerated T4 or T5 limbs on a laser scanning confocal microscope, exploiting the cuticle's autofluorescence to observe the sensory organs in the ventral arrays (figure 2.5). In the combined dataset of SEM and confocal data (n=29), we find that the length of the propodus is ~18% smaller in the regenerated limbs compared with their unamputated contralateral limbs. The propodus of regenerated limbs harbour the similar numbers of setae units on their ventral array as size-matched unamputated limbs of 350-450 μm in length. We need to sample

unamputated limbs with smaller propodus (<350 µm) and regenerated limbs with a larger propodus (>450 µm) to extend this comparison over a wider range of limb sizes.. These data suggest that the regenerated limbs bear similar numbers of setae as size-matched unamputated limbs. However, we need to image and quantify setae from more unamputated limbs, especially smaller ones, to clarify this correlation.

In our experiments the site of amputation is the distal part of the carpus (see figure 2). Thus, distal elements in the carpus (such as the crown and the combs) are removed by the cut while proximal elements are retained. We find that this partly regenerated podomere bears all the patterns of sensory organs found in T4 and T5 limbs, including a crown (with 2.41 type-2 lamellate setae on average, s.d. 1.06, n=17; figure 5a), anterior comb (with 1.21 type-2 lamellate setae on average, s.d. 0.43, n=18), posterior comb (with 3.28 type-2 lamellate setae on average, s.d. 0.99, n=17), and ventral setae (2.50 sets of twin and/or type-1 lamellate setae on average, s.d. 0.70, n=18) with a composition that largely resembles the composition found in unamputated limbs (figure 2.6).

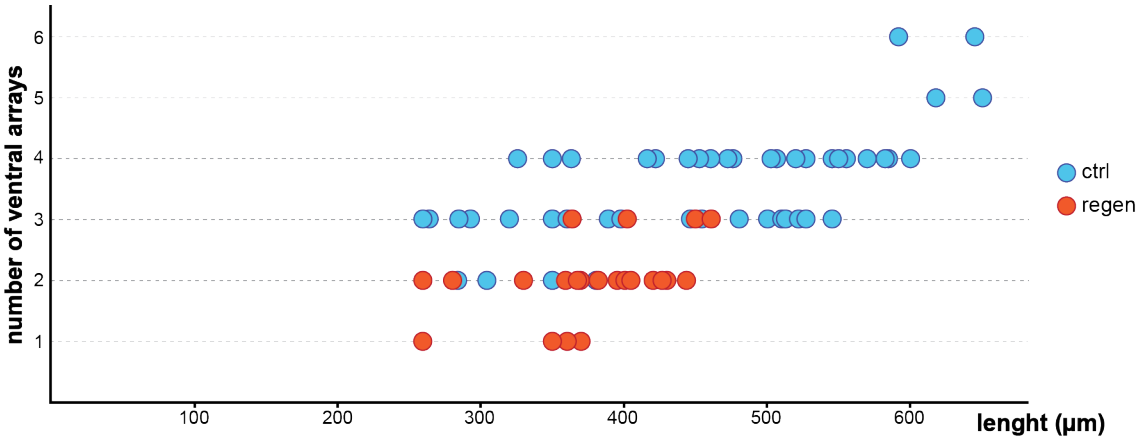


Figure 2.5: Correlation between the number of ventral arrays and the length of the podomere for propodus.
Scatter plot of the size of the propodus versus the number of ventral arrays of unamputated control (blue circles) and regenerated (orange circles) legs.

In addition to the recovery of all types of large setae, we find that the microsetae and the pores which are distributed on the limb surface are also restored following limb regeneration. We did not quantify the density and distribution of microsetae and pores.

Putative sensory functions of different types of setae

Stainings using an antibody for acetylated-tubulin, which labels nerve axons, as well as transgenic animals expressing lyn-tdTomato (Alwes et al., 2016), which allows us to visualise some neurons, show that all the types of setae that I described are innervated, as expected of sensory organs (figure 2.7).

Based on what we know from other arthropods, we expect these external sensory organs to have mechanosensory and/or chemosensory functions. Although it is not easy to identify the precise sensory functions of each type of seta, we could speculate on their functions based on their morphology and some previously defined functions of homologous setae in other crustaceans. To begin with, the presence of a pore on the apex of a seta is indicative of a chemosensory function (Brandt, 1988; Garm and L. Watling, 2013), therefore, we can consider type-1 lamellate setae, curved setae, and microsetae as putative chemosensory organs.

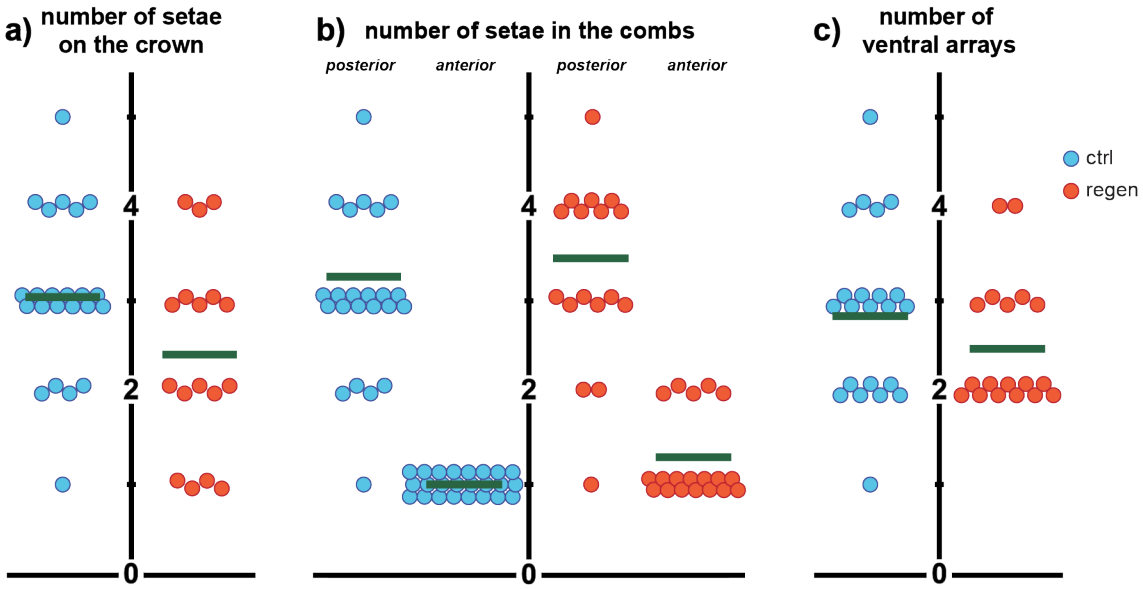


Figure 2.6: Quantification of the setae on the carpus of uninjured and regenerated legs. Quantification of the number of seta in the (a) crown and (b) comb groups, and (c) number of arrays on the ventral side of the carpus in unamputated and regenerated legs. Every point represents the seta on a single carpus (blue for control, orange for regenerated legs). Green bars correspond to the mean values.

Conversely, setae that bear setules along their shaft, like plumose setae, have never been shown to be chemosensitive (Garm and L. Watling, 2013). They either have non-sensory functions, such as generating stroking power for swimming or used as combs for moving food into the mouth, or they can have mechanosensory functions. In *Parhyale* pereopods, it is unlikely that they would have a structural function since there is only a single plumose seta positioned on the distal part of the leg. However, they could serve as mechanosensory organs, since two rows of long setules on the shaft makes them very sensitive for waterborne vibrations and might also provide information on the direction of the signal (Fields et al., 2002; Vedel, 1985).

The comb and crown groups consist of rows of type-2 lamellate setae of different length. With such organisation, they could provide 3D sensation of mechanical stimuli collectively, as rodent whiskers do (Carvell and Simons, 1990).

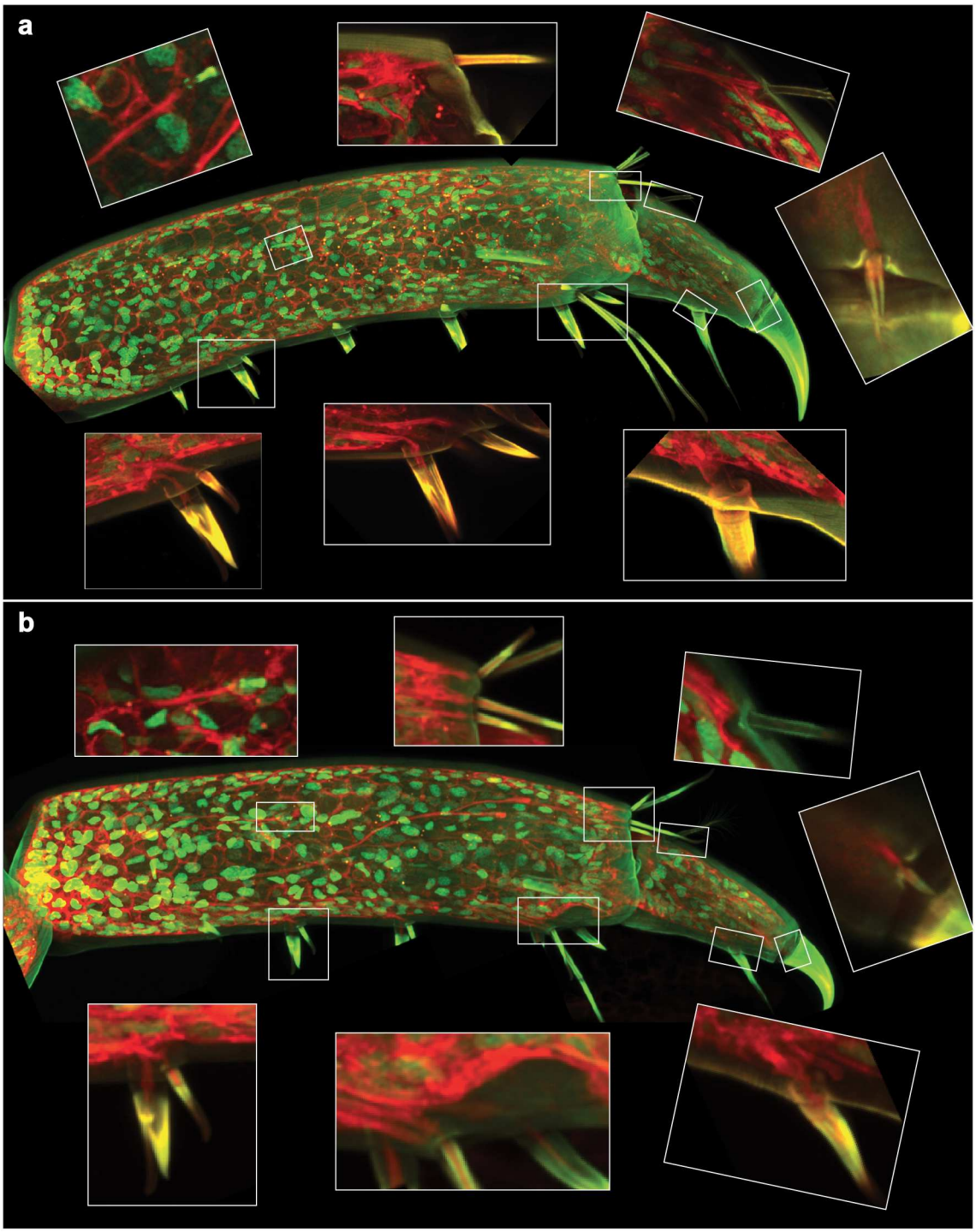


Figure 2.7: All setae of both uninjured and regenerated limbs are innervated.
(a) An uninjured and (b) a regenerated T5 limb from the same individual carrying the PhHS-lyn::tdTomato-2A-H2B::EGFP transgene, which expresses lyn-tagged tdTomato (marking cell outlines, in red) and histone-EGFP (marking cell nuclei, in green). The tdTomato label allows neurons to be tracked from the setae on the surface of the epidermis to the main nerve of the limb. In the insets, clockwise from top left: microsetae, type-2 lamellate seta, plumose seta, hooked seta, curved seta, cuspidate seta, lamellate-1 and twin setae, and type-2 lamellate setae. The axons of the neurons cannot be seen in this focal plane.

The dactylus has no muscles and possibly also no chordotonal organs. The cuspidate seta beneath the dactylus-propodus junction might be important to perceive the bending of the dactylus.

Twin setae have two components, resembling the cuspidate and lamellate setae. The lamellate part of twin setae is likely to have a function in chemoreception, as suggested by the terminal pore on the apex, and the cuspidate part could have mechanosensory functions. Apart from putative mechanosensory functions, the sturdy cuspidate setae might also perform mechanical functions, such as clinging to the substrate or manipulating food (Garm, 2004). Considering that *Parhyale* spends most of its life clinging on a substrate, the robust twin setae on the ventral side of the legs could be used for attachment to the substrate.

Recovery of sensory function after regeneration

All the setae of regenerated legs are innervated, similar to unamputated legs (figure 2.7), which suggests that, besides recovering their detailed morphology, regenerated *Parhyale* limbs may have also recovered their sensory functions. To test this hypothesis we developed a simple assay for the mechanosensory functions of limb setae. We found that touching the ventral row or comb setae of T4 or T5 limbs with a solid object evokes stereotypical escape response (Video 2.1). The assay is robust: 95% (confidence interval (CI) 90-98%) of stimulations to the carpus comb setae resulted in a response in T4 and T5 limbs of immobilised animals (n=22, figure 2.8, table S2.2). After amputating one of these limbs, either T4 or T5, at the carpus-propodus junction, and the assay was repeated 6 days post amputation (dpa): the response rate of the amputated limbs dropped to 8% (CI 3-25%), whereas the unamputated T4 or T5 control limbs in the same animals remained highly responsive, with an average of 98% (CI 92-100) response rate. The unresponsiveness of the amputated limbs was expected since the sensory organs at the distal part of the limb stump are thought to de-differentiate and become disconnected from the setae (figure 2.8a, see chapter 3 for details). Therefore the setae located on the distal part of the stump are not expected to respond to the stimuli at the later stages of regeneration. Once limb regeneration was complete and the animals molted, the response rate of both control and regenerated limbs were the same, 88% (CI 74-95%), which is comparable to the response rate of before amputation. These results suggest that the mechanosensory function is fully restored during leg regeneration.

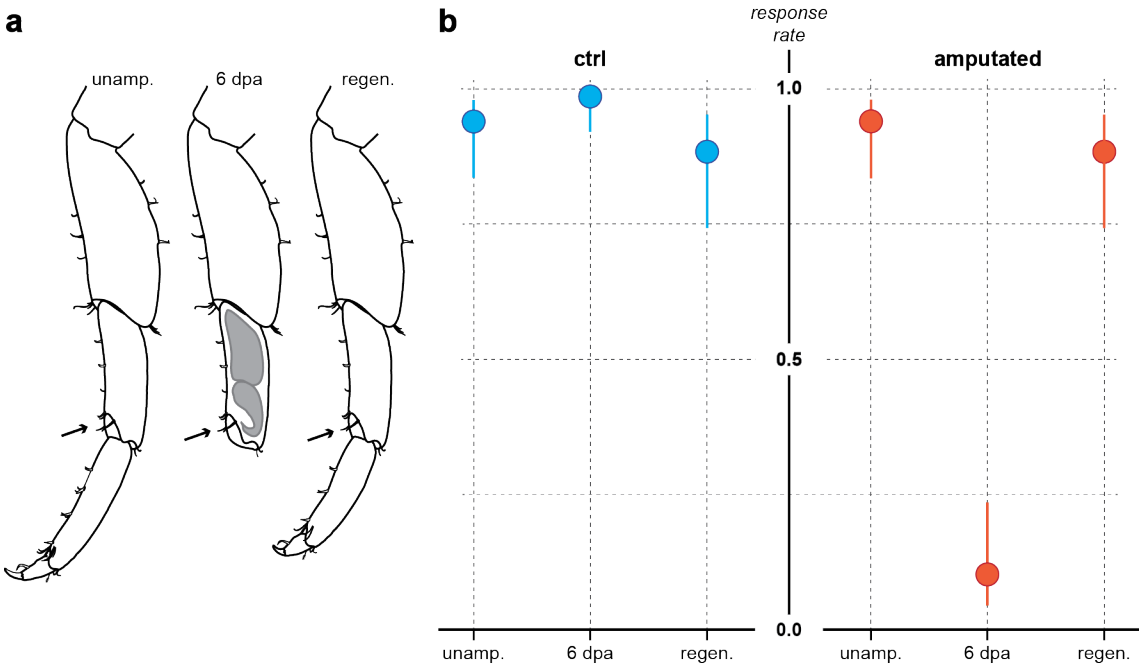


Figure 2.8: Recovery of mechanosensory function after regeneration.
(a) The comb setal groups of carpus (black arrows) of T4 and T5 legs from the same animals were stimulated before amputation, 6 days after amputation, and after regeneration. (b) The response rate after stimulation was measured in unamputated control legs (blue circles) and amputated legs (orange circles). The circles represent the mean response rate and the bars indicate standard errors.

Transformation of type-2 lamellate setae into twin setae

The ventral row units consist of two types of setae, twin and type-2 lamellate setae. In unamputated legs there is always a central twin seta present in each ventral row unit of the propodus, and in some cases twin setae also appear laterally, in the positions usually occupied by lamellate setae. This is seen more frequently in posterior units than in anterior ones, and may reflect the degree or maturity of these units (as defined by their size and number of setae per unit). Moreover, in newly regenerated legs, the central twin seta is sometimes replaced by a lamellate seta. These observations suggest that there may be a temporal transition from lamellate setae to twin setae as these elements mature. A closer look at the morphology of twin setae supports this hypothesis: we find that some of the twin seta on the regenerated legs are smaller than the ones on control limbs. In these setae the lamellate part of the seta has a normal size whereas the cuspidate part is considerably smaller (figure 2.9). Putting these two observations together, we suggest that some of the type-2 lamellate setae turn into twin setae as the limb develops; the smaller setae observed on the regenerated limbs could be the intermediate forms.

Homologues of twin seta have been previously identified in isopods (Brandt, 1988) and hyalellid amphipods (Zimmer et al., 2009). But neither of these studies provides insights about development of these structures. A closer investigation of twin

seta both on developing and regenerating Parhyale limbs might help to investigate this process in detail.

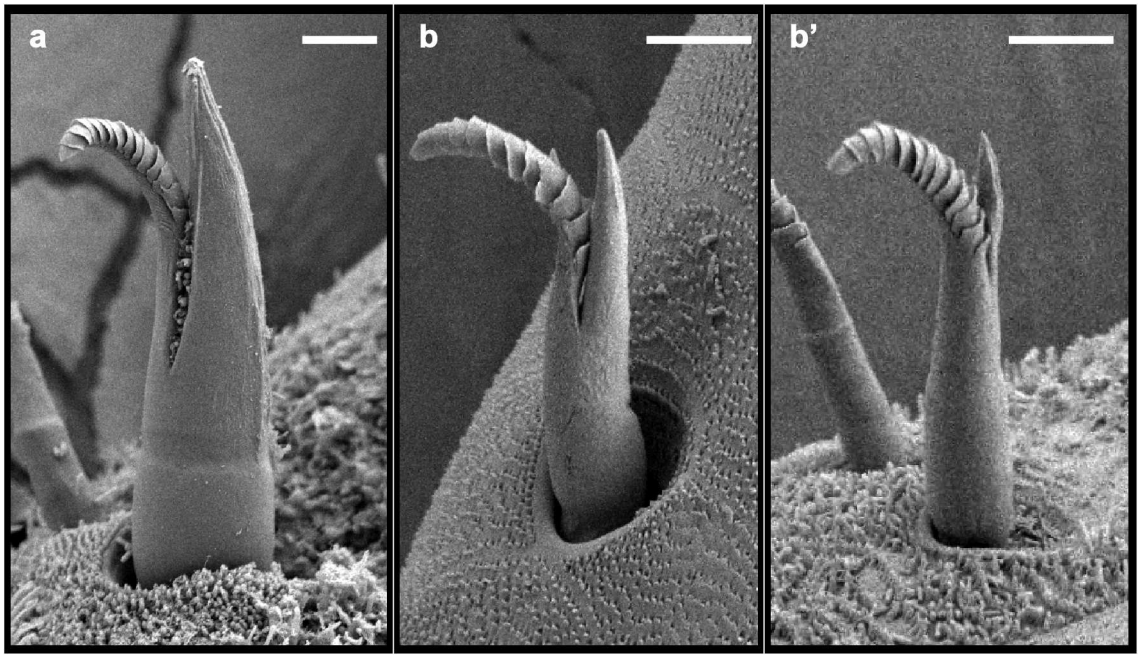


Figure 2.9: Morphology of twin setae of unamputated and regenerated legs.
SEM images of twin seta from (a) unamputated and (b-b') regenerated limbs. The width of regenerated twin seta is smaller than the mature ones, and their cuspidate part is underdeveloped. The scale bars are 5μm.

Conclusions

The morphology and the distribution of setae on limbs are distinctive features of crustacean species and have been widely used as taxonomic characters. In this study, we identified eight different types of setae on the distal part of *Parhyale* pereopods by SEM. Most of these setae are arranged on the legs in stereotypic and well-defined patterns, which present excellent microanatomical markers to assess the fidelity of regeneration. Newly regenerated limbs are smaller than their unamputated counterparts, most likely as a result of regeneration taking place in the leftover limb stump where space is limited, as described above. Yet, they harbor all of the eight types of setae, in their specific patterns, but in reduced numbers compared with unamputated limbs. Our data suggests that the smaller number of setae found on regenerated legs is linked with the reduction of limb size.

The setae we identified are all innervated, both in unamputated and regenerated legs, supporting the notion that they all function as sensory organs. Using a simple behavioral assay, we showed that regeneration restores the sensory functions of at least one type setae.

Our data shows that the leg regeneration is capable of restoring the morphology and the function of legs. The peripheral sensory organs perceive the sensory stimuli and transmit them to CNS thanks to synchronised work of several cell types, therefore the recovery of sensory function during regeneration confirms that the cellular diversity required for this function is also recovered. This might be linked to the fact that the *Parhyale* legs are important sensory units and regeneration of this function is crucial for these animals. Alternatively, the precise regeneration of the sensilla could be the product of an embryonic developmental programme, generating a highly stereotypic pattern of sensilla on the limb epidermis that is precisely replayed during adult limb regeneration. Independent of the underlying cause, the regenerated legs are near-perfect copies of the original ones, both morphologically and functionally.

Materials and Methods

Parhyale culture

A wild type *Parhyale* stock kept for many generations in the laboratory was raised as described previously (Browne et al., 2005). Prior to imaging, animals were kept individually for 3 months in the dark to prevent the build-up of algae. Animals were anaesthetised using 0.02% clove oil.

Scanning Electron Microscopy (SEM)

Adult animals of various sizes were selected and the T4 and T5 limbs on their right side were amputated at the distal part of the carpus (see dashed line, figure 2). The limbs on the contralateral size were left unamputated as controls. Following the first molt after amputation, the animals were anaesthetised, washed in artificial sea water (ASW, specific gravity 1.02), fixed for 2 hours in 1% glutaraldehyde (Electron Microscopy Sciences #16300) in ASW, then washed in ASW and re-fixed for 2 hours in 1% osmium tetroxide (Electron Microscopy Sciences #19150) in ASW. Animals were then washed in ASW for 1 hour. All steps were carried out at room temperature.

The regenerated T4 and T5 and their contralateral unamputated limbs were carefully removed from the fixed individuals, dehydrated by washing in increasing concentrations of ethanol (in ASW) and stored in 90% ethanol. The samples were handled on the proximal parts of the limbs in order not to damage the distal structures. Subsequently they were washed 3 times in absolute ethanol (Merck, 1009832500), subjected to critical point drying in a Leica EM CPD300 critical point dryer, mounted on specimen holders, and coated with gold in a Polaron SC7640 sputter coater. Imaging was performed on a JEOL 6700F scanning electron microscope at the electron microscopy facility of the Stazione Zoologica Anton Dohrn, Naples.

Laser Scanning Confocal Microscopy

Unamputated animals were anaesthetized and fixed at room temperature for 15 minutes in 3.6% formaldehyde (VWR, 20909.290) in ASW. After washing in ASW, the T4 and T5 limbs were dissected and fixed for an additional 15 minutes in 3.6% formaldehyde in ASW. Samples were then washed with ASW for 1 hour, washed in 50% glycerol in phosphate buffered saline (PBS) for 30 minutes, transferred to 70% glycerol and mounted. The cuticle autofluorescence was observed under Zeiss LSM 800 laser scanning confocal microscope, using a 488 nm excitation laser and a 400-730 nm detection window.

Immunostaining

Large males were amputated at the distal part of the carpus of T4 and T5 limbs, as described earlier. After the first molt following regeneration they were anesthetized and fixed at room temperature for 10 minutes in 3.6% formaldehyde in ASW. The carpus and propodus podomeres of regenerated and unamputated control limbs were dissected and cut in half to improve antibody penetration. The samples were re-fixed for 15 minutes in 3.6% formaldehyde in PBS, washed for 1 hour in PBS with 0.1% Triton X-100 (PTx), and incubated for 1 hour at room temperature in PBS with 0.1% Triton X-100, 0.1% sodium deoxycholate, 5% bovine serum albumin and 0.5% normal goat serum (PAXD1). The samples were then incubated for 3 days at 4°C, with 1:1000 dilution of mouse monoclonal 6-11B-1 antibody for acetylated tubulin (Sigma T6793, RRID: AB_477585) in PAXD1, and washed overnight with PTx at 4°C. The samples were incubated with the secondary antibody (1:1000 dilution of anti-mouse IgG Alexa 488 (Life Technologies A11001, RRID: AB_2534069) in PAXD1 for 3 days at 4°C. Then the samples were washed overnight in PTx at 4°C and mounted in Vectashield mounting medium (Vector Labs H-1000). The samples were imaged on a Zeiss LSM 800 laser scanning confocal microscope.

Testing mechanosensory function

Adult males (n=22) were immobilised to standard 60mm Petri dishes from their body, T4, and T5 limbs with a surgical glue (2-octyl cyanoacrylate, Dermabond) as described by (Alwes et al., 2016). The day after gluing, petri dishes were carried under the stereoscope individually, and after waiting 15 minutes the comb groups on the carpus of T4 and T5 limbs were stimulated twice with a fine tungsten needle and their responses were recorded (supp. table S2.2). The animals were left untouched for 15 minutes in between two stimulations, to calm them down. After the experiment, either the T4 (n=10) or T5 (n=12) limb of each animal was amputated from the carpus-propodus junction. The stimulation experiment was repeated on the same animals 6 days after the amputation, stimulating the comb groups of the carpus of both the amputated and the control limbs, twice with 15 min intervals. After the second round of experiments, the animals were left to molt, and one day post molting they were glued again on the Petri dishes. One day after they are glued, the experiment is repeated on the carpus comb groups of unamputated control and regenerated limbs, twice per limb with 15 minutes intervals.

The statistical analysis of the results were performed with R, using generalised linear mixed models. Initial analysis revealed that experiments performed on T4 and T5

limbs are identical (either they are considered as random or fixed effects), therefore the data coming from either limbs were pooled together for further analysis. Following analysis revealed that treating the two replicas of each experiment as separate data points or taking an average of them does not affect the results. As a result, each data point was taken individually as double the amount of animals were stimulated once (n=44). The source codes and the results of the analysis can be found in the supplementary material.

Acknowledgements

We thank the electron microscopy facility of the Stazione Zoologica Anton Dohrn in Naples for hosting us to perform scanning electron microscopy, especially Franco Iamunno for expert technical support, and the Centre Technologique des Microstructures of the Université Claude Bernard Lyon 1 for their help in sample preparation. We thank Marie Semon for performing the statistical analyses of the behavioral experiments. This work was supported by a doctoral fellowship from Boehringer Ingelheim Fonds and by the European Research Council, under the European Union Horizon 2020 program (grant ERC-2015-AdG #694918).

Author contributions

CC and MA conceived the study; CC performed and analysed the experiments; CC and MA wrote the manuscript

Supplementary Information

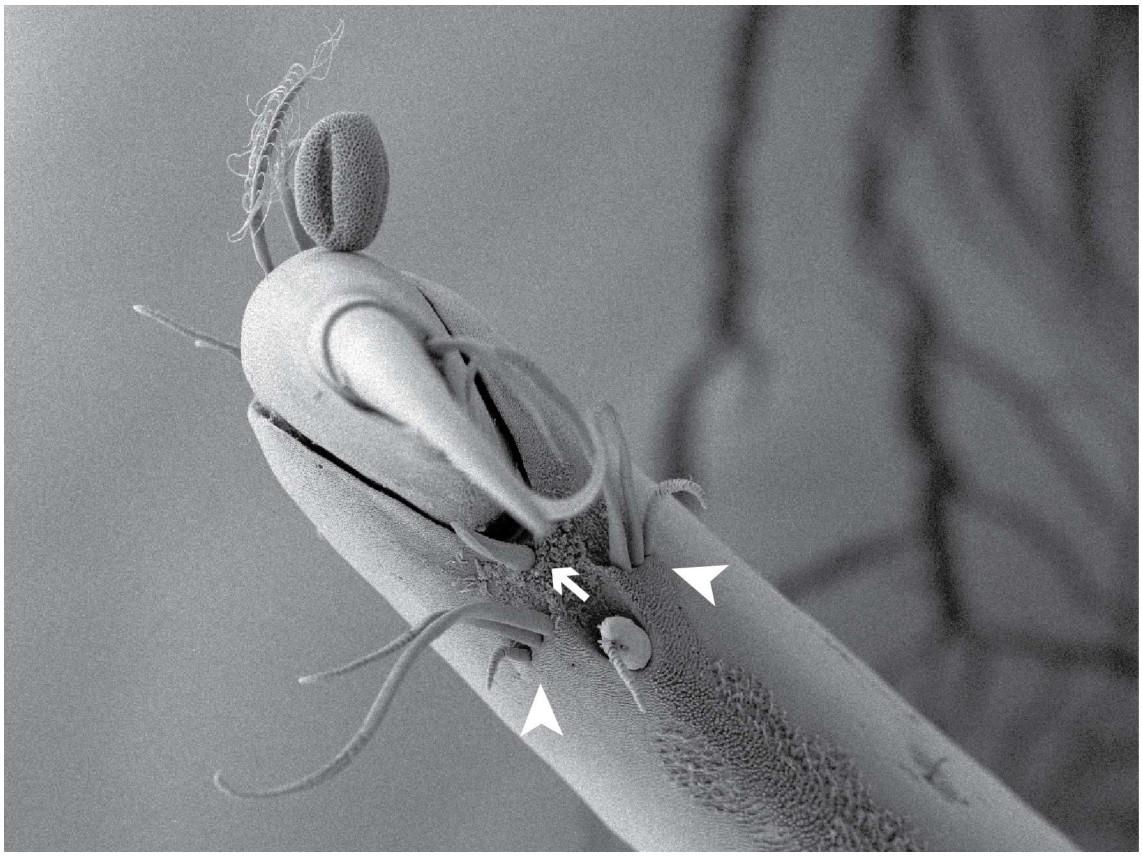


Figure S2.1: Ventral view of the propodus podomere.
There are two comb groups (white arrowheads) on the ventral-distal part of the propodus and a single cuspidate seta (white arrow) situated in between them.

Table S2.1: Quantification of the seta of regenerated and uninjured legs imaged with SEM.

				DACTYLUS			PROPODUS															CARPUS											
								ventral-distal region					setal composition of the ventral arrays										ventral-distal region			setal composition of the ventral arrays									
			Sample ID	curved seta	hooked seta	plumose seta	podomere lenght (µm)	# setae in the anterior comb	central cuspidate seta	# setae in the posterior comb	# of setae on the crown	1st group	2nd group	3rd group	4th group	5th group	6th group	7th group	total # of groups in the ventral arrays	total # of setae on the ventral arrays	total # of setae on the podomere	# of setae in the anterior comb	# of setae in the posterior comb	# of setae on the crown	1st group	2nd group	3rd group	4th group	5th group	6th group	total # of groups in the ventral arrays	total # of setae in the ventral arrays	total # of setae on the podomere
Ctrl	Male	T4	5.1.1	p	p	p	600	6	p	4	3	1 twin	3 lam + 1 twin	2 lam + 1 twin	2 lam + 1 twin	1 lam + 1 twin	np	np	5	11	24	4	1	4	1 twin + 3 lam	1 twin	2 lam + 1 twin	2 lam + 1 twin	np	np	4	10	20
			5.1.2	p	p	p	650	5	p	4	7	1 twin	3 lam + 1 twin	2 lam + 1 twin	1 twin	2 lam + 1 twin	1 twin	np	6	13	29	3	1	4	3 lam	1 lam	2 lam + 1 twin	2 lam	1 lam + 1 twin	1 lam	6	12	20
			5.1.3	p	p	p	527	4	p	4	4	1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	8	20	2	1	3	3 lam	1 lam	2 lam	2 lam	np	np	4	8	14
			5.1.4	p	p	p	350	3	p	3	3	1 twin	3 lam + 1 twin	2 lam + 1 twin	np	np	np	np	3	8	17	2	1	3	1 lam	1 lam	1 lam	np	np	np	3	3	9
			5.1.5	p	p	p	527	4	p	4	3	1 twin	3 lam + 1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	5	12	23	2	1	3	2 lam	1 lam	2 lam	3 lam	1 lam	np	5	9	15
			5.1.6	p	p	p	585	3	p	3	3	1 twin	1 lam + 1 twin	3 lam + 1 twin	2 lam + 1 twin	1 lam + 1 twin	np	np	5	12	21	1	1	3	3 lam	1 twin	2 lam	3 lam	np	np	4	9	14
			5.1.7	p	p	p	550	3	p	3	3	2 twin	2 lam + 1 twin	2 lam + 1 twin	2 lam + 1 twin	3 twin	np	np	5	14	23	3	1	4	1 twin + 1 lam	1 twin	2 lam + 1 twin	1 lam + 1 twin	np	np	4	8	15
			5.1.8	p	p	p	520	4	p	3	4	1 twin	1 lam + 1 twin	2 lam + 1 twin	2 lam + 1 twin	1 lam + 1 twin	np	np	5	11	22	3	1	3	3 lam	2 lam	3 lam	3 lam	1 lam	np	5	12	19
		T5	5.2.1	nv	nv	p	510	nv	nv	nv	5	nv	nv	nv	nv	np	np	np	4	nv	nv	3	1	3	3 lam	1 lam	1 lam + 1 twin	2 lam	2 lam	np	5	10	16
			5.2.2	p	p	p	526	5	p	3	3	2 twin	4 lam + 1 twin	2 lam + 1 twin	1 lam + 1 twin	1 twin	np	np	5	13	22	3	1	3	3 lam	1 lam	1 lam	3 lam	np	np	4	8	15
			5.2.3	p	p	p	592	5	p	2	4	1 twin	3 lam + 1 twin	2 lam + 1 twin	1 lam + 1 twin	2 lam + 1 twin	1 twin	1 twin	7	15	26	3	1	5	3 lam	1 twin	2 lam + 1 twin	3 lam	2 lam	np	5	12	21
			5.2.4	p	p	p	304	4	p	2	2	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	13	4	1	3	3 lam	1 twin	1 lam	np	np	np	3	5	13
			5.2.5	p	p	p	500	4	p	3	4	1 twin	3 lam + 1 twin	2 lam + 2 twin	2 twin	np	np	np	4	11	22	3	1	4	1 twin + 2 lam	1 twin	2 lam + 1 twin	2 lam + 1 twin	np	np	4	10	18
			5.2.6	p	p	p	460	2	p	2	2	1 twin	3 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	13	19	3	1	3	3 lam	1 twin	1 lam	2 lam	np	np	4	7	14
			5.2.7	p	p	p	582	5	p	3	6	1 twin	3 lam + 1 twin	3 lam + 1 twin	1 lam + 1 twin	1 lam + 1 twin	np	np	5	13	27	5	1	4	2 lam	2 lam	4 lam	1 lam	np	np	4	9	19
			5.2.8	p	p	p	nv	3	p	3	4	1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	8	18	3	1	3	3 lam	1 twin	1 lam	np	np	np	3	5	12
	Female	T4	5.5.1	p	p	nv	nv	3	nv	1	3	1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	8	15	3	1	3	3 lam	nv	nv	nv	nv	nv	nv	nv	nv
			5.5.2	p	p	p	nv	3	p	2	2	1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	8	15	4	1	2	1 lam	2 lam	3 lam	1 lam	np	np	4	7	15
			5.5.3	p	p	p	nv	4	p	1	3	1 twin	1 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	7	15	3	1	2	2 lam	1 lam	2 lam	np	np	np	3	5	11
			5.5.4	p	p	p	nv	2	p	1	2	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	10	2	1	2	2 lam	1 lam	1 lam	np	np	np	3	5	9
		T5	5.6.1	p	p	p	274	2	p	3	2	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	12	4	1	1	1 lam	1 lam	np	np	np	np	2	2	8
			5.6.2	p	p	p	350	3	p	2	3	1 twin	2 lam + 1 twin	2 lam + 1 twin	2 lam + 1 twin	np	np	np	4	10	18	4	1	3	2 lam	1 lam	1 lam	np	np	np	3	4	12
			5.6.3	p	p	p	283	2	p	2	2	1 twin	2 lam + 1 twin	1 lam + 1 twin	1 twin	np	np	np	4	7	13	3	1	2	2 lam	1 lam	2 lam	np	np	np	3	5	11
	Mean							3.59		2.64	3.35								4.35	9.86	19.27	3.04	1.00	3.04							3.86	7.50	14.55
	StDev							1.1		1.0	1.3								1.0	3.1	5.1	0.9	0.0	0.9							0.9	2.9	3.8
Regenerated	Male	T4	5.3.1	p	p	p	420	3	p	2	3	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	13	3	2	3	2 lam	1 lam + 1 twin	1 lam + 1 twin	2 lam	np	np	4	8	16
			5.3.2	p	p	p	443	3	p	2	4	1 twin	2 lam + 1 twin	1 lam + 1 twin	np	np	np	np	3	6	15	4	1	3	2 lam	2 lam + 1 twin	2 lam	np	np	np	3	6	15
			5.3.3	p	p	p	260	2	np	1	2	2 lam	1 twin	np	np	np	np	np	2	3	8	1	1	1	1 lam	1 lam	2 lam	np	np	np	3	4	7
			5.3.4	p	p	p	400	2	p	2	4	1 twin	1 lam + 1 twin	1 lam + 1 twin	1 twin	np	np	np	4	6	14	4	1	3	1 twin + 2 lam	1 twin	1 lam + 1 twin	3 lam	1 lam	np	5	10	18
			5.3.5	p	p	p	400	2	p	2	3	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	13	3	1	4	3 lam	1 lam	2 lam + 1 twin	3 lam	1 lam	np	5	11	19
			3.4.T4	p	p	p	nv	2	p	4	3	1 twin	2 lam + 1 twin	1 lam + 1 twin	1 twin	np	np	np	4	7	16	2	2	nv	2 lam	2 lam + 1 twin	2 lam + 1 twin	2 lam	np	np	4	10	nv
			3.5.T4	p	p	p	nv	2	nv	nv	5	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	nv	3	2	3	3 lam	2 lam	2 lam + 1 twin	3 lam	np	np	4	11	
		T5	5.4.1	p	p	p	367	2	p	3	2	1 twin	2 lam + 1 twin	1 twin	np	np	np	np	3	5	12	4	1	3	2 lam	1 lam + 1 twin	2 lam + 1 twin	1 lam	np	np	4	8	16
			5.4.2	p	p	p	401	3	p	2	2	1 twin	2 lam + 1 twin	1 lam + 2 twin	1 twin	np	np	np	4	8	15	5	1	2	2 lam	2 lam + 1 twin	2 twin	np	np	np	3	7	15
			5.4.3	p	p	p	363	2	p	2	3	1 lam	2 lam + 1 twin	1 twin	1 twin	np	np	np	4	6	13	3	2	2	2 lam	1 lam + 1 twin	4 lam	1 lam	np	np	4	9	16
			5.4.4	p	p	p	370	2	p	1	1	1 lam	1 twin	np	np	np	np	np	2	2	6	4	1	2	2 lam	2 lam	2 lam	np	np	np	3	6	13
			3.4.T5	p	p	p	nv	2	p	3	5	1 twin	2 lam + 1 twin	2 lam + 1 twin	1 twin	np	np	np	4	8	18	4	1	4	1 lam	3 lam	2 lam + 1 twin	np	np	np	3	8	17
			3.5.T5	p	p	p	nv	2	p	2	4	1 lam	1 twin	np	np	np	np	np	2	2	10	4	2	4	1 twin + 1 lam	2 lam + 1 twin	2 lam + 1 twin	np	np	np	3	8	18
	Female	T4	5.7.1	p	p	nv		2	p	1	2	1 lam	1 twin	np	np	np	np	np	2	2	7	3	1	1	1 lam	1 lam	1 lam	np	np	np	3	3	8
			5.7.2	p	p	p		2	p	1	1	1 lam	1 twin	np	np	np	np	np	2	2	6	2	1	2	1 lam	1 lam + 1 twin	1 lam	np	np	np	3	4	9
		T5	5.8.1	p	p	p		1	p	1	1	1 lam	1 twin	np	np	np	np	np	2	2	5	3	1	1	2 lam	1 lam	1 lam	np	np	np	3	4	8
			5.8.2	p	p	nv		1	p	1	1	1 lam	1 twin	np	np	np	np	np	2	2	5	3	1	1	2 lam	1 lam	1 lam	np	np	np	3	4	8
			5.8.3	p	p	p		2	p	1	2	1 lam	1 twin	np	np	np	np	np	2	2	7	4	1	2	2 lam	1 lam	2 lam	np	np	np	3	5	12
	Mean							2.06		1.82	2.67								2.83	4.33	10.76	3.28	1.28	2.41							3.50	7.00	13.44
	StDev							0.5		0.9	1.3								0.9	2.2	4.3	1.0	0.5	1.1							0.7	2.6	4.2

p: present **np:** not present **nv:** not visible

Table S2.2: The results of mechanosensation experiments.
Each line represents an individual, presence of behavioral response after the stimuli is recorded as “1” and no-response is recorded as “0”.

The limb amputated	Sample ID	amputated legs						control Legs					
		unamputated		6 dpa		regenerated		unamputated		6 dpa		regenerated	
T4	C#1	1	1	0	0	1	1	1	1	1	1	1	1
	C#4	1	1	0	0	1	1	1	1	1	1	1	1
	C#6	1	1	0	0	1	1	1	1	1	1	1	1
	C#13	1	0	0	0	0	0	1	1	1	1	0	0
	C#14	1	1	1	1	1	1	1	1	1	1	1	1
	C#15	1	1	0	0	1	1	1	1	1	1	1	1
	C#25	1	1	0	0	1	0	1	1	1	1	0	1
	C#28	1	1	0	0	1	1	1	1	1	1	1	1
	C#34	1	1	0	0	1	1	1	1	1	1	1	1
	C#39	1	1	0	0	0	1	1	0	1	1	1	1
T5	C#3	1	0	0	0	1	1	1	1	1	1	1	1
	C#5	0	1	0	0	1	1	1	1	1	1	0	1
	C#10	1	1	0	1	0	1	1	1	1	1	1	1
	C#11	1	1	0	0	1	1	1	1	1	1	1	1
	C#17	1	1	0	0	1	1	1	1	1	1	1	1
	C#18	1	1	0	0	1	1	0	1	0	1	1	1
	C#24	1	1	0	0	1	1	1	1	1	1	1	0
	C#27	1	1	0	0	1	1	1	0	1	1	1	1
	C#29	1	1	1	1	1	1	1	1	1	1	1	1
	C#31	1	1	0	0	0	0	1	1	1	1	0	0
	C#33	1	0	0	0	1	1	1	1	1	1	1	1
	C#37	1	1	0	1	1	1	1	0	1	1	1	1

3. Continuous live imaging of *Parhyale* limb regeneration

Introduction

Parhyale is an attractive model for the live recording of limb regeneration for several reasons. To begin with, *Parhyale* has a hard and rigid exoskeleton, or cuticle, which is also transparent. After amputation, the rigid cuticle confines the regenerating tissue within the leftover limb stump and transparency makes it possible to image through the cuticle to visualize limb regeneration. An elegant method for live imaging and for cell tracking in regenerating *Parhyale* limbs had been established already (Alwes et al., 2016), taking advantage of these features, using transgenic animals that express nuclear-localized fluorescent proteins in all cells under the control of a heat-shock promoter. For live recordings, the limbs and the body of the transgenic animals are immobilized in a glass-bottom petri dish with surgical glue. This method ensures the animals can move most of their body, including their pleopods, whose activity is required for breathing, while the imaged limbs are completely immobile. This has proved to be a powerful technique, generating high-quality datasets where individual nuclei can be tracked throughout the recording to generate cell lineages.

Table 3.1: List of live recordings of *Parhyale* limb regeneration, used to track cell lineages (modified from **(Alwes et al., 2016)**).

		start of recording (hpa)	*duration of the recording (hours)*	*timelapse interval (min)*	*voxel dimensions x:y:z (µm)*	*number of z slices*
#1	*PhHS>H2B::mRFPRuby*	24	40	30	0,32: 0,32: 4,85	22
#3	*PhHS>H2B::mRFPRuby, DC5>dsRed*	48	63	45	0,26: 0,26: 3,32	22
#4	*PhHS>H2B::mRFPRuby*	46	64	30	0,27: 0,27: 2,16	28
#13	*PhHS>H2B::mRFPRuby, PhHS>eGFP-PhGemN*	6	88	30	0,32: 0,32: 3,76	20

Parhyale limb regeneration is relatively rapid compared to other established models used to study limb regeneration; complete leg regeneration in *Parhyale* can be achieved in 1-2 weeks, compared with 3-4 weeks for fin regeneration in zebrafish and 8-12 weeks for leg regeneration in the axolotl. Therefore, considered together with the established cell tracking approach, *Parhyale* is an attractive model to identify the progenitors of limb regeneration.

Previously, cell tracking was carried out on short live recordings of regenerating limbs, covering 2-3 days of the process (table 3.1). In order to construct the complete lineage trees of terminally differentiated cells of the regenerated limbs, we need to extend the live imaging technique to cover the complete time course regeneration, from beginning to end.

Several fluorescent microscopy techniques have been developed in recent decades. Three of these techniques are widely used for live imaging in 3D, namely laser scanning confocal microscopy (LSCM), spinning disc confocal microscopy (SDCM), and light-sheet fluorescence microscopy (LSFM), therefore they are potentially useful to record *Parhyale* limb regeneration. In LSCM the same objective is used to focus a laser beam on the sample and to collect the emitted fluorescence signal from the specimen. A dichroic filter, placed between the objective lens and the detector, reflects the laser on the specimen and lets the emitted light reach the detector. An adjustable pinhole placed in front of the detector removes out-of-focus fluorescence, only allowing the passage of light coming from a specific z plane (figure 3.1a). Therefore, each position of the specimen is illuminated several times, and each time the emitted light from a different z position is collected. Scanning the sample in xyz generates a 3D image of the specimen. The major drawback of LSCM is the high levels of phototoxicity. Each time a voxel is imaged, a cone of illumination excites the tissue only on the z position of the imaged voxel, but throughout the cone of illumination. Therefore, if an object is imaged in 25 optical sections (z sections), each voxel is illuminated 25 times but imaged only once. This situation combined with relatively low scanning speed, exposes the specimen to light more compared to the other imaging methods.

In SDCM, the confocal images are generated in a fundamentally different way. Two discs, one with ~20.000 pinholes and another with a matching pattern of microlenses, are placed co-axially between the light source and the objective. During the image acquisition, each of these microlenses focuses the incoming laser to a corresponding pinhole. Multiple positions of the specimen are illuminated simultaneously and the emitted light is collected simultaneously from multiple positions. The rapidly spinning discs allow rapid sampling of the image space. A dichroic mirror placed between the two discs directs the emitted light to the detectors (figure 3.1b). This method has three major advantages over classical LSCM. First, the detection of the signal is utilized by EM-CCD cameras, which are significantly more rapid than the PTD detectors used in LSCM. Second, images can be acquired much faster with a low signal to noise ratio (SNR) since the detection sensitivity is very high. As a result,

SDCM can generate the same image 10-100 times faster than the LSCM, which decreases the amount of photodamage. However, these advantages come at a cost since there are some major drawbacks to SDCM. First, the highly sensitive EM-CCD cameras record only 512 x 512 pixels, which results in a low-resolution image at the center of the field of view (FOV). Moreover, this FOV is significantly smaller than the one of LSCM, which makes it very challenging to image large specimens with SDCM (figure 3.2). Another problem stems from the fact that there are several pinholes located on the disc close to each other, allowing out-of-focus light from different focal planes to enter adjacent pinholes. As a result, the axial resolution of SDCM is significantly lower than in LSCM.

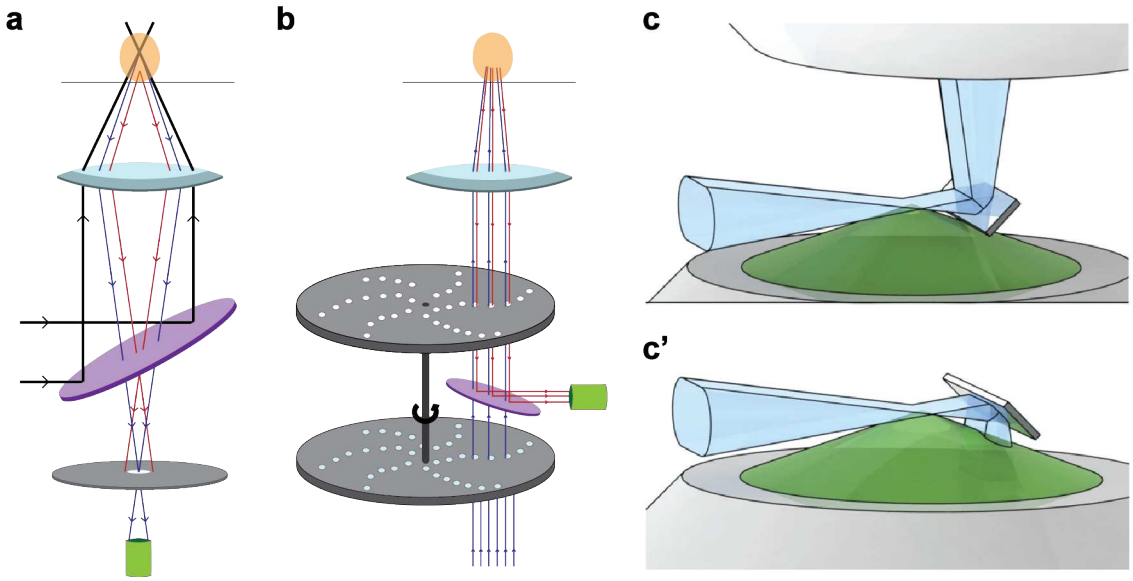


Figure 3.1: 3D imaging techniques that could be used to image regenerating Parhyale limbs.
(a) Illustration of a laser scanning confocal microscopy setup. The laser beam (black lines) is directed to the sample (orange) via a dichroic mirror (purple) and objective lens (turquoise). Emitted fluorescent light from a desired focal plane (blue lines) is collected by the detector (green) while out-of-focus light (red lines) is excluded by a pinhole (gray). (b) Illustration of a spinning disc confocal microscope setup. The laser (blue lines) is first passing through two co-axial spinning discs with matching arrays of microlenses and pinholes, and then directed to the specimen by an objective. Emitted fluorescent light (red lines) is collected through the same pinholes and directed to the detector by a dichroic mirror. Illustrations of (c) reflected light-sheet microscopy and (c') single objective selective plane illumination microscopy setups. In both of these techniques, a light sheet (turquoise) is directed to the sample using a mirror. In the first technique, there are two objectives (shown in grey), one to direct the light sheet and the other for detection, whereas the latter technique is using the same objective for both purposes. The cone of emitted fluorescence captured by the objective is shown in green. Panel c-c' is from (Power and Huisken, 2017).

The last technique, LSFM, uses a light sheet to illuminate the sample from a direction that is perpendicular to the angle of detection. The detection objective, placed perpendicular to the light sheet, collects the fluorescent light emitted from the sample. This mode of illumination ensures that, instead of illuminating all of the axial positions and collecting the signal from only one, the sample is only illuminated in the detection

plane. This way, all of the signals can be detected without filtering, since there is no out-of-focus light in the system. In this method, the samples can be imaged with minimal light exposure compared to confocal methods for two reasons: every xyz position of the sample is illuminated once during imaging. Many variants of this method have been developed and become widely used in live imaging of embryonic development (Reviewed in (Colombelli and Lorenzo, 2014; Power and Huisken, 2017)). Imaging *Parhyale* limbs with LSFM is technically challenging because it is necessary to immobilise the specimen by attaching the imaged limb on a cover slip. The presence of coverslip is incompatible with the imaging setup of many LSFM techniques, such as SPIM, used to image large samples like *Parhyale* legs. However, two variants of this technique are potentially applicable for *Parhyale* limb regeneration, since they do not require a light source placed perpendicular to the imaging objective: reflected light-sheet microscopy (RLSM) and single objective selective path illumination microscopy (soSPIM). In RLSM, both the light source and detection objectives are placed perpendicular to the cover slip but on the opposite sides of the specimen, and the light is directed to the specimen using a microprism (Gebhardt et al., 2013; Greiss et al., 2016). In soSPIM, a single objective is used both for illumination and detection. A mirror placed next to the sample directs the light sheet to the sample ((Galland et al., 2015), figure 3.1c-c').

Among the above-mentioned microscopy techniques, the LSFM variants are the most promising ones to record the complete regeneration process of *Parhyale* limbs. However, these are relatively new techniques, and the components are not commercially available. A microscope, adjusted to *Parhyale*, should be assembled *de novo* and custom-made imaging components, like glass-bottom Petri dishes with microprisms or mirrors, should be produced. These tasks require some engineering skills we currently lack in our lab. On the other hand, SDCM is offering great advantages for long-term live imaging but the constraints on the FOV and resolution are very restrictive for imaging and tracking cells during limb regeneration. Therefore, I have decided to use SLCM, as Alwes *et al.* did, to generate complete recordings of *Parhyale* limb regeneration with a resolution that allows cell tracking. This method has the highest risk of damaging the tissue by light exposure; therefore I will especially focus on limiting the light exposure during imaging.

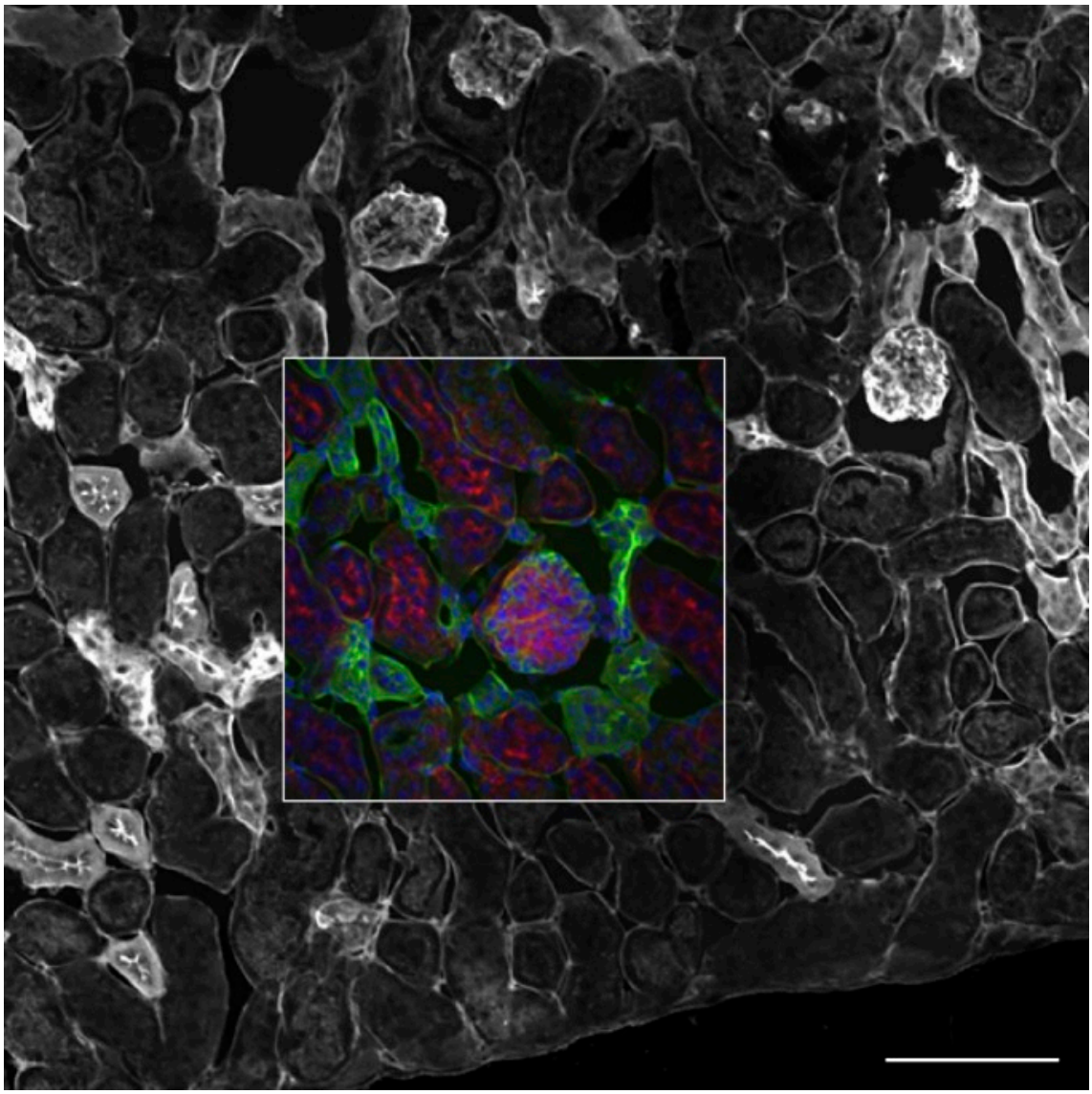


Figure 3.2: Comparison of the fields of view (FOV) for a spinning disc and a laser scanning confocal microscopy.
A fixed mouse kidney section was imaged with the same 20x/0.75 NA objective lens on a Yokogawa spinning-disk confocal equipped with a 512 x 512 pixel EM-CCD camera (inner square), using the full FOV available on an Olympus FluoView 1000 CLSM (outer square; Tokyo, Japan). Scale bar, 100 μm (from (Jonkman and Brown, 2015)).

Results and discussion

Extending and optimizing continuous live imaging of regenerating limbs

There are two main requirements of a good dataset for cell tracking during limb regeneration. First, all of the cells of regeneration should be visible and in focus throughout the recording. And second, the quality and the resolution of the image should be good enough to identify cells in space (to distinguish the cells from each other in a given frame) and in time (to identify a given cell in two consecutive time points). In pursuance of the goal of generating complete recordings of limb regeneration with such qualities, I optimized several parameters of sample preparation and image acquisition.

I have started with the optimization of the FOV. *Parhyale* can regenerate their limbs independent of the amputation position and the regeneration takes place in the distal part of the leftover stump. Thus, the amputation plane dictates both the location and also the size of the regenerated tissue. In a previous study, Alwes *et al.* performed limb amputations within the carpus of T4 or T5 legs and imaged limb regeneration in the distal 100-200μm of the limb stump (Alwes et al., 2016) which was small enough to fit in the FOV with a 20x objective. However, such amputation is not optimal for live imaging for two main reasons. First, when the amputation plane is located proximally on the carpus, the regenerated tissue generally spans the carpus-merus junction and the cuticular architecture hinders imaging due to autofluorescence and the folding of tissues at the limb joints. Second, such amputations lead to the regeneration of three podomeres: carpus, propodus, and dactylus. Given that space is already limited by the presence of the joint, the regenerating tissue folds extensively (figure 3.3b). Considering that axial resolution is generally worse than lateral resolution in confocal microscopy, such folding decreases our ability to resolve and to track cells. I tried several different amputation planes on T4 and T5, to address these problems, and found that amputations at the distal part of the carpus, just the posterior of the pit that is located on the ventral side of the joint, give the best chances to image the regenerating limb clearly without interference from the joints. With distal amputations on the carpus, regeneration is still taking place at the distal 150-200μm of the stump, but now this happens entirely within the carpus, which presents a smooth surface through which we can image the regenerating limb. Under these conditions, autofluorescence from the cuticle is reduced and the regenerating tissue is folding considerably less (figure 3.3c).

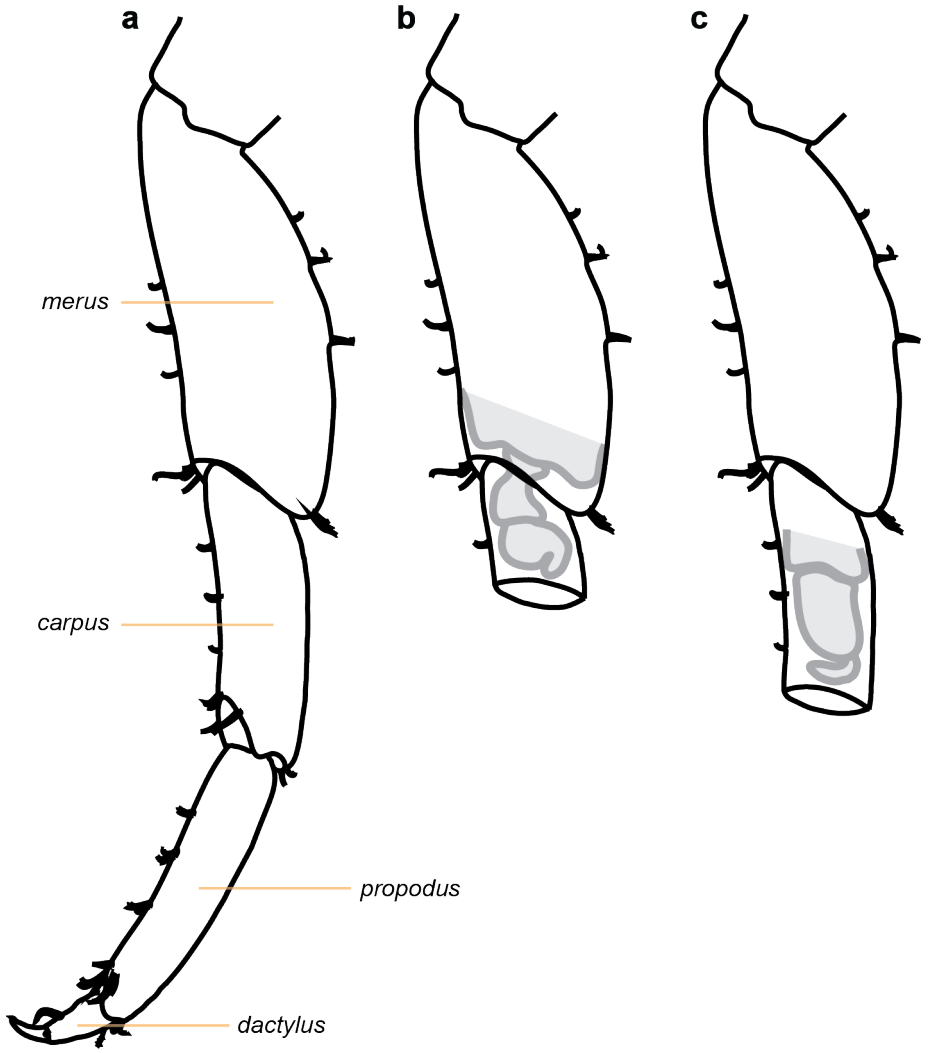


Figure 3.3: The effect of the amputation plane on the localization and the shape of the regenerating limbs.
(a) Illustration of an unamputated T4 or T5 limb and the limb regeneration when amputation is performed on (b) the proximal or (c) the distal part of the carpus.

My next step was to improve the image quality. What is considered a good quality image is usually context-dependent. In collaboration with my colleague Ko Sugawara, we focused on improving two major parameters: the contrast ratio (CR) and the SNR of the images. These parameters are defined by the following equations: CR = $|FG|/|BG|$ and *SNR* = $|FG| - |BG|/stdev|BG|$ (Ulman et al., 2017), where |FG| is the mean foreground signal, the average signal intensity of all nuclei in a given image, and |BG| is the mean background signal, the average signal intensity of all areas except the nuclei. In our case, the background signal results largely from the autofluorescence of the cuticle, light scattering in the tissues, and the surgical glue used to stabilize the limbs. Foreground signal depends on the expression levels of the fluorescent marker, the heat-shock intervals, and the mounting of the sample. For a given microscopy setting, an image with a higher CR value is an indication of the

specimen having high levels of fluorophore expression and being properly mounted. SNR measures the image noise, which is arbitrary variation in brightness, generally due to the technical apparatus like sensors (Gonzalez and Woods, 2018). Thus, having a high SNR value is a good indication of the optimization of the image acquisition settings with minimal artifacts.

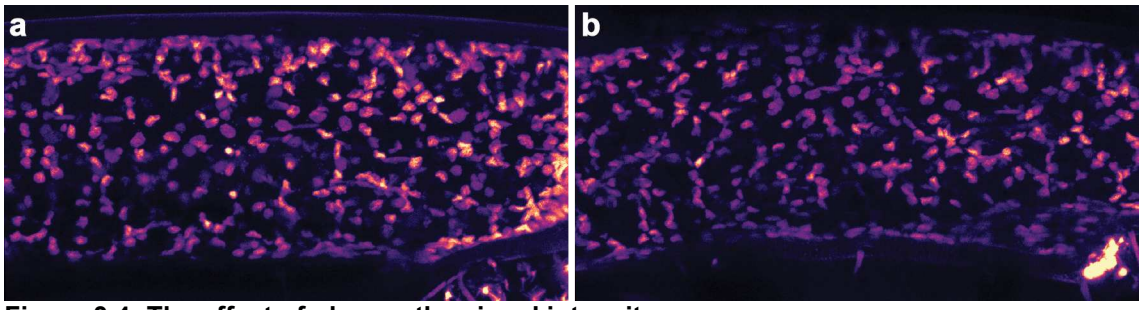


Figure 3.4: The effect of glue on the signal intensity.
Maximum projection of the images of carpus podomere of (a) T4 and (b) T5 unamputed limbs of the same animal, under the same 20x objective (Zeiss 420650-9901-000, N.A. 0.8). The glue layer between the limb and the glass is ~4µm thicker in the T5 limb, resulting in lower signal.

There are two critical steps in sample preparation that are specific to our experiments. First, the animals need to be immobilized on a glass coverslip (generally a glass-bottomed petri dish) using surgical glue (Alwes et al., 2016). The amount of glue between the limb and the glass is very critical: we need to apply a sufficient amount of glue to attach the animal, but using too much increases light scattering and reduces the signal (figure 3.4). The second important point is to ensure having high levels of fluorescent proteins in the sample throughout the imaging. I have been using transgenic animals expressing a fusion between histone H2B and the fluorescent protein mRFPRuby, driven by a heat shock promoter (Alwes et al., 2016). The animals are heat-shocked at 37°C for 45min to induce transgene expression. Under those conditions, the transgene is transcribed during ~2 hours following the heat shock (Pavlopoulos et al., 2009) but the fluorescent signal persists for many days in the absence of cell divisions (figure 3.5). The durability of the signal is a result of the stability of the histones, and also shows that photobleaching of the fluorescent proteins is minimal in our imaging setup.

However, the fluorescent signal declines rapidly as the cells divide, presumably by dilution due to the growth in cell/tissue mass. Consequently, following an initial heat-shock performed 16-24 hours before amputation; additional heat-shocks are required during the live recordings, depending on their duration. Our confocal microscope is equipped with a temperature-controlled stage that makes it possible to heat-shock the animals in situ, without disrupting the imaging. Frequent heat-shocks ensure a good signal but may also induce premature molting and increase mortality. After several

trials, I found that a good interval for heat-shocks is every 36-48 hours once the cell proliferation phase has started.

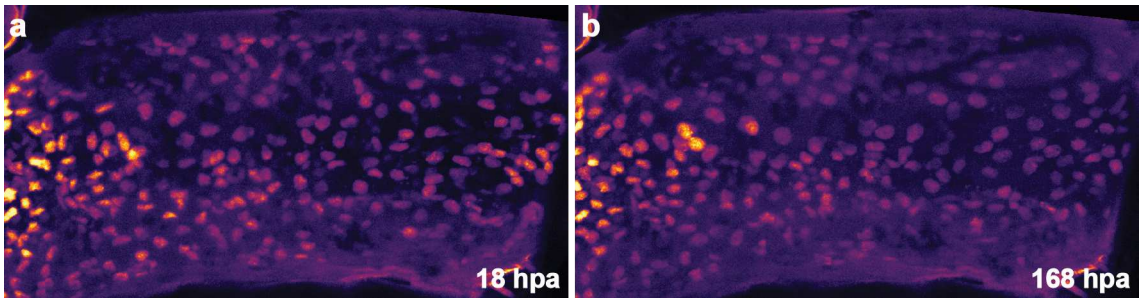


Figure 3.5: The stability of the H2B::mRFPRuby protein.
Two frames from a continuous recording of an amputated but non-regenerating limb, imaged under a 20x objective (Zeiss 420650-9901-000, N.A. 0.8), show that the fluorescent signal is comparable (a) in the first and (b) the seventh day of post-amputation even though the animal was not heat-shocked in between. hpa: hours post-amputation.

Having established the conditions for obtaining well-mounted specimens, I focused on optimizing the image acquisition. Our Zeiss LSM 800 confocal microscope is equipped with three GaAsP detectors and since the live imaging is performed with a single type of fluorophore, I could use either of these detectors. I tested and compared the efficiency of these detectors by imaging the same sample with identical settings using each detector. The images revealed that the second detector undoubtedly has better performance (figure 3.6). That result was surprising, considering that detector 2 should not be the one having the shortest optical path or the least number of optical components, like mirrors and variable secondary dichroics (VSDs). This may reflect an error in the numbering of detectors in our microscope.

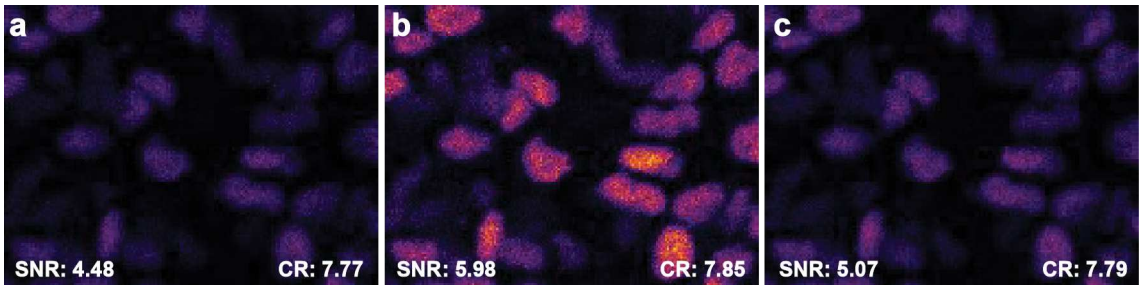


Figure 3.6: Testing the efficiency of the confocal detectors.
The same frame is simultaneously imaged with detectors (a) 1, (b) 2, and (c) 3, using identical settings (20X magnification, Zeiss 420650-9901-000, N.A. 0.8). SNR: signal to noise ratio, CR: contrast ratio.

In a laser scanning confocal, a laser beam scans through the sample, detecting the fluorescence emitted at each point of focus. There are two important parameters related to this process: imaging speed and averaging. The first refers to the amount of time the detector spends at each point during image acquisition (giving rise to each image pixel); the second refers to the option of scanning each pixel multiple times and taking an average of the measured signal. Scanning slower, as well as multiple times,

increases the image quality but also the light exposure on the sample. Imaging speed and averaging are expected to have similar effects with regard to light exposure: doubling the scan time per pixel should be equivalent to doubling the number of images captured for averaging. Thus, setting these parameters will involve a trade-off between image quality and phototoxicity.

To optimize these parameters, I tested different averaging and scan speed combinations and measured the quality of the resulting images (figure 3.7a). As expected, there is a strong correlation between the SNR and the scanning speed and averaging: increasing the imaging time of each pixel, either by slower scanning or by increased averaging, consistently raises the SNR of the image. The relationship with CR is more complex, but the CR tends to be higher in images acquired with slower scanning (figure 3.7b). Keeping in mind the need to reduce minimize light exposure on the samples, I have generated short live recordings to assess the tracking performance with some of these settings: a speed of 2.06 µs per pixel with an averaging of 2 images, or 1.03 µs per pixel with an averaging of 4 are the fastest settings that generate images of sufficient quality for manual or semi-automated tracking. In my subsequent live recordings, I decided to record at 2.06 µs per pixel, averaging over 2 images.

Finally, I have worked to determine the resolution needed to support efficient cell tracking. Previously, Alwes *et al*. generated live recordings with different spatial and temporal resolutions (see Table 1 in (Alwes et al., 2016)). In collaboration with Ko Sugawara, I used these image data, along with some short recordings that I generated, to identify the effect of resolution on tracking performance. Alwes *et al*. recorded images of 1024 x 1024 pixels, each pixel measuring 0.20-0.32 µm in length and width, which is sufficient both for manual and semi-automated tracking. However the axial (z) and temporal resolutions of those recordings were not optimal. In my experience, to reliably discriminate nuclei in the z axis, the step size should be lower than 2.5 µm. For Ko Sugawara's auto-tracking algorithm, it is desirable to have a lateral:axial resolution ratio 1:4, or in the worst case 1:8. This means having a step size of 1.24 µm or 2.48 µm, for lateral pixel dimensions of 0.31 µm. Regarding the temporal resolution, I found that imaging intervals longer than 20-25 minutes increased the risk of losing track of cells during mitosis, either by manual or semi-automated tracking. For example, I tried to reproduce some of the nuclei trackings in recording #4 of Alwes *et al*. and there were some discrepancies between my cell lineages and those reported by Alwes *et al*., mainly because it is challenging to trace the nuclei during cell division with 30min time

resolution. For this reason, I have chosen to set the imaging interval to 20 minutes. I did not choose shorter time intervals to minimize photodamage.

One of the limitations in the images I acquired, is the low image quality when imaging deep in the tissue. It is only possible to get good quality images of the first 20-25 μm from the surface of the limbs, which accounts for approximately half the depth of these thoracic limbs at the level of the carpus. I have been using a 20x dry objective. The limb tissue has a much higher refractive index, expected to be between 1.30-1.40 as measured for cells (Liang et al., 2007), and the mismatch is causing a dramatic decrease in image quality in deeper sections. Likewise, the mismatch in refractive indices also causes the limb to appear thicker than it is in reality in the axial dimension, (Diaspro et al., 2002; Egner and Hell, 2006). In order to improve image quality, I have tried to use a 40x objective (Zeiss 420862-9970-790, N.A. 1.2) with silicone oil immersion, which has a refractive index of 1.40. Since the lateral and axial resolutions of the images taken with the 20x objective were good enough, I used the same z step size, and a 0.5x zoom with the 40x objective, to obtain a similar magnification and FOV. In order not to lose any signal due to the thinner optical sections, I adjusted the pinhole to 2 Airy units. 3D images I captured with these settings, indeed had better quality compared to the ones taken with the 20x objective, therefore I tried to use these settings to perform live recording of regenerating limbs. However, in all of my trials (19 different settings on 4 animals in total) either the imaged tissue died or the amputated limbs failed to regenerate. I attribute this to increased phototoxicity.

Taking all these criteria and restrictions into account, I have established standard conditions for reliably imaging the entire limb regeneration process, from amputation to molting. My recordings typically span 4-12 days, acquired with a dry 20x/NA 0.8 objective (Zeiss 420650-9901-000) at 1024 x 1024 pixel resolution, using 0.15 - 0.50% laser power with 650-750 mV detector gain, and with 20-minute intervals (Table 3.2). Using these settings I managed to image complete limb regeneration of 12 animals among 19 experiments, which accounts for a 63% success rate. Among the 7 unsuccessful experiments, in 4 animals molted before starting to regenerate, in 1 the animal detached itself from the cover slip, and in 2 the animals did not start to regenerate after 3 days so I stopped imaging. Overall, the success rate is comparable to the experiments of Alwes *et al*., where the survival rate was 80%. My recordings are of considerably longer duration and often capture two limbs per animal.

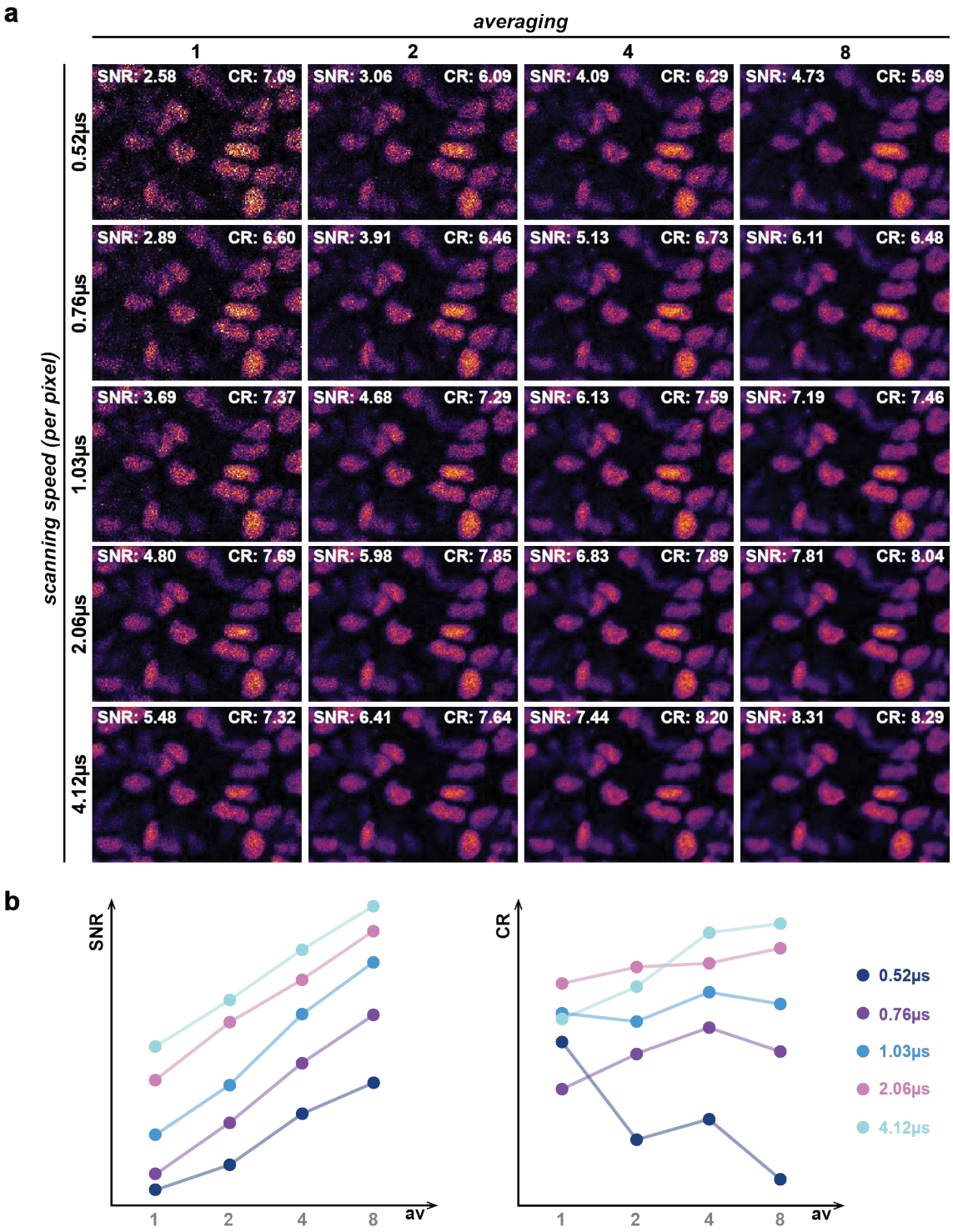


Figure 3.7: The effect of scanning speed and averaging on image quality.
(a) Images acquired with different scanning speed and averaging settings, using a 20x objective (Zeiss 420650-9901-000, N.A. 0.8). (b) Plotting the effect of averaging and scanning speed on SNR and CR. All of the images were acquired with the same parameters, while varying scan speed and averaging. SNR: signal to noise ratio, CR: contrast ratio, av: averaging.

A complete overview of limb regeneration

The different stages of limb regeneration (Konstantinides and Averof, 2014) and corresponding cell behaviours during each stage (Alwes et al., 2016) have been described. With complete recordings, I can observe individual limbs going through all of these phases and I can expand the existing description to include the later stages of regeneration. I outline here my observations on this process, in relation to what was described by Alwes *et al.*

Table 3.2: The list of complete live recordings of *Parhyale* limb regeneration.

		imaged limb	*duration of the recording (hours)*	*voxel dimensions x:y:z (µm)*	*number of z slices*
li13	*PhHS>H2B::mRFPRuby*	T4 & T5	168	0,31: 0,31: 2,48	11
li15	*PhHS>H2B::mRFPRuby, DC5>H2B::eGFP*	T5	110	0,31: 0,31: 1,80	15 - 18
li16	*PhHS>H2B::mRFPRuby, DC5>H2B::eGFP*	T4 & T5	231	0,31: 0,31: 1,24	24
li24*	*PhHS>H2B::mRFPRuby, DC5>H2B::eGFP*	T4 & T5	113	0,31: 0,31: 1,24	30
li25*	*PhHS>H2B::mRFPRuby*	T4 & T5	190	0,31: 0,31: 1,24	30
li27	*PhHS>H2B::mRFPRuby, DC5>H2B::eGFP*	T4 & T5	264	0,31: 0,31: 1,24	30
li30*	*PhHS>H2B::mRFPruby*	T4 & T5	189	0,31: 0,31: 1,24	27
li34*	*PhHS>H2B::mRFPRuby*	T4 & T5	237	0,31: 0,31: 1,24	27
li36*	*PhHS>H2B::mRFPRuby*	T5	145	0,31: 0,31: 1,86	18
li37*	*PhHS>H2B::mRFPRuby*	T4 & T5	137	0,31: 0,31: 1,86	17
li38	*PhHS>H2B::mRFPRuby*	T4 & T5	210	0,31: 0,31: 1,86	17
li40*	*PhHS>H2B::mRFPRuby*	T4	165	0,31: 0,31: 1,86	15

*Recordings labelled with asterisk were stopped before molting to proceed with antibody stainings (see chapter 6 for details). For all of the recordings, the time resolution is 20 minutes. The animal in li38 recording died while molting after the regeneration is completed.

The first response following the amputation is wound closure, mediated by haemocytes adhering to the wound surface. This is a very rapid process and the temporal resolution of my recordings is not sufficient to record in detail, yet it is still possible to observe the boundary forming between the amputated limb and the outside in the early frames of the recordings (Video 3.1, 0-22 hpa). In the next phase, the wound surface turns into a melanised scab. In the recordings, we can observe fluorescent cells (which probably represent haemocytes) becoming integrated in the scab. After the scab is formed the epithelial cells start to migrate towards the wound re-

establishing an epithelial layer between the scab and the limb tissue (Video 3.2, 17-50 hpa). In the earliest phases, motile cells (probably representing macrophages) can be observed moving rapidly and making contact with the wound surface and with each other, below the wound surface (Video 3.2, 0-50 hpa). I can also observe several apoptosis events (seen as nuclei that condense and break up), especially towards the end of this phase (Video 3.2, 25-75hpa).

The following phase is marked with extensive cell-proliferation. The cells that lie close to the wound area start to divide rapidly and the new tissue begins to form. The newly forming tissue detaches from the cuticle and the new limb starts to take shape (Video 3.2, 60-100 hpa). The segments of the new limb are visible from relatively early stages in this phase. Gradually cell proliferation rates decline as the limb continues to grow. In some recordings, the growing podomeres can be seen pulsing with a periodicity of 6-7 hours (Video 3.2, 110-146 hpa). The latter movement had not been previously described and its significance is unclear. Finally, all the cell movements and divisions cease. Terminal differentiation of most cell types is completed during this phase. The newly regenerated limb becomes exposed when the animal molts.

Conclusions

In this study, I have improved the method of live recording originally developed by (Alwes et al., 2016), and extended it to allow imaging of the complete limb regeneration process. The current method is robust, and the resulting datasets are of high quality, allowing cell tracking of the majority of recorded cells in the surface of the limb. One of the datasets that I generated (T4 limb, recording li13) has been used for extensive ell tracking using the semi-automated cell-tracking tool described in Appendix B. The performance was very satisfying and considering this dataset is one with the lowest axial resolution, I expect the performance of semi-automated tracking to be better in the other datasets. Apart from lineage tracking, these live recordings are being used to quantify the contribution of different cell behaviors, such as oriented cell divisions and cell rearrangements, to morphogenesis in regenerating limbs by my colleague Severine Urdy.

The quality of the resolution of the images are sufficient for the purpose of my project, which aims to trace the progenitors of individual cells, from their terminal differentiated state back to their precursors in the blastema and the freshly amputated limb. With the current setup, I am limited to analysing the cells that are close to the surface of the limb since the image quality and the axial resolution decline deeper in the tissue. To be able to image deeper tissues, a better optical setup is required. I have tried to use a 40x/NA1.2 objective with silicone oil immersion, but the phototoxicity was too high. Maybe a 25x objective again with the same immersion oil would work better. Another option is to explore other imaging techniques like LSFM.

One practical limitation of these live recordings is the time needed to complete each recording. Considering the time span of regeneration (and failed experiments), each dataset requires on average ~10 days. At each time point, imaging requires ~2 minutes of acquisition time and the microscope stays idle for ~18 min until the next time point. This implies that it should be possible to image multiple samples simultaneously. I have made several trials to achieve this, and my next goal is to optimize this to enable simultaneous recording of multiple specimens.

Video 3.1: https://www.dropbox.com/s/1efutuc3d6qhbij/Video 3.1.avi?dl=1

Video 3.2: https://www.dropbox.com/s/5z0j0se6pcz9hs3/Video 3.2.avi?dl=1

4. CRISPR-mediated knock-in approach to generate cell type-specific marker lines of *Parhyale* peripheral sensory organs

Introduction

In recent years, highly precise genome editing has become available with versatile technologies that are applicable to diverse species. These approaches are re-shaping many fields in the life sciences and provide a deeper understanding of a range of biological processes like embryonic development and cancer progression. In this study, my objective was to use these gene-editing technologies to insert a reporter cassette into the genomic loci of putative marker genes, aiming to generate cell type-specific marker lines to study the regeneration of the peripheral sensory organs of *Parhyale* limbs.

The emergence of precise genome editing

Introducing precise changes into genomes has been a great interest of scientists for a long time. In the mid-1980s, the first targeted editing in eukaryotic genomes, initially in yeast (Rothstein, 1989) and later in mammalian cells (Smithies et al., 1985; Thomas et al., 1986), was reported, based on homologous recombination. However, the efficiency of this technique was very low, only 1 in 10^3-10^7 cells were transformed, making it applicable only for the experimental systems where screening large numbers of cells is possible, such as cultured mammalian cells or yeast. The frequency of homologous recombination, nonetheless, is greatly increased when a double-stranded break (DSB) is introduced at the genomic locus of interest (Szostak et al., 1983). However, techniques to introduce a DSB in a targeted way were not available for another decade.

First, special restriction enzymes recognizing 14-40bp long DNA sequences, called meganucleases, were used to cut the genome in a desired target locus (Rouet et al., 1994). However, naturally occurring meganucleases provide a very limited target choice and designing them *de novo* is very challenging (Takeuchi et al., 2011). Therefore meganucleases did not become very popular for genome engineering. Later, programmable endonucleases were engineered by combining DNA recognizing zinc-finger proteins with restriction enzymes. These engineered zinc finger nucleases (ZFNs), contain several zinc-finger domains (ZF), each one targeting a 3-nucleotide DNA motif. By generating ZFNs with different combinations of ZFs, in theory, any

location in the genome could be targeted (Bibikova et al., 2001). However, there is no ZF for each possible 3-nucleotide motif, and the target sequence of some ZFs varies depending on the neighboring ZFs. This leads to a decrease in target locus predictability and combined with a costly design process, ZFNs did not have a wide use (Sander et al., 2011). Later, these problems were addressed with the development TAL effector nucleases (TALEN). Instead of ZF, these new proteins contained TAL effector (TALE) DNA-binding modules. There is a different and highly specific TALE module for each of the A, T, C, and G nucleotides, which makes it possible to custom design arrays of TALENs to recognize different DNA motifs (Christian et al., 2010). This highly versatile method was superior to ZFNs since it is easier to design highly precise nucleases for almost any sequence (Reviewed in (Gaj et al., 2016)). The invention of TALENs was shadowed by another method developed only a few years later, CRISPR/Cas9. This new method only requires designing a short RNA, rather than a new protein, to target different DNA sequences, therefore it gained popularity very quickly and started a genome-editing revolution (Jinek et al., 2012).

CRISPR revolution

The CRISPR/Cas9 genome editing technique is relying on an adaptive immune mechanism of archaea and bacteria (Mojica et al., 2005). Many prokaryotic genomes contain “clustered regularly interspaced short palindromic repeats” (CRISPR) in which the short sequences from invading viruses (called protospacer sequences) are incorporated and transcribed as CRISPR-derived RNAs (crRNAs). These loci are associated with proteins called CRISPR associated proteins (Cas), which are involved in degrading viral DNA together with the crRNAs. The Cas nucleases can cleave the viral DNA only if the crRNA binding site is adjacent to a specific DNA sequence, called the protospacer-adjacent motif (PAM). Since the crRNA sequence in the CRISPR loci does not have a neighboring PAM, the bacterial genome is protected from being cleaved (Gilles and Averof, 2014).

Several Cas proteins have been identified to date but only two of them, Cas9 and Cas12a, have nuclease activity. Cas9 from *Streptococcus pyogenes* (spCas9) is the major protein used for genome editing because it requires a rather simple PAM sequence, NGG, and relatively short protospacer sequence of 20 bp (reviewed in (Adli, 2018)). To be functional, the spCas9 protein needs to be associated with two RNA molecules, the crRNA, including the protospacer sequence, and the trans-activating RNA (tracrRNA) which has an invariable sequence. These two can be fused into a single RNA, called sgRNA, which makes the experimental setup simpler ((Jinek et al., 2012), figure 4.1).

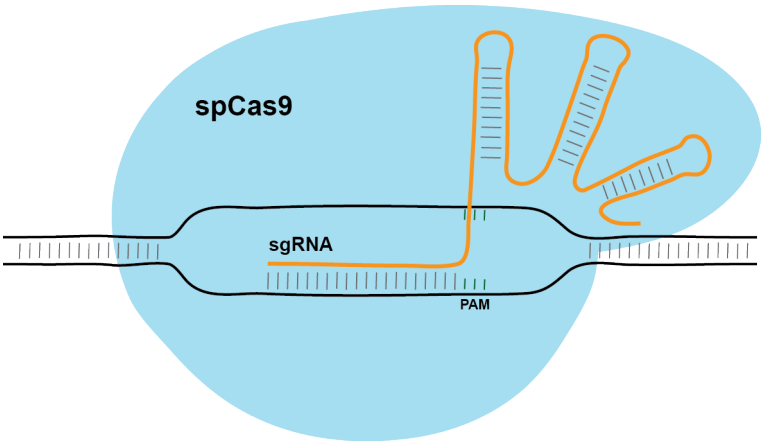


Figure 4.1: Cas9-sgRNA nucleoprotein complex-DNA interaction, as engineered for gene editing
An spCas9 protein (blue) and sgRNA (orange) complex targeting the genomic sequences that are complementary with the 5' 20 bp of the sgRNA. The target locus should also contain a PAM sequence (NGG) immediately downstream of the region targeted by the gRNA for targeting by the Cas9-sgRNA complex. Once the spCas9 protein binds to the DNA target, it introduces a double-strand break 3 bp downstream of the PAM sequence.

Design and synthesis of CRISPR-mediated genome editing reagents

CRISPR-mediated genome editing requires two reagents to be delivered into the cells, the Cas nuclease and the sgRNA. The only decision to make at this step is to choose the type of Cas nuclease to use. There are several plasmids available with the coding sequences of different Cas9 and Cas12a variants, codon-optimized for different species or carrying tags such as nuclear localization signals. Cas nucleases can be delivered in the form of expression/helper plasmids, as *in vitro* transcribed capped mRNA or as purified proteins, which are commercially available.

Designing the sgRNA is a critical step of the CRISPR technique. For spCas9, in theory, any 20 bp DNA sequence followed by a PAM sequence can be used as a target. However, numerous studies have revealed that the success rate of genome editing varies widely for different targets. This may result from the local chromatin structure and the resulting accessibility of the target locus (Janssen et al., 2019), or the nucleotide composition of the sgRNA, which could produce different secondary structures (Moreno-Mateos et al., 2015). The exact molecular basis of the effect of the protospacer/target sequence on genome editing has yet to be discovered (Anzalone et al., 2020). Based on genome-wide screens and other experimental results, several algorithms have been developed to predict sgRNA performance, but none of these can predict the sgRNA efficiency with high precision (Cui et al., 2018).

Once the protospacer sequence is chosen and the sgRNA is designed, it can be delivered into cells in the form of an expression/helper plasmid or as *in vitro* synthesized RNA. Each of these methods has its limitations. Within eukaryotic cells

sgRNAs are typically expressed using RNA Pol III promoters, therefore the helper plasmid should contain the U6 snRNA promoter or equivalent, which imposes the requirement that transcripts start with a G. Alternatively, *in vitro* sgRNA synthesis can be performed either enzymatically or chemically. For enzymatic synthesis, *in vitro* transcription is usually performed using the T3, T7, or SP6 bacteriophage RNA polymerases, which generate transcripts starting with GG (for T3 and T7) or GA (for SP6). These alternatives impose slightly different constraints on the target sequence. Chemical synthesis does not constraint the target sequence any further than having an adjacent PAM sequence. There are several methods to chemically synthesize the sgRNAs (Kelley et al., 2016) but are expensive.

Using a helper plasmid to deliver sgRNA is easier since it does not require synthesizing and handling the sgRNA. However, suitable promoters, like U6, have not been characterized in many species. Chemical synthesis of sgRNAs is convenient since it does not restrict the protospacer sequence choice, but it is costly. The most common method of synthesis is *in vitro* transcription using bacteriophage polymerases. Therefore, the targeted genomic sequences are, theoretically, limited to $GGN_{18}NGG$ or $GAN_{18}NGG$ when spCas9 is used. However, there are several studies reporting that a 5' mismatch of the sgRNA is well tolerated (Gilles and Averof, 2014; Moreno-Mateos et al., 2015). Indeed, within a given target genomic locus, some protospacer sequences could be favorable, over canonical $GGN_{18}NGG$ or $GAN_{18}NGG$ ones, even if they are shorter than 20bp or carry mismatches at the 5' end (Moreno-Mateos et al., 2015). This provides more flexibility in choosing a target sequence. The sgRNAs with 5' mismatches or bearing a protospacer sequence shorter than 20bp are referred to as non-canonical sgRNAs.

DSB repair pathways: how to insert transgenes into genomes

Once a DSB break is introduced into its genome, a cell either repairs it or goes through apoptosis (Kaina, 2003). There are two major molecular mechanisms to repair a DSB: non-homologous end joining (NHEJ) and homology-directed repair (HDR). In NHEJ, the two DNA ends are ligated to each other. This process is error-prone and can introduce point mutations, small indels, and occasionally even larger sequence modifications to the repair site. HDR uses a homologous DNA sequence as a repair template and results in a seamless repair (reviewed in Wynman and Kanaar, 2006). Both pathways can be 'hacked' to insert foreign DNA into the genome (reviewed in (Gilles and Averof, 2014), figure 4.2).

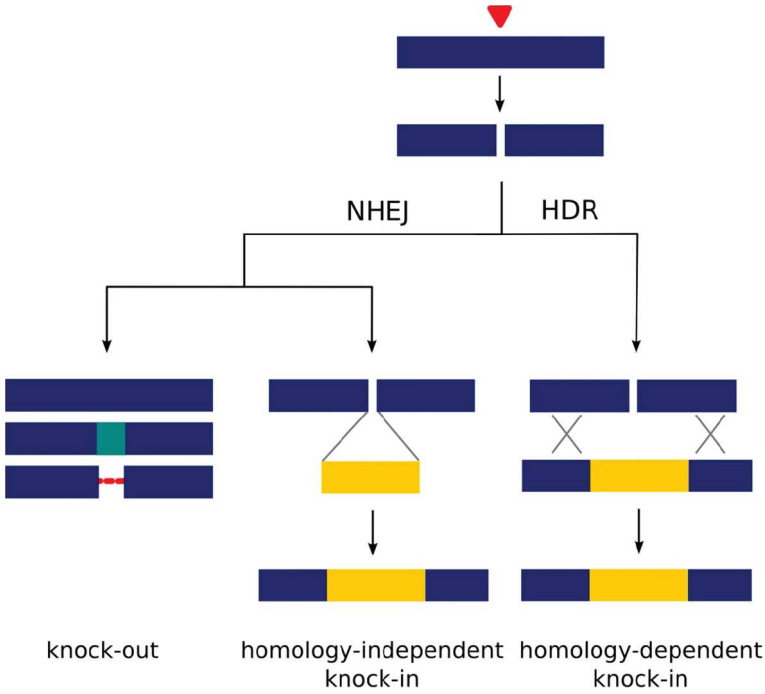


Figure 4.2: Gene targeting strategies using targeted double-strand breaks.
When chromosomal DNA is cleaved (red arrowhead), the resulting double-strand break is repaired by non-homologous end joining (NHEJ) or by homology-dependent repair (HDR). NHEJ may result in perfect rejoining, of the ends, or in the introduction of point mutations and indels (knock-out). NHEJ may also join exogenous linear DNA (yellow) to the broken ends of the chromosome (homology-independent knock-in); the orientation and reading frame in these insertions is random, unless directed by complementary overhangs. HDR repairs the double-strand break by precise copying of a repair template carrying an exogenous sequence (yellow) flanked by sequences with homology to the targeted locus (in blue) (homology-dependent knock-in). The repair template usually consists of circular plasmid DNA with long homology arms (from (Gilles and Averof, 2014)).

If CRISPR-mediated genome editing will be used for knock-ins, a repair template should be prepared and delivered into the cells along with the gRNA and Cas protein. For HDR, the template is usually a double-stranded DNA molecule with the DNA insert flanked by sequences that are homologous to the target locus. The template is often provided as a circular plasmid, but there are also approaches using linear templates (for example (Stinson et al., 2020)). For NHEJ, the transgene is often injected as circular plasmid DNA bearing one or two gRNA targets, so that it will be linearised within the cell by the Cas9/gRNA complex (Adli, 2018). There are also variants of these knock-in methods that use single-stranded DNA as a template (Bai et al., 2020). Recently, CRISPR-based prime editing was developed based on a Cas9 fusion protein with reverse transcriptase activity, in which the repair template is a short RNA molecule fused to the sgRNA (Anzalone et al., 2019).

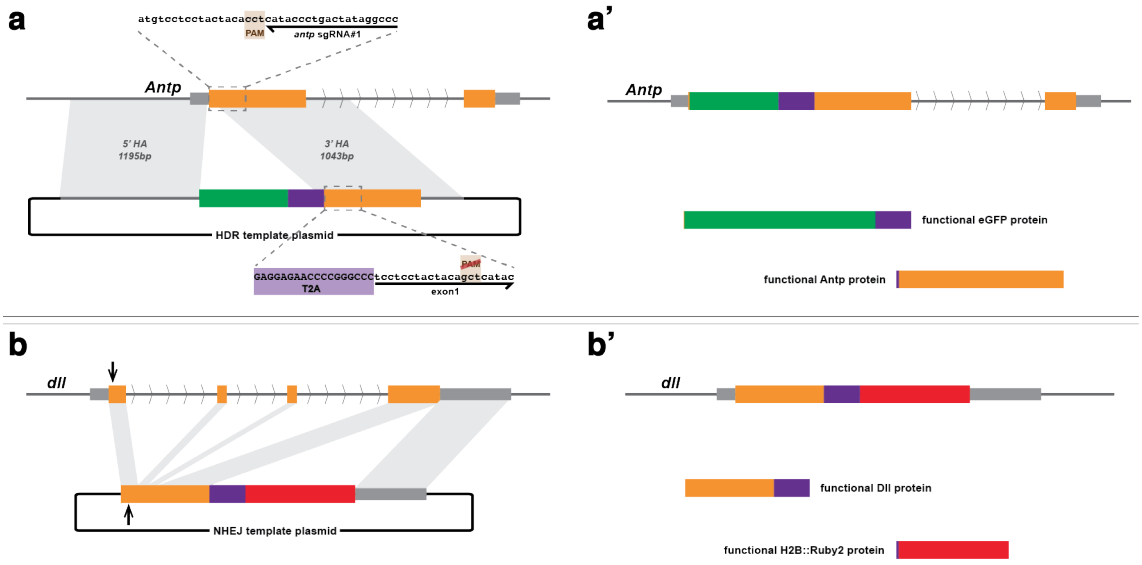


Figure 4.3: CRISPR-mediated knock-ins performed in two Parhyale genes
(a) Illustration of Antp genomic locus with exons (orange bars) and UTRs (gray bars) labeled. The sgRNA is targeting a sequence just downstream of the start codon. The HDR repair template plasmid carries two homology arms homologous to sequences upstream and downstream of the target locus, surrounding a reporter cassette bearing coding sequences of eGFP (green) and the T2A ribosome skipping peptide (purple). The PAM sequence on the HDR plasmid is altered. **(a')** Once the transgene is inserted upstream of the Antp coding sequence, translation of this locus yields two independent peptides: eGFP and Antp (modified from (Serano et al., 2016)) **(b)** Illustration of Dll genomic locus with exons (orange bars) and UTRs (gray bars) labeled. The sgRNA is targeting a sequence downstream of the start codon (black arrow). The NHEJ repair plasmid carries the complete coding sequence of Dll followed by the T2A (purple) and eGFP (green) coding sequences. The repair plasmid bears the target sequence of the sgRNA (black arrow) therefore it is linearised after injection by the CRISPR machinery. **(b')** Once the transgene is inserted, translation of this locus yields two independent peptides: Dll and eGFP (modified from (Kao et al., 2016)). The locus retains the plasmid backbone and the endogenous Dll sequences downstream of the knocked-in construct (not shown).

CRISPR-mediated genome editing in *Parhyale*

Parhyale embryos are accessible from the 1-cell stage, therefore the transgenesis reagents can be directly injected into the early zygote. There are four published studies using CRISPR-mediated genome editing in *Parhyale*: two reporting knock-outs (Clark-Hachtel and Tomoyasu, 2020; Martin et al., 2016) and two knock-ins (Kao et al., 2016; Serano et al., 2016). In these studies, spCas9-NLS protein and canonical sgRNAs were used. For CRISPR-mediated knock-ins, a reporter cassette containing eGFP and the T2A ribosome skipping sequence was inserted in the *Parhyale Dll* (Kao et al., 2016) and *Antp* (Serano et al., 2016) gene loci (figure 4.3). The *Antp* knock-in relied on an HDR-based approach, using a repair plasmid carrying the T2A ribosome skipping sequence followed by the eGFP coding sequence, surrounded by ~1kb homology arms corresponding to sequences upstream and downstream of the sgRNA target region. The PAM sequence was altered in the repair plasmid to prevent it from being targeted by CRISPR, which is essential for HDR mediated knock-ins. The repair construct was

designed such that, once the reporter cassette is inserted at the 5' end of *Antp*, separate Antp and eGFP proteins are produced from the target locus (figure 4.3a-a'). For the *Dll* knock-ins, the NHEJ pathway was used to insert a plasmid carrying the complete coding sequence of *Dll* followed by the T2A ribosome skipping sequence and the eGFP coding sequence at the 5' end of the *Dll* locus. The PAM sequence was left intact in the repair plasmid to promote linearization of the plasmid (figure 4.3b-b').

Results and Discussion

Antp knock-in as a positive control

The first step of CRISPR-mediated knock-ins (KI) is the design, cloning, or synthesis of the sgRNA, and cloning of the repair template plasmid. To establish the pipeline of CRISPR-mediated Knock-ins and make sure it works in my hands I decided to repeat a knock-in performed previously in the *Parhyale Antp* gene (Serano et al., 2016) as a positive control. Arnaud Martin kindly provided the HDR template plasmid, I cloned and synthesized the sgRNA used in that study. I used the same reagents and similar microinjection protocol as Serano *et al*. and screened the late embryos for eGFP expression. I obtained two eGFP-expressing embryos out of 256 screened embryos, but both of them died before hatching. The knock-in ratio was comparable to the original study, 0.8% in my injections vs. 1% in the original study, but I observed a significantly lower rate of survival among the injected embryos, 23% vs. 70% (Table 4.1).

Table 4.1: Quantifications of the injections for *Antp* knock-in.

***Antp* HDR**	*Injection Date*	*# of embryos injected*	*# of embryos screened*	*# of positives embryos*
sgRNA #1	14.06.2016	80	12	0
	15.06.2016	146	17	0
	16.06.2016	92	10	0
	17.06.2016	103	10	0
	20.06.2016	195	43	0
	21.06.2016	41	2	0
	22.06.2016	367	114	1
	23.06.2016	98	48	1
	Total	**1122**	**256** (22.8%)	**2** (0.8%)

For *Parhyale* transgenesis, typically, the embryos are injected at the one-cell stage, which has a duration of ~4 hours at 26˚C. In the original *Antp* knock-in study, the positive animals had been injected in the early stages of the one-cell stage (Arnaud Martin, personal communication). For this reason, I initially injected embryos within ~1-hour post-fertilization. However, the embryos are very fragile during this stage and the majority of deaths (~75%) occurred the first hours following the injection. Therefore, I decided to inject embryos around 2-3 hours post-fertilization in subsequent experiments.

When an embryo is injected for CRISPR knock-ins but has no eGFP expression, there are three possible scenarios. First, the CRISPR machinery did not

target the genome, so no DSB was introduced. Second, a DSB was introduced but it was repaired without inserting the eGFP from the repair template (most likely through NHEJ). Third, the transgene may have been inserted, but eGFP is not expressed or the levels of expression are too low to detect. For the *Antp* KI experiments, I already know that once the insertion is made the eGFP expression levels are detectable. To test the other two possibilities, I performed the T7 endonuclease assays on the injected embryos that did not show any eGFP fluorescence. The aim of this experiment was to assess whether the injected Cas9 nuclease complex was active in the different sets of injections.

T7 endonuclease assay can be used to detect the presence of polymorphic indels in the CRISPR target site (Mashal et al., 1995). In these experiments, I pooled 10-50 injected embryos that had shown no eGFP expression and extracted their genomic DNA. Using PCR, I amplified a 610 bp long DNA fragment around the sgRNA target site and purified the amplicon. Denaturing and then annealing these DNA fragments leads to the formation of heteroduplexes with nucleotide mismatches, resulting from polymorphisms (indels) in the target sequence. Then I treated the samples with T7 endonuclease, an enzyme that cuts the double-stranded DNA specifically in mismatched regions, and ran these reactions on an agarose gel. The results show that many of the pools of eGFP-negative embryos have indels in the *Antp* sgRNA target locus since the T7 endonuclease treatment cut the DNA into two fragments of ~250 and ~350 bp as expected (Figure 4.4).

Designing sgRNAs for new targets

The results of *Antp* experiments showed that CRISPR-mediated knock-in experiments are working in my hands. To target the genes relevant to my project, first I wanted to explore the best way to design efficient sgRNAs for *Parhyale*. In two recent studies seven *Parhyale* homeobox genes, *abdA*, *AbdB*, *Antp*, *Dfd*, *Dll*, *Scr*, and *Ubx*, were targeted by CRISPR knock-out. For each gene, two sgRNAs were used and their efficiencies were assessed by quantifying the resulting mutant phenotypes (Kao et al., 2016; Martin et al., 2016). I tested eight existing sgRNA prediction algorithms (Doench et al., 2014; Moreno-Mateos et al., 2015; Park et al., 2015; Prykhozhij et al., 2015; R et al., 2015; Wong et al., 2015; Xu et al., 2015) on the above-mentioned genes. I evaluated the predictions of these algorithms by two criteria: whether the predicted sgRNAs included those that had been proven to experimentally, and whether the prediction scores correlated with the relative efficiencies of the sgRNAs that were tested (supplementary table S4.1). Two of these tools, CRISPRScan, and sgRNAdesigner, appeared to perform better than the others. I decided to use

CRISPRscan for two reasons: first, this sgRNA scoring algorithm has been developed with the data coming from both vertebrate and invertebrate species, and second, it can also predict non-canonical sgRNAs, which increases the chances of finding efficient sgRNAs in a relatively small genomic locus.

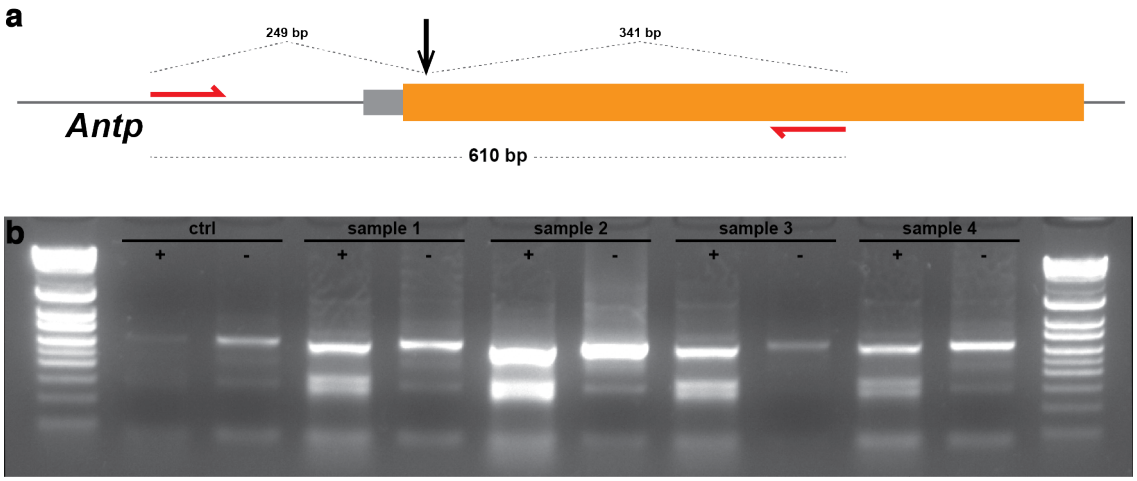


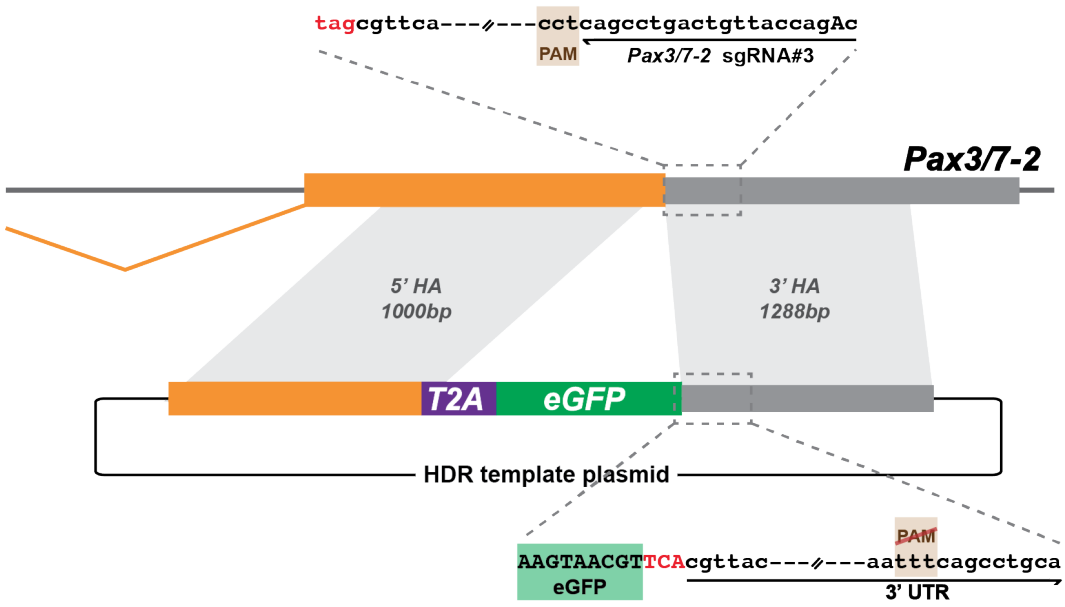
Figure 4.4: T7 endonuclease assay on animals injected for the *Antp* knock-ins.
(a) 5' of the *Antp* genomic locus showing the first intron with the 5'UTR (gray) and the CDS (orange). Two primers (red arrows) complementary to 249 bp upstream and 341 bp downstream sequences of the sgRNA target site (black arrow), amplifying a 610 bp DNA. **(b)** The PCR fragments from uninjected control animals (first two lanes) and animals injected with *Antp* CRISPR-mediated knock-in reagents (lanes 2-9) visualized on an agarose gel. The lanes labeled as (-) were not treated with T7 endonuclease and they show a single band just above the 600 bp band of the ladder. The lanes labeled as (+) were treated with T7 endonuclease and they have additional two bands around 250 bp and 350 bp size, except the control embryos.

Targeting *Pax3/7-2*

As a first target, in collaboration with my colleague Marco Grillo, I focused on the marker gene for the satellite-like cells, *Pax3/7-2*. Satellite-like cells are progenitors of the muscles during limb regeneration (Konstantinides and Averof, 2014) and Marco Grillo was investigating the cellular dynamics of myogenesis during limb regeneration. I designed a sgRNA targeting a sequence 100 bp downstream of the stop codon, within the 3'UTR of the gene. I co-injected this sgRNA with Cas9 protein in a small number of embryos and later extracted genomic DNA from these injected embryos to perform a T7 endonuclease assay, which confirmed that the sgRNA is functional. I then generated an HDR template plasmid that contains homology arms of ~1 kb upstream and ~1,3 kb downstream of the stop codon, surrounding a T2A-eGFP transgene. I modified the PAM of the target sequence on the plasmid so that it will not be targeted (figure 4.5a). I co-injected this plasmid, the sgRNA, and Cas9 protein in ~2500 embryos and screened the surviving ~1000 embryos during the final stages of embryogenesis for eGFP expression. I found no eGFP positive embryos (figure 4.5b). I extracted the genomic DNA from these negative embryos and confirmed that a DSB was introduced to the target genomic locus using the T7 endonuclease assay. Next, to

test whether some of these embryos might carry an integration giving undetectable levels of eGFP expression, I performed a PCR on the genomic DNA using a forward primer binding immediately upstream of the 5' homology arm and a reverse primer on the T2A coding sequence. The primers were individually confirmed to be functional first. This pair is expected to amplify a 1200 bp fragment only if the KI construct has integrated into the target locus. The absence of a PCR band in 12 samples of pooled embryos (including 1062 embryos in total) indicated that the knock-in did not take place.

a



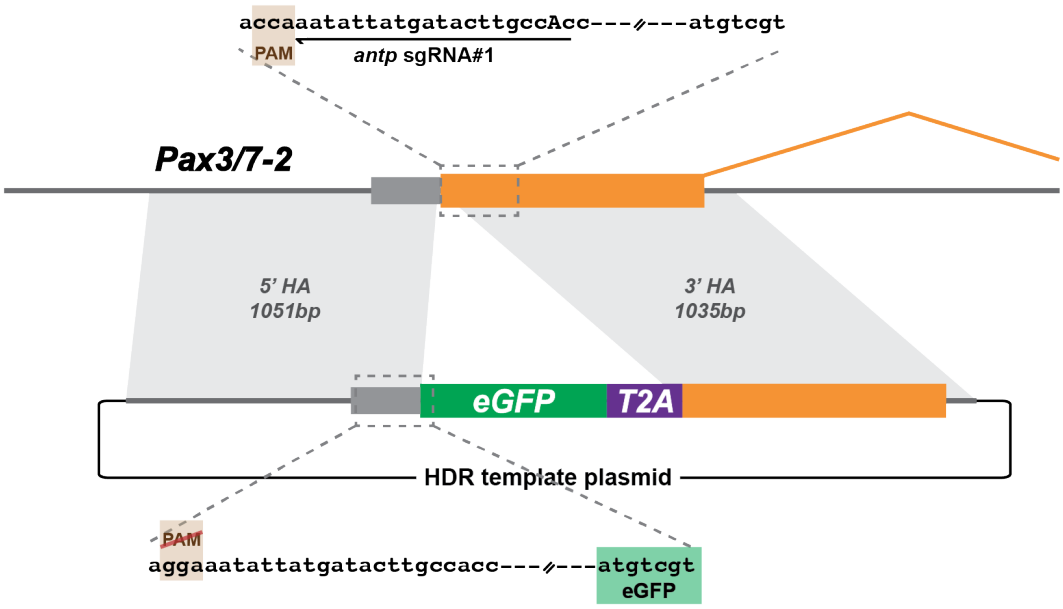
b

***Pax3/7-2* HDR**	*Injection Date*	*# of embryos injected*	*# of embryos screened*	*# of positive embryos*
sgRNA #3	05.07.2016	96	35	0
	06.07.2016	125	55	0
	07.07.2016	242	134	0
	08.07.2016	365	125	0
	09.07.2016	320	101	0
	15.08.2016	136	58	0
	16.08.2016	297	162	0
	17.08.2016	315	105	0
	18.08.2016	279	130	0
	19.08.2016	318	157	0
	Total	**2462**	**1062** (43%)	**0**

Figure 4.5: CRISPR-mediated knock-in to the 3' of the *Pax3/7-2* locus through HDR.
(**a**) Illustration of the 3' of the *Pax3/7-2* genomic locus indicating the coding sequence (orange bar) and the 3'UTR (gray bar). The sgRNA is targeting 100 bp downstream of the stop codon (red letters), and the repair template plasmid has 1000 bp upstream and 1288 bp downstream sequences of the stop codon as homology arms, surrounding a T2A-eGFP reporter cassette. The plasmid has a modified PAM sequence for the sgRNA. (**b**) Quantifications of the injected embryos screened for eGFP expression.

In a second attempt, I tried to perform KI into the 5' of the *Pax3/7-2* gene. To avoid deleterious frameshift mutations, I designed a sgRNA targeting the 3'UTR, 102 bp upstream of the start codon, of the gene. After confirming the activity of this sgRNA, via injection in a small number of embryos and a T7 endonuclease assay, I designed an HDR template plasmid carrying homology arms that extend 1051 bp upstream and 1035 bp downstream of the start codon, with an eGFP-T2A construct in between (figure 4.6a). I co-injected these two constructs with the Cas9 protein in 1576 embryos and screened 819 surviving embryos at the S28 stage for eGFP fluorescence. I found no eGFP expressing embryos (figure 4.6b). T7 endonuclease assays and PCR analysis, as described above, revealed that a DSB was introduced in the genome of the injected embryos but the transgene was not inserted.

a



b

Pax3/7-2 HDR	*Injection Date*	*# of embryos injected*	*# of embryos screened*	*# of positive embryos*
sgRNA #1	02.08.2016	148	56	0
	03.08.2016	252	132	0
	04.08.2016	315	150	0
	05.08.2016	421	209	0
	06.08.2016	440	272	0
	Total	**1576**	**819** (52%)	**0**

Figure 4.6: CRISPR-mediated knock-in to the 5' of the *Pax3/7-2* locus through HDR.
(a) Illustration of the 5' of the *Pax3/7-2* genomic locus indicating the 5'UTR (gray bar) and the coding sequence (orange bar). The sgRNA is targeting 102 bp downstream of the start codon. The repair template plasmid carrying 1051 bp upstream and 1035 bp downstream sequences of the stop codon as homology arms, surrounding a T2A-eGFP reporter cassette. The plasmid has a modified PAM sequence for the sgRNA. (**b**) Quantifications of the injected embryos screened for eGFP expression.

Both attempts to insert a transgene into the *Pax3/7-2* locus have failed. In both cases, I managed to introduce a DSB in the target locus but CRISPR-mediated KI through the HDR pathway was unsuccessful. The efficiency of KI is strongly correlated to the efficiency of the sgRNAs and is target-dependent (Liu et al., 2019). Instead of trying new sgRNAs for this gene, I decided to move on to genes that are more relevant for my project on sensory organ regeneration.

Selecting putative sensory organ markers

Arthropod peripheral sensory organs consist of several cell types. When I started this project, we had no information on the molecular/genetic basis of *Parhyale* sensory organ development. At that time, I also did not have single-cell RNAseq datasets that could help to identify genes expressed in the sensory organ cells. Therefore, I followed a candidate gene approach and screened the literature to find marker genes identified in other species. I aimed to pursue genes that have been shown to mark a specific cell type in more than one species. Doing so, I practically limited my search to two cell types, neurons, and glia, since there are not many studies focusing on the other cells of the sensory organs, beyond *Drosophila melanogaster*. I focused on three candidate marker genes for neurons, *futsch*, *brp*, and *elav*; and two for glia, *gcm* and *repo*.

Initially, I focused on identifying the putative neuronal marker genes. Futsch is a microtubule-binding protein that is important for synapse and neuromuscular junction formation, whose function is conserved from mammals to insects (Hummel et al., 2000). Brp is a cytoskeletal protein required for the release of synaptic vesicles. The function and also the neuronal expression of both genes is conserved in mammals, nematodes, and flies (Wagh et al., 2006). Finally, Elav is an RNA binding protein required for CNS development and function. *Elav* gene belongs to the RRM family, which includes different paralogues in different species. Even though the evolutionary conservation of the neuronal expression is debated, there are neuron-specific members of this family in many species (Colombrita et al., 2013).

I identified the *Parhyale* orthologues of these genes in the genome and transcriptomes (table 4.2), cloned corresponding gene fragments, and performed RNA *in situ* hybridization experiments in stage 21-23 embryos, to examine whether they are expressed in developing neurons (*in situ* hybridization is not yet established in adult limbs, due to penetration issues caused by the chitinous exoskeleton). In embryos, I could detect strong *brp* expression in the CNS and in small clusters of cells on the legs, antennae, and also on the dorsal parts of the thorax, which matches the expected distribution of neuronal cells. *Futsch* is also highly expressed in the CNS but the

expression is missing on the limbs. The lack of *futsch* expression on the limbs could be the result of the staging: we cannot perform *in situ* hybridization on late embryos since they start to deposit their cuticle, which gives a strong background signal and prevents us from deciphering the expression pattern of the genes. Lastly, *elav* is also highly expressed in CNS and in several cells on the limbs. This expression pattern suggests that the *elav* paralogue that I cloned is expressed in non-neuronal cell types (figure 4.7).

Table 4.2: The list of *Parhyale* transcripts targeted in this study.

gene name	*source*	*transcript ID*
Antp	NCBI GenBank	JQ952581
Pax3/7-2	trancriptome r3.1 (Kao et al., 2016)	c32936_g1_i1
brp	Hunt+Rosata (Hunt et al., 2019)	PH.k21.comp4294_seq9
futsch	trancriptome r3.1 (Kao et al., 2016)	c10291_g1_i1
elav	trancriptome r3.1 (Kao et al., 2016)	c146164_g1_i1
gcm	trancriptome r3.1 (Kao et al., 2016)	c224532_g1_i1
repo	trancriptome r3.1 (Kao et al., 2016)	c37156_g1_i2
sens	trancriptome r3.1 (Kao et al., 2016)	c230020_g1_i1
cut	Hunt+Rosata (Hunt et al., 2019)	PH.k21.comp1009_seq18

Next, I explored the two putative glial marker genes. *Gcm* is a zinc-finger transcription factor essential for gliogenesis both in the CNS and PNS in *Drosophila* (Van De Bor et al., 2000), but the role of the gene is very diverse in vertebrates and unrelated to the nervous system (Hanaoka et al., 2004). Even though the data from vertebrates are not very encouraging, I wanted to check whether the expression pattern of *gcm* is conserved among arthropods. And finally, *repo* is a homeobox transcription factor required for glial fate in *D. melanogaster* and expressed in glial cells in *T. castaneum* (Trebels et al., 2020). I identified the homologs of these genes in the *Parhyale* genome/transcriptomes but, unlike the results I obtained with the putative neuronal markers, I did not succeed in obtaining good *in situ* hybridization signals for *gcm* and *repo*. Therefore, I decided to proceed with CRISPR-mediated KI experiments focusing on a neuronal marker gene.

Targeting the neuronal marker *futsch*

Based on the embryonic expression patterns, *brp* and *futsch* were my best candidates as neuronal markers. Unfortunately, I found only partial gene sequences for these candidates in our transcriptome assemblies (Kao et al., 2016). For *brp* it was not possible to determine the 5' or the 3' ends of the coding sequence. For *futsch*, I found a partial transcript including the stop codon and the 3'UTR. Since inserting a transgene

in the middle of the coding sequence is likely to disrupt the function of a gene, I chose the 3' end of the *futsch* coding sequence as my target for a CRISPR KI.

I designed two sgRNAs targeting the 3'UTR of *futsch*: one is located 87 bp and the other 28 bp downstream of the stop codon. I confirmed the functionality of these sgRNAs via T7 endonuclease assay. Then, I designed an HDR template plasmid containing homology arms of 1000 bp upstream and 1103 bp downstream of the stop codon, with disrupted PAM sequences, and the T2A and eGFP coding sequences fused in frame with the *futsch* coding sequence (figure 4.8a). I injected ~2500 embryos with each of the sgRNAs and screened ~1350 survivors during the late embryonic stages. I did not find any eGFP expressing embryos (figure 4.8b). Further analyses with the T7 endonuclease assay and PCR (as described earlier) confirmed that the sgRNAs were active, but showed no evidence of KI.

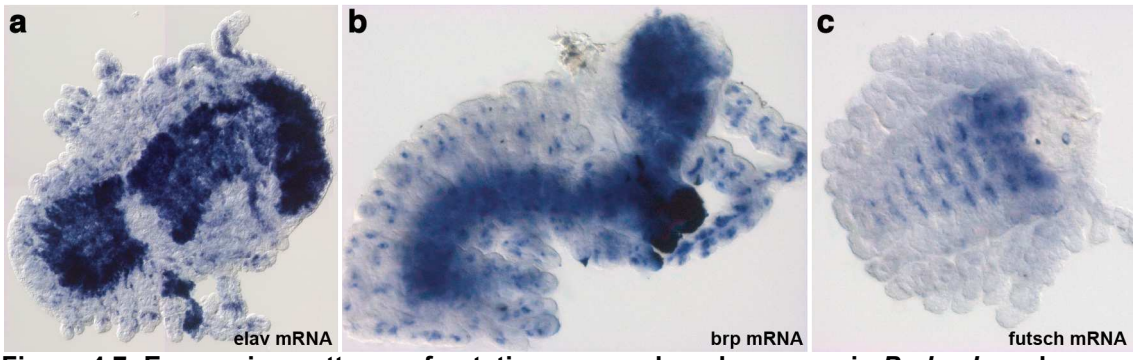


Figure 4.7: Expression patterns of putative neuronal marker genes in *Parhyale* embryos. RNA *in situ* hybridization reveals that **(a)** the *elav* transcript is expressed predominantly in the CNS of stage S21 embryos, and in clusters of cells on the body and developing limbs. **(b)** Likewise, there are small clusters of cells on the developing legs, antennae, and body segments of stage S23 embryos expressing the *brp* transcript. **(c)** The *futsch* transcript is only present in the CNS cells of stage S21 embryos.

Given the activity of sgRNAs (as revealed by T7 endonuclease assays) by the lack of any evidence for homology-dependent repair, I decided to switch to an NHEJ-mediated KI approach. I designed five more sgRNAs targeting sequences shortly upstream of the stop codon, and I co-injected these along with the plasmid I had generated as a template for HDR. In this plasmid, the PAM sequences of the new targets were intact, so that the plasmid was also expected to be cleaved in addition to the genomic target. The cleaved plasmid could then be used for NHEJ mediated KI (figure 4.9a), as described in previous studies (reviewed in (Gilles and Averof, 2014)). I injected ~3500 embryos for this NHEJ approach, using single sgRNAs, but again I obtained no eGFP expressing embryos (figure 4.9b). Analysis of the genomic DNA of the injected embryos, as performed previously, affirmed that DSB was introduced for every sgRNA, but no KI events could be detected either by fluorescence or by PCR.

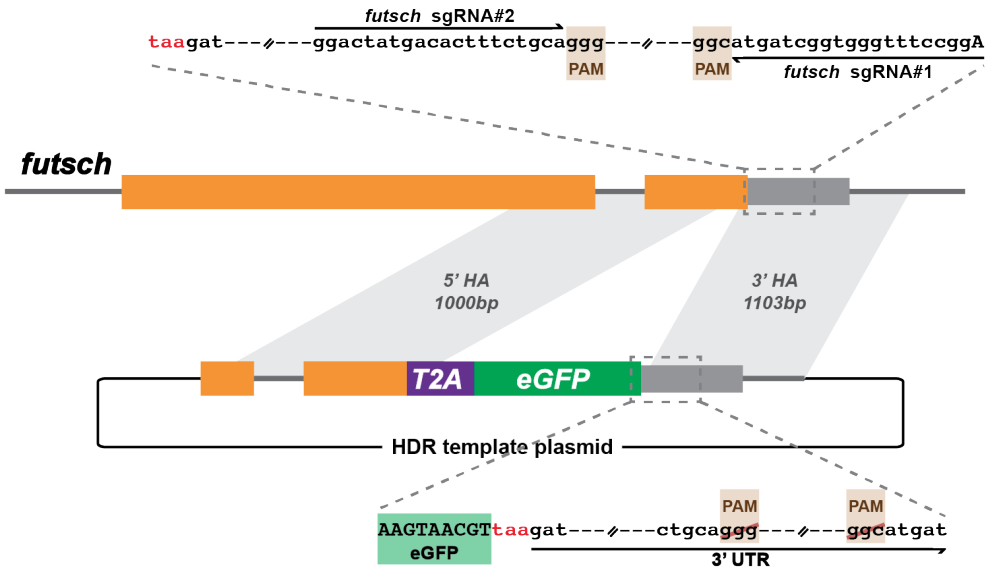
Testing sgRNA efficiencies on *cut*, *sens,* and *elav*

Both for *futsch* and *Pax3/7-2* genes, CRISPR-mediated KI attempts have failed. In every experiment, I managed to introduce DSBs but the transgene was not inserted in the genome. This suggests that the problem is unlikely to be due to technical problems, like RNA degradation or inactivity of the Cas9 nuclease.

In both of the previously reported *Parhyale* CRISPR-mediated KI examples, highly efficient sgRNAs were used, yielding knock-out phenotypes in 70% of individuals injected with the *Antp* sgRNA (Martin et al., 2016) and in 76% of individuals injected with the *Dll* sgRNA (Kao et al., 2016). In the experiments I described above, I did not quantify the efficiency of the sgRNAs that I used (the T7 endonuclease assay does not provide accurate quantification of cleavage efficiency and I was not able to assess the frequency of knockout phenotypes). As the cleavage efficiency of sgRNAs could be decisive for obtaining KI at detectable frequencies, I decided to take an experimental approach in which the efficiency of CRISPR-mediated KO phenotypes can be measured. I searched for target genes that manifest an easily recognizable phenotype when mutated, to be able to assess the efficiency of sgRNAs causing deleterious mutations on those genes. My approach was to find very efficient sgRNAs, that give a very high frequency of KO phenotypes, and then to design KI experiments based on these sgRNAs.

The most striking phenotype that is easy to score in peripheral sensory organs is the loss of the external sensilla (the bristles of flies or the setae of *Parhyale*). Knocking out the *sens*, which is required to maintain SOP identity (Blochlinger et al., 1990; Jafar-Nejad et al., 2003), and *cut*, which is required for the development of external sensilla (Blochlinger et al., 1990), genes of *D. melanogaster* results in this phenotype. I identified the *Parhyale* homologs of these genes in the *Parhyale* transcriptome and probed their expression patterns in embryos by RNA *in situ* hybridization. Both genes are expressed in small clusters of cells in the developing limbs (figure 4.10), which could be the cells of developing sensory organs. After confirming these genes are present in adult limb transcriptomes, which could suggest their persistent expression in terminally differentiated cells, I decided to target them with CRISPR.

a

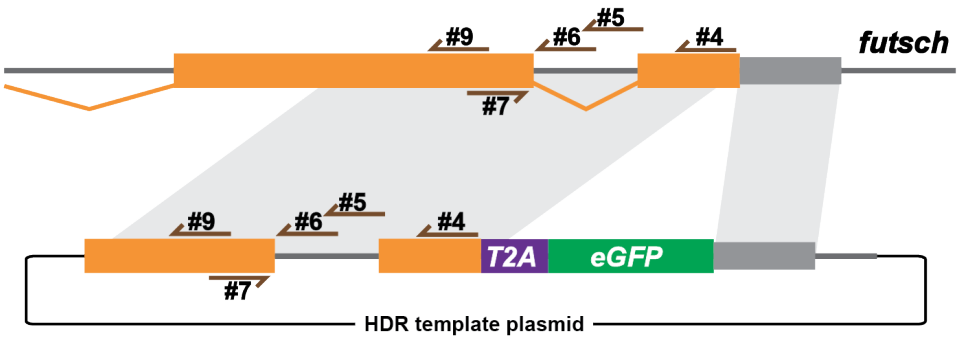


b

futsch HDR	Injection Date	# of embryos injected	# of embryos screened	# of positive embryos
sgRNA #1	29.08.2016	336	71	0
	30.08.2016	224	117	0
	31.08.2016	218	95	0
	01.08.2016	217	107	0
	05.09.2016	660	410	0
	06.09.2016	350	175	0
	07.09.2016	492	190	0
	08.09.2016	150	71	0
	Total	**2647**	**1236** (47%)	**0**
sgRNA #2	11.10.2016	261	142	0
	12.10.2016	317	135	0
	13.10.2016	193	86	0
	14.10.2016	241	162	0
	18.10.2016	245	114	0
	19.10.2016	541	296	0
	20.10.2016	505	268	0
	21.10.2016	358	152	0
	Total	**2661**	**1355** (51%)	**0**

Figure 4.8: CRISPR-mediated knock-in to the 3' of the *futsch* locus through HDR.
(a) Illustration of the 3' of the futsch genomic locus with the 3'UTR (gray bar) and the coding sequence (orange bars) are indicated. The sgRNAs are targeting 28 and 87 bp downstream of the stop codon (red letters). The repair template plasmid carrying 1000 bp upstream and 1103 bp downstream sequences of the stop codon as homology arms, surrounding a T2A-eGFP reporter cassette. The plasmid has modified PAM sequences. (b) Quantifications of the injected embryos screened for eGFP expression.

a



b

futsch NHEJ	Injection Date	# of embryos injected	# of embryos screened	# of positive embryos
sgRNA #4	31.01.2017	24	22	0
	01.02.2017	98	68	0
	02.02.2017	90	53	0
	07.02.2017	303	88	0
	08.02.2017	79	58	0
	09.02.2017	150	86	0
	Total	**744**	**375** (50%)	**0**
sgRNA #5	18.04.2017	65	31	0
	19.04.2017	214	99	0
	20.04.2017	196	76	0
	21.04.2017	244	113	0
	Total	**719**	**319** (44%)	**0**
sgRNA #6	02.05.2017	182	103	0
	03.05.2017	231	116	0
	04.05.2017	322	160	0
	Total	**735**	**379** (52%)	**0**
sgRNA #7	17.05.2017	66	12	0
	18.05.2017	232	161	0
	19.05.2017	397	198	0
	Total	**695**	**371** (53%)	**0**
sgRNA #9	31.05.2017	122	84	0
	01.06.2017	130	51	0
	02.06.2017	254	114	0
	03.06.2017	261	129	0
	Total	**767**	**378** (49%)	**0**

Figure 4.9: CRISPR-mediated knock-in to 3' of the *futsch* locus through NHEJ.
(a) Illustration of the 3' of the *futsch* genomic locus with the 3'UTR (gray bar) and the coding sequence (orange bars) are indicated. Five sgRNAs designed to target the last 1 kb of *futsch* coding sequence, and the repair template plasmid carrying 1 kb upstream of the stop codon and the 3' UTR of the *futsch* surrounding a T2A-eGFP reporter cassette. **(b)** Quantifications of the injected embryos screened for eGFP expression.

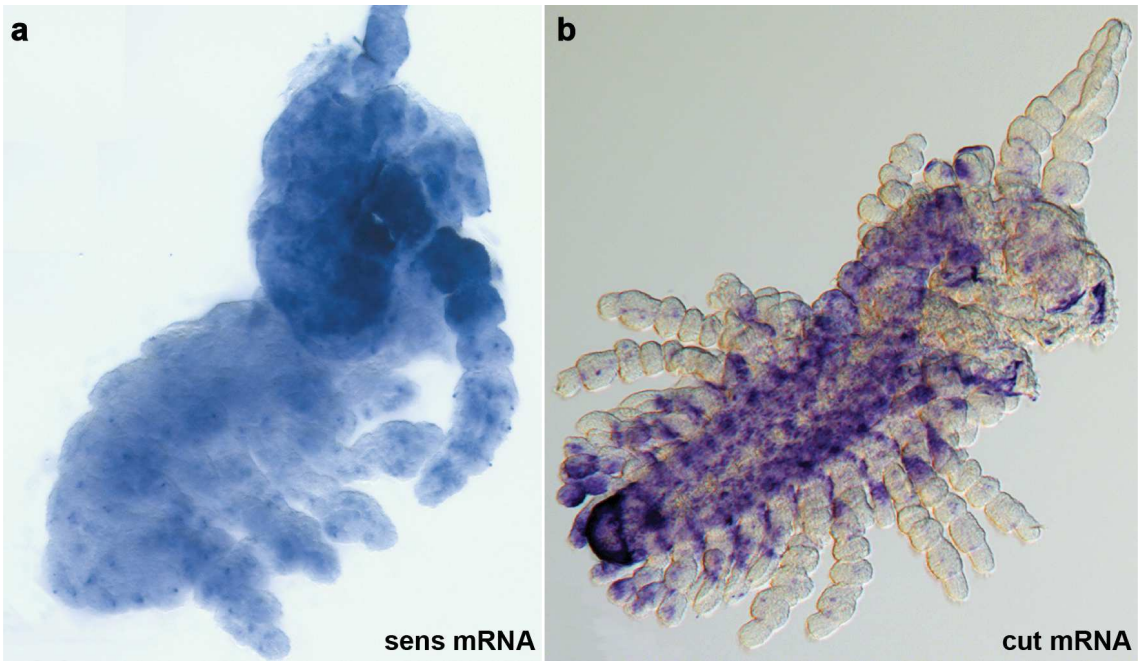
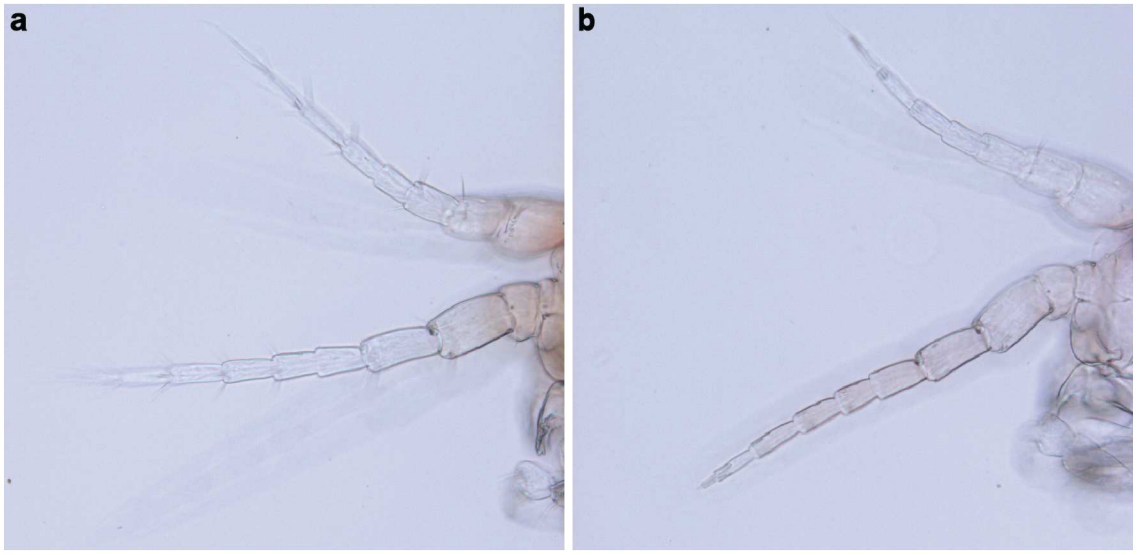
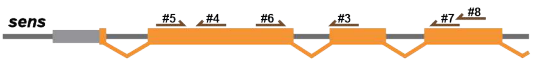


Figure 4.10: Expression of *sens* and *cut* in *Parhyale* embryos.
RNA *in situ* hybridization revealed that **(a)** *sens* transcript is expressed in the head, and in small clusters of cells in the body and the limbs of stage S22 embryos. **(b)** *cut* transcript is highly expressed in the VNC and gills of stage S23 embryos, there are also isolated cells expressing it on the legs.

Parhyale Cut protein is ~1300 aa long and the coding sequence (CDS) of the gene is spanning five exons. The first three exons are relatively small and encode the first ~100 aa of the protein. I designed two sgRNAs targeting the third, and one targeting the first exon (supp figure S4.1a). Sens protein is ~1000 aa long and the CDS is spanning seven exons, and the well-conserved Zn finger domain is positioned between 760-820 aa. Thus, I designed six sgRNAs targeting the first half of the CDS (figure 4.11). I co-injected each of these sgRNAs along with Cas9 protein and prepared cuticle preps from injected hatchlings and from non-injected controls. I screened the animals for the loss of setae mainly focusing on the antennae, which have very prominent and regular arrays of setae. All the sgRNAs tested, except one, gave the loss of setae phenotype in the antennae (figure 4.11). The frequency of phenotypes was low: the most efficient of these sgRNAs showed loss of antennal setae in only 12% of surviving hatchlings. Mortality among the injected animals was ~65%, compared with ~50% in the *futsch* knock-in injections. (figure 4.11). Since I could not find a sgRNA with a KO efficiency comparable to that of the *Antp* and *Dll* sgRNAs, I decided not to continue with the KI experiments for these two genes.



c



sens knock-out	Injection Date	# of embryos injected	# of embryos screened	# of mutant embryos
sgRNA #3	27.11.2017	138	49	14
	28.11.2017	348	112	14
	30.11.2017	170	86	0
	Total	**656**	**247** (38%)	**28** (11%)
sgRNA #4	05.12.2017	149	33	4
	06.12.2017	171	34	4
	Total	**320**	**67** (21%)	**8** (12%)
sgRNA #5	06.12.2017	91	17	0
	07.12.2017	132	19	2
	Total	**223**	**36** (16%)	**2** (6%)
sgRNA #6	13.07.2018	178	98	13
	15.07.2018	93	36	3
	20.07.2018	151	52	6
	Total	**422**	**186** (44%)	**22** (12%)
sgRNA #7	16.07.2018	61	17	0
	19.07.2018	246	85	0
	27.07.2018	299	102	0
	Total	**606**	**204** (34%)	**0**
sgRNA #8	17.07.2018	125	52	4
	30.07.2018	269	126	9
	31.07.2018	313	173	15
	Total	**707**	**351** (50%)	**28** (8%)

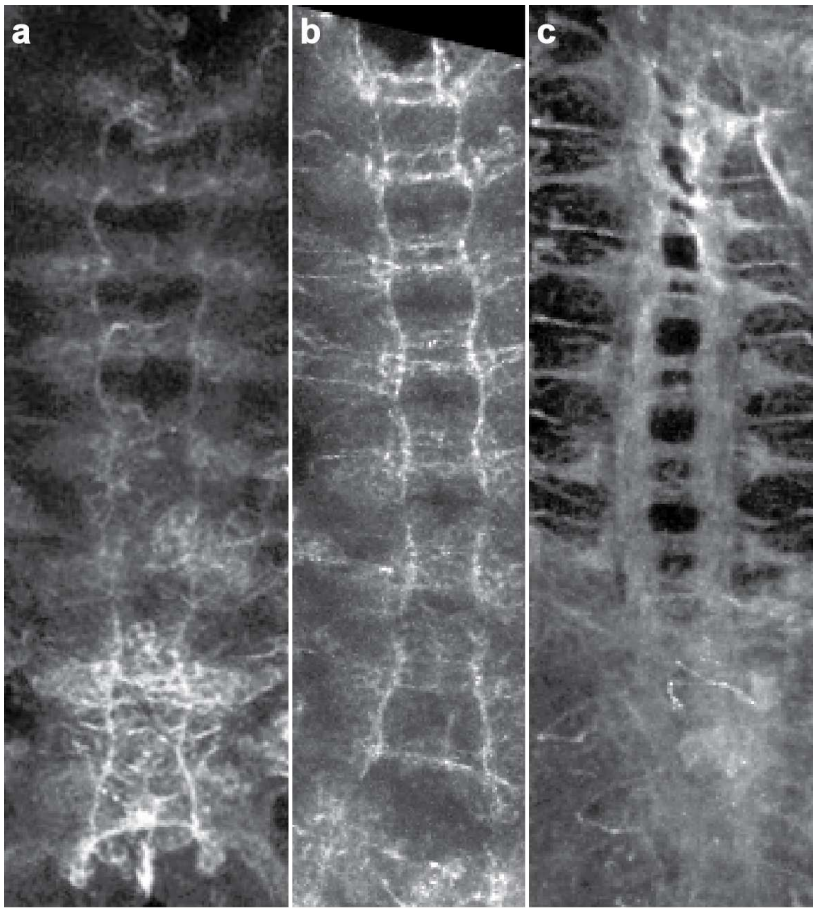
d

cut knock-out	Injection Date	# of embryos injected	# of embryos screened	# of mutant embryos
sgRNA #1	26.12.2017	62	17	0
	27.12.2017	93	22	2
	28.12.2017	114	32	6
	29.12.2017	125	45	7
	Total	**394**	**116** (29%)	**15** (13%)
sgRNA #2	01.02.2018	102	23	5
	02.02.2018	147	35	7
	03.02.2018	210	107	6
	Total	**459**	**165** (36%)	**18** (11%)
sgRNA #7	13.02.2018	113	0	0
	14.02.2018	119	51	0
	15.02.2018	207	101	0
	16.02.2018	173	105	0
	Total	**612**	**211** (34%)	**0**

Figure 4.11: Knock-out of *sens* and *cut* leads to loss of setae.
(a) Control *Parhyale* hatchlings have several seta on their antennae. **(b)** These are lost in individuals injected with sgRNAs targeting either *sens* or *cut* genomic loci (here shown for *sens* KO). Illustrations of the 5' of the **(c)** *sens* and **(d)** *cut* genomic loci, indicating the 5'UTR (gray bar), the coding sequence (orange bars), and the sgRNAs targeting the CDS (brown arrows). The tables show the frequencies of loss-of-setae phenotype among the injected animals.

In my final CRISPR experiments, I tried to knock-out the *elav* gene. Expression in *Parhyale* embryos suggested that this gene might be expressed in the nervous system but there are possibly other cell types in the developing limbs expressing *elav* (figure 4.7a). Even though it would not be ideal to label multiple cell types for tracking the progenitors during limb regeneration, I thought I could still distinguish the neurons from the other cells by their morphology. Therefore, marking the *elav* expressing cells could be still useful to identify the sensory neurons. Null mutations in *elav* cause morphological defects in the embryonic ventral nerve cord in *D. melanogaster* (Campos et al., 1985). I designed two sgRNAs targeting the coding sequence of the

elav gene and confirmed they are inducing a double-strand break via the T7 endonuclease assay. After injecting ~250 embryos with each sgRNA along with Cas9 protein, I dissected the surviving embryos and stained them with an antibody for acetylated tubulin antibody to reveal the structure of their ventral nerve cord. I could not detect any morphological defects (figure 4.12).



d

elav knock-out	Injection Date	# of embryos injected	# of embryos screened	# of mutant embryos
sgRNA #3	05.02.2018	56	12	0
	06.02.2018	85	23	0
	07.02.2018	122	29	0
	Total	**263**	**64** (24%)	**0**
sgRNA #4	08.02.2018	104	26	0
	09.02.2018	72	19	0
	10.02.2018	58	17	0
	Total	**234**	**62** (26%)	**0**

Figure 4.12: Injecting embryos with *elav* sgRNAs does not alter the morphology of their VNC.
The VNC morphologies of **(a)** uninjected control, **(b)** *elav* sgRNA#3 and **(c)** *elav* sgRNA#4 injected stage S23 embryos were revealed by immunostaining for acetylated tubulin; there are no apparent differences. **(d)** Illustration of the 3' of the *elav* genomic locus, indicating the 3'UTR (gray bar), the coding sequence (orange bars), and the target sequences of the two sgRNAs (brown arrows). The tables indicate the screening results of the injected animals.

Overall lessons from CRISPR experiments

In this study, I tried to establish CRISPR/Cas9 mediated transgenesis in *Parhyale* for the generation of animals with cell type-specific markers. I managed to generate putative knock-in embryos in the initial experiments I performed on the *Antp* gene, yet, I could not get the knock-ins to work for *Pax3/7*, *futsch*, and *elav* genes. Additionally, I tried to knock-out the *sens, cut*, and *elav* genes but the penetrance of the resulting phenotypes was low.

CRISPR-mediated knock-ins have two successive steps: introducing a cut to the genome and inserting a transgene through a DSB repair pathway. In two previous studies, knock-ins were performed through either NHEJ or HDR pathways. This indicates that both pathways are active in early *Parhyale* embryos and can be used for knock-ins. However, the efficiency of knock-ins can vary significantly depending on the targeting method, the target locus and the efficiently of the sgRNA; in the published experiments it was ~1% for *Antp* targeted by HDR (Arnaud Martin, personal communication) and ~6% for *Dll* targeted by NHEJ (Kao et al., 2016).

Several different methods have been developed to enhance the frequency of knock-ins, such as inhibiting components of the NHEJ pathway to favor HDR (reviewed in (Liu et al., 2019)) and using single-stranded repair templates (for example (Bai et al., 2020)). Such methods could be beneficial to increase the knock-in rates in *Parhyale*. I would test these approaches using *Antp* or *Dll* targeting as a control, to confirm whether they enhance the knock-in rates also in *Parhyale*.

In all of my knock-in attempts, I confirmed that the targeted loci were cut but the transgenes were not incorporated into the genome. Since the knock-in event is dependent on the efficiency of the genome cutting, a limiting step for my experiments could be the efficiency of DNA cleavage. I assessed the in vivo activity of each sgRNA via the T7 endonuclease assay, which is not quantitative, meaning that I do not have an estimate of the efficiency of each sgRNA. There are several possible explanations why the sgRNAs were not efficient, and based on this I can propose ways to design better sgRNAs.

The *Parhyale* genome has high levels of polymorphism, which is potentially inhibitory for homology-based approaches like CRISPR (Kao et al., 2016). HDR mediated knock-in efficiencies are expected to decrease as the polymorphism in the target loci increases since polymorphisms would decrease the homology between the repair template and the genome, and the *Parhyale* genome is highly polymorphic (Kao et al., 2016). For this reason, designing experiments utilizing the NHEJ repair pathway

could be beneficial for future experiments. To minimize polymorphism in the target loci, I have used the Chicago-F strain (Parchem et al., 2010). This is an inbred strain originating from a single cross. Considering all of the founding individuals are the progeny of a single couple, there could be up to four haplotypes of each gene. For uninjected controls, when the target loci of uninjected embryos are amplified and treated with T7 endonuclease there is no cleavage (as can be seen in figure 4.5, the first two lanes), which shows that there are no short indels in the target loci. Therefore, we need a more sensitive and preferably a quantitative method to assess the polymorphisms in the target loci. In a previous study on *Parhyale*, researchers reported that they assessed the polymorphisms in target loci before designing their sgRNAs but did not reveal how (Clark-Hachtel and Tomoyasu, 2020). One way of achieving this is amplifying the locus of interest by PCR from genomic DNA coming from multiple individuals, and performing next-generation sequencing (as performed in (K. Crawford et al., 2020)). Alternatively, we can take advantage of several deeply sequenced RNAseq datasets generated recently in our lab (Chiara Sinigaglia, unpublished). The reads from these datasets can be used to identify polymorphic regions in the targeted transcripts.

Even though sequence complementarity is necessary for sgRNA functionality, a perfectly matching sgRNA does not necessarily result in an efficient genome cleavage. There are several additional factors, like DNA accessibility, the methylation of the target locus, and the nucleotide composition of the sgRNA (Cui et al., 2018), that influence cleavage efficiency. In the two previous reports of CRISPR-mediated knock-ins in *Parhyale*, the same sgRNAs were initially used to generate null mutants of the target genes and by quantifying the resulting phenotypes, the researchers assessed the efficiencies of the sgRNAs and used only the highly efficient ones for knock-in experiments (Kao et al., 2016; Serano et al., 2016). I tried to follow this approach for *sens*, *cut*, and *elav* genes but most of the animals I injected with sgRNAs to knock-out these genes did not exhibit the expected phenotype. My interpretation is that the sgRNAs tested may not have been efficient and/or that the penetrance of the phenotypes was low. The mutant phenotype screening approach is not applicable to many genes because mutations in many genes do not result in easy-to-score phenotypes, or the penetrance of these mutations is low.

A good method for quantifying the *in vivo* efficiency of the sgRNAs is needed and this could be achieved by next-generation sequencing. A sgRNA can be co-injected with Cas9 protein in *Parhyale* embryos, and the efficiency of genome targeting could be assessed by next-generation sequencing of the target locus. ~100 injected

embryos would give sufficient information about the efficiency of each sgRNA, and 1-2 days of injection would generally yield that number of embryos. This way, the efficiency of several sgRNAs could be assessed within a short time.

When I started to look for marker genes, it was not possible to identify the marker genes of sensilla cells directly, without taking a candidate gene approach. Combined with RNA*ish* on *Parhyale* embryos, I managed to identify several putative marker genes for neurons. However, this approach is limiting, because there are not many putative marker genes identified for other cell types of peripheral sensory organs. The second drawback of this approach is that currently, RNA*ish* is only applicable in embryos before they start to deposit their cuticle, which leaves us ignorant about the expression patterns of the putative marker genes in adult limbs. Currently, we have a single nucleus RNA-seq method established for adult *Parhyale* limbs and very good quality datasets generated in our lab (Alba Almazán, unpublished). These datasets will be valuable for identifying genes expressed in terminally differentiated sensory organ cells of *Parhyale*. Using this approach it will be possible to identify many more marker genes in a more reliable way.

Supplementary Information

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs.
The sgRNAs used in the previous studies are highlighted; the more efficient one to generate knock-outs with green and the other one with red.

CRISPRScan

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
TGACATGTTCGAGGCCACGTTGG	58	GGGTGGCGGCGGAGAGACGGCGG	88	GGGTATGAGGTGTAGTAGGAGG	82	GCCACTACTCACTTCCCCAGGGG	70	TGGCGGGGTACAGTAACATGGGG	78	GCGGCTGCAGCTGGGAGTCTTGG	69	GTTGGTGGTGTTGTTGGTGTTGG	82
GAATCCGTGCACTCAACCTACGG	58	GATCACGGTGGCGACGGCGGCGG	85	AGGGTATGAGGTGTAGTAGGAGG	81	GAGCCCATGGATTGATGTTGCGG	68	GATGACGGACTGGGAGCGTGAGG	73	AGCCGCTGCCGTGGCCGCCACGG	68	TGGAGTCATGTGGTCTTGGTGGG	74
AGTCTGTGATGGACATCCAAGGG	57	GCGGCGGAGAGACGGCGGATAGG	79	GGGCCTATAGTCAGGGTATGAGG	71	TGAGTAGTGGCCGTAATGAGCGG	65	GAATAGGAGGCGAGGTGAGGCGG	71	AGGCTCCATACTGTTGTAGTGGG	59	TGCTGACTGGCCCCATCCAACGG	71
GACGCAGCCATGGCACTAAACGG	55	CGGCATCATGAAGCAGTGGGTGG	77	CGGAGTGTGAGGATGTTGCGGGG	70	GCCCCTGGGGAAGTGAGTAGTGG	63	GTCACAGAGAAGCCGGAGGGCGG	71	AGCTTCAACCGTGGCGGCCACGG	58	GTGGTGTTGTTGGTGTTGGACGG	67
GGTGCCGTAGGTTGAGTGCACGG	51	GGTGGCGGCGGAGAGACGGCGG	77	TGCGGAGTGTGAGGATGTTGCGG	70	TGGTAGGACTCATCGGGCGGAGG	61	GTGACATCAGATGACGGACTGGG	65	GGCAAGGCTGTGCTTCCAGTAGG	54	GGGTTTTACGGAGGCGGTTCAGG	66
GGTTGAGTGCACGGATTCAAAGG	48	GTGGCGCTGTATACGAAGAGCGG	72	TGATGACCGTGCGGAGTGTGAGG	70	TGCATGGTATCTCTCGCATAGGG	61	GGGTACAGTAACATGGGGTACGG	64	GCTCCATACTGTTGTAGTGGGGG	46	GGAGTCATGTGGTCTTGGTGGG	64
GCATCAGCTGGCATCCAACGTGG	47	TGGCGGCATCATGAAGCAGTGGG	72	GAGTACCAAGTTAGTCAAGGCGG	67	GGTAGGACTCATCGGGCGGAGGG	60	TGTGAATAGGAGGCGAGGTGAGG	62	AGCAGCGCCATAATAATCTGCGG	45	AGTGGCGGCAAAGACCAGAATGG	63
GGTCGAGGGATGTCATTCCTTGG	47	GAGACGGCGGATAGGATGAACGG	70	GTTGCGGGGGCCTATAGTCAGGG	65	TGGAGGGCATGTATGAATTTTGG	60	CGAGGTGAGGCGGTGGCTGTTGG	60	TGTGCTTCCAGTAGGTGTATAGG	44	GATAGGGCCCATGCCGGGGCTGG	62
GCCACAGACATGTCGCAATACGG	46	GTGGCGGCATCATGAAGCAGTGG	65	GTTAGTCAAGGCGGTGGCTGCGG	65	GGTAGGACTCATCGGGCGGAGG	53	GGCGGGGTACAGTAACATGGGG	59	GGCTCCATACTGTTGTAGTGGG	42	GACTCAGCGTTGGGCACTACGGG	61
GAGTGTGACTTCGCATCAGCTGG	44	TGGCGACGGCGGCGGAATTGTGG	65	TGAGCGTAATAAACATGGGGGGG	65	GAATTTTGGTAGGACTCATCGGG	50	GCCGCCATGTACGGACAGAACGG	58	GGCCTGCTATGGCCGAGTGCAGG	41	CGGATCCAATTTCTGGTCGAGGG	60
GCCATAGAGAACCAATGGCGTGG	42	AGTGGGTGGCGGCGGAGAGACGG	64	TGGTACTCGTCGGGTGGTTGCGG	62	TGTATTGTAGTAGTCTGCATTGG	50	GGCTTCCCCGCCGCCATGTACGG	57	GGCTCCATACTGTTGTAGTGGGG	38	GTTGTTGTTGGTGGTGTTGTTGG	60
GGCGTGGATTAGCAGCAAATTGG	41	GGCATCATGAAGCAGTGGGTGG	62	TGACTAACTTGGTACTCGTCGGG	60	GGGTCTTACATTATGAACCCGGG	46	AGCGTGAGGGTACCCTGCCAAGG	57	GGAAGCACAGCCTTGCCAAATGG	36	GGGGCTGGATGGAATACATGCGG	59
TGCGAAGTCACACTCACATATGG	39	GCGGATAGGATGAACGGTGGTGG	62	GATCGTAGGGAGGGTAACGAGGG	59	TGATGTTGCGGCATGATACATGG	44	TGATGTCACAGAGAAGCCGGAGG	56	TGACTTTTTAGCTTCAACCGTGG	34	GGGCTGGATGGAATACATGCGG	59
CGGATATTCCAAGATACCCTTGG	38	CGACGGCGGCGGAATTGTGGCGG	59	GGAGTGTGAGGATGTTGCGGGG	59	TGAATTTTGGTAGGACTCATCGG	42	GGTACAGTAACATGGGGTACGG	55	GTCCTGCACTCGGCCATAGCAGG	33	AGCACGGCCTGTCTGCACTCCGG	55
GGCACCGTCTCCACATATACCGG	33	GGCGGCATCATGAAGCAGTGGG	59	TGGTGAGCGTAATAAACATGGGG	58	GGAGGGCATGTATGAATTTTGG	42	GTTACGGGTACTCCTTGGCAGGG	51	GAAGCTAAAAAGTCAGCAGCTGG	32	GCACCTGCTAATGCCATCCACGG	55
TGTCCATCACAGACTGCTCAGGG	33	CGCCACCGTGATCGAGGAGCAGG	51	GGTTGCGGGGGCCTATAGTCAGG	57	CGGCCACTACTCACTTCCCCAGG	42	GGAACTGTGAATAGGAGGCGAGG	50	AGCTGGAGTTCCATTTGGCAAGG	28	GGTTTTACGGAGGCGGTTCAGG	54
GGATTAGCAGCAAATTGGAACGG	30	CGGAATATGCAAAGGCGCCTGGG	50	GTCAGGGTATGAGGTGTAGTAGG	55	GGGCATGTATGAATTTTGGTAGG	34	TGTACGGACAGAACGGTTACGGG	50	GCCACGGCAGCGGCTGCAGCTGG	26	GTAAAACCCGCCCTGGAGAAGGG	52
GCACGGATTCAAAGGATTAGTGG	28	GTCCTCCTGCTCCTCGATCACGG	50	GGTACTCGTCGGGTGGTTGCGG	55	GGTCTTACATTATGAACCCGGG	31	GATGTCACAGAGAAGCCGGAGGG	44	GCAGTCGCCGCAGATTATTATGG	17	GGATCCAATTTCTGGTCGAGGG	52
GATTAGTGGCGTTGCTCTGATGG	26	GTCGCTCGAGGCCGCTGCCCAGG	48	GTGAGCGTAATAAACATGGGGGG	52	GATACCATGCAATATAGTCACGG	31	GTTCTGTCCGTACATGGCGGCGG	43			GGGCCCATGCCGGGGCTGGATGG	52
GGATATTCCAAGATACCCTTGG	22	GGCGACGGCGGCGGAATTGTGG	48	GGCCTATAGTCAGGGTATGAGG	51	GGCCACTACTCACTTCCCCAGG	25	GTAACCGTTCTGTCCGTACATGG	37			GGACTCAGCGTTGGGCACTACGG	50
		CGCCGTCGCCACCGTGATCGAGG	40	GCGGAGTGTGAGGATGTTGCGGG	50	GGCCACTACTCACTTCCCCAGGG	24	GGTTACGGGTACTCCTTGGCAGG	35			GGTGCCAAAGATATGTCCGTTGG	46
		GGAATATGCAAAGGCGCCTGGG	39	GGTGAGCGTAATAAACATGGGGG	47	GGGTCTTACATTATGAACCCGG	23	GTTGTTACCAAAGAACGGCTTGG	34			GGTGACGTACTAGGGTCTGGAGG	46
				GGTGAGCGTAATAAACATGGGG	46	AGGGTCTTACATTATGAACCCGG	20	AGTACCAACAGACTATGGCGGGG	34			GATTTGGAGTCATGTGGTCTTGG	43
				GTACTCGTCGGGTGGTTGCGGGG	41	GGCATGTATGAATTTTGGTAGG	15	GTACTGTTGGAACTGTGAATAGG	32			GGCCCATGCCGGGGCTGGATGG	43
				AGCCACCGCCTTGACTAACTTGG	41			TGTGACATCAGATGACGGACTGG	32			TGCCCAACGCTGAGTCCAAGTGG	41
				GGTACTCGTCGGGTGGTTGCGGG	39			GAACGGTTACGGGTACTCCTTGG	28			GTCATGTGGTCTTGGTGGGTCGG	41
				GGTGTAGTAGGAGGACATGTTGG	36			TGTACCCCGCCATAGTCTGTTGG	25			GCAAATGGATTACGGCGACGAGG	41
				CGATCGTAGGGAGGGTAACGAGG	35			CGTTCTTTGGTAACAACTCTTGG	12			TGCAGACAGGCCGTGCTGACTGG	39
				GACGAGTACCAAGTTAGTCAAGG	29							GCATGTATTCCATCCAGCCCCGG	39

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

CasDesigner

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
AGTCTGTGATGGACATCCAAGGG	81.6	CGACGGCGGCGGAATTGTGGCGG	83.7	CTAACTTGGTACTCGTCGGGTGG	73.1	ACCATGCAATATAGTCACGG	0.7355	TGTACCCCGCCATAGTCTGTTGG	82	GGCCTGCTATGGCCGAGTGCAGG	79.1	GTGGTGTTGTTGGTGTTGGACGG	81.6
TAGTAGTTGCACAATTTTGAAGG	80.5	TGGCGACGGCGGCGGAATTGTGG	80	GGGCCTATAGTCAGGGTATGAGG	70.8	GCCACTACTCACTTCCCCAG	0.734	TGTGACATCAGATGACGGACTGG	81.5	GCTCCATACTGTTGTAGTGGGGG	76.7	TTCGAGCCCTTCTCCAGGGCGGG	81.1
TTGTGCAACTACTAGTAACATGG	78.9	CGGCATCATGAAGCAGTGGGTGG	76.9	TGAGCGTAATAAACATGGGGGGG	69.4	GATACCATGCAATATAGTCA	0.6851	ACAGTACCAACAGACTATGGCGG	77.8	CTGCGCTAACGTCCTGCACTCGG	76.6	TGCAGACAGGCCGTGCTGACTGG	80.6
CAGTCTGTGATGGACATCCAAGG	78.6	GTGGCGGCATCATGAAGCAGTGG	73.7	GTGAGCGTAATAAACATGGGGGG	64.9	TAGGGCATGACCTGCCCCTG	0.6771	GAAGTCCGCGTAACAGCGAAGGG	77.4	TAGGCTCCATACTGTTGTAGTGG	73.5	CCAACGCTGAGTCCAAGTGGAGG	78.6
TGTCCATCACAGACTGCTCAGGG	77.3	TGGCGGCATCATGAAGCAGTGGG	72.9	CGGAGTGTGAGGATGTTGCGGGG	64	TTTTGGTAGGACTCATCGGG	0.6738	CCAACAGTACCAACAGACTATGG	76.4	AATTCGATGTCGGCCTGCTATGG	73.2	CAATTTTCGAGCCCTTCTCCAGG	78.5
CAGCCCTGAGCAGTCTGTGATGG	77	ACGGCGGATAGGATGAACGGTGG	71.9	TACTCGTCGGGTGGTTGCGGGGG	63.3	TGAGTAGTGGCCGTAATGAG	0.6589	TGTGAATAGGAGGCGAGGTGAGG	76	GTCCTGCACTCGGCCATAGCAGG	72.5	GTAAAACCCGCCCTGGAGAAGGG	78.5
CTTGAGTCCGCTTGATACTTTGG	73.4	AGTGGGTGGCGGCGGAGAGACGG	70.9	TTGACTAACTTGGTACTCGTCGG	62.9	TTGTAGTAGTCTGCATTGGA	0.6465	AGAAGTCCGCGTAACAGCGAAGG	75.2	TCTCCCCCACTACAACAGTATGG	71.9	GCACCTGCTAATGCCATCCACGG	78.4
CATTTTACCAAAGTATCAAGCGG	73	GTGGCGCTGTATACGAAGAGCGG	68.9	TAAACATGGGGGGGCATACGAGG	62.8	GGGTCTTACATTATGAACCC	0.6429	CAGTACCAACAGACTATGGCGGG	75	AGGCTCCATACTGTTGTAGTGGG	68.2	TGAAAGGATTTGGAGTCATGTGG	77.8
ACCGTATTGCGACATGTCTGTGG	72.8	ATACGAAGAGCGGTCGCTCGAGG	68	ACACCTCATACCCTGACTATAGG	62.7	GAGCCCATGGATTGATGTTG	0.6353	ACGCGGACTTCTCCGCCCTCCGG	74.9	GGCTCCATACTGTTGTAGTGGGG	67.2	TTTCGAGCCCTTCTCCAGGGCGG	77.4
GCCACAGACATGTCGCAATACGG	72.3	CATCATGAAGCAGTGGGTGGCGG	66.3	GGTTGCGGGGGCCTATAGTCAGG	62.7	ACCACCGTGACTATATTGCA	0.6343	GAATAGGAGGCGAGGTGAGGCGG	74.4	TATGGAGCCTATACACCTACTGG	64.8	CAAAGATATGTCCGTTGGATGGG	76.2
CAATTATCCCTGTCACAAAATGG	72	GCGGCGGAGAGACGGCGGATAGG	65.4	GTACTCGTCGGGTGGTTGCGGGG	62.6	GGGCATGTATGAATTTTGGT	0.6221	AAGAAGATATACAAGACACAGGG	74.2	TGTGCTTCCAGTAGGTGTATAGG	63.9	TGCTGACTGGCCCCATCCAACGG	76.2
ATGGACATCCAAGGGTATCTTGG	70.2	GAGACGGCGGATAGGATGAACGG	65.3	TGACTAACTTGGTACTCGTCGGG	62.6	GGCCACTACTCACTTCCCCA	0.5998	GTGACATCAGATGACGGACTGGG	72.8	GGCAAGGCTGTGCTTCCAGTAGG	55.3	GCTTATCGCTTCCCTCCACTTGG	76.1
ATGTCCATCACAGACTGCTCAGG	69.8	CGGAATATGCAAAGGCGCCTGGG	64.8	AAACATGGGGGGGCATACGAGGG	62.5	GCCCCTGGGGAAGTGAGTAG	0.5838	AGTACCAACAGACTATGGCGGGG	72.4	AGCTGGAGTTCCATTTGGCAAGG	55.3	CAACGCTGAGTCCAAGTGGAGGG	76.1
GATTAGTGGCGTTGCTCTGATGG	67.5	GGGTGGCGGCGGAGAGACGGCGG	64.4	GGTACTCGTCGGGTGGTTGCGGG	62.3	TGCATGGTATCTCTCGCATA	0.5744	CAAGAAGATATACAAGACACAGG	72.4	GGAAGCACAGCCTTGCCAAATGG	54.5	CATCCACGGATAGAAAGTTGTGG	75.9
TTGGTGCCATAGAGAACCAATGG	66.5	CCGGAATATGCAAAGGCGCCTGG	63.8	GTTGCGGGGGCCTATAGTCAGGG	60.7	ATGCCGCAACATCAATCCAT	0.5621	CGTTCTTTGGTAACAACTCTTGG	71.3	TCAGCAGCTGGAGTTCCATTTGG	50.5	TTGGATGGGGCCAGTCAGCACGG	75.3
CCCTCGACCATTTTGTGACAGGG	66.3	GCGGATAGGATGAACGGTGGTGG	62.5	TGGTACTCGTCGGGTGGTTGCGG	60.6	GGTAGGACTCATCGGGCGGA	0.5519	GTCACAGAGAAGCCGGAGGGCGG	68.7	AGCAGCGCCATAATAATCTGCGG	49.9	AATTTTCGAGCCCTTCTCCAGGG	75.2
GGTCGAGGGATGTCATTCCTTGG	65.9	CATGAAGCAGTGGGTGGCGGCGG	60.6	ATTATTCATAAGCCGATCGTAGG	60.1	ATTTCGGGTCAACGTAGCCC	0.5473	TTCTCTGTGACATCAGATGACGG	68.5	GAAGCTAAAAAGTCAGCAGCTGG	48.9	CGTAAAACCCGCCCTGGAGAAGG	75.2
CGGATATTCCAAGATACCCTTGG	65	GTCGCTCGAGGCCGCTGCCCAGG	59.2	AGCCACCGCCTTGACTAACTTGG	59	AGGGTCTTACATTATGAACC	0.5467	GTTCCAACGGACGCAATATTTGG	68.4	GCAGTCGCCGCAGATTATTATGG	45.3	AAAGATATGTCCGTTGGATGGGG	74.9
ACATGTCTGTGGCACCGGACTGG	64.9	ATGCAAAGGCGCCTGGGCAGCGG	59.2	GCGGAGTGTGAGGATGTTGCGGG	58.7	ATTGTAGTAGTCTGCATTGG	0.5396	CAGCTGCAGACTCGAGTATATGG	67.7	AGCCGCTGCCGTGGCCGCCACGG	34.6	ATACCACAACTTTCTATCCGTGG	74
TCCCTCGACCATTTTGTGACAGG	64.9	GTCCTCCTGCTCCTCGATCACGG	58.8	TCATAAGCCGATCGTAGGGAGGG	57.9	CTACTCACTTCCCCAGGGGC	0.5358	GGAACTGTGAATAGGAGGCGAGG	67.6	AGCTTCAACCGTGGCGGCCACGG	32.8	CACAACTTTCTATCCGTGGATGG	73.5
ACGTGTCGATATACCCAGTCCGG	64.5	CGCCGTCGCCACCGTGATCGAGG	58.8	GTTAGTCAAGGCGGTGGCTGCGG	57.7	ATAGGGCATGACCTGCCCCT	0.5344	ATATACAAGACACAGGGTGCCGG	66.9	CTTTTTAGCTTCAACCGTGGCGG	32.5	CCAAAGATATGTCCGTTGGATGG	72.6
CATGTCTGTGGCACCGGACTGGG	64.1	CCAGGCGCCTTTGCATATTCCGG	57.2	CTCGTTACCCTCCCTACGATCGG	57.1	TGTATTGTAGTAGTCTGCAT	0.5318	GATGTCACAGAGAAGCCGGAGGG	66.2	TGACTTTTTAGCTTCAACCGTGG	23.1	GTCATGTGGTCTTGGTGGGTCGG	72.5
GAGTGTGACTTCGCATCAGCTGG	64	CAAAAGGGCGGCTGCCACATCGG	56.5	TTATTCATAAGCCGATCGTAGGG	57.1	TGATGTTGCGGCATGATACA	0.5302	TCTTGCCCTTCGCTGTTACGCGG	66.2			CTCCACTTGGACTCAGCGTTGGG	71.7
TCCCTGTCACAAAATGGTCGAGG	63.5	GCATATTCCGGCCAAAAGGCGG	49.1	GGTGAGCGTAATAAACATGGGGG	56.1	CATGCCGCAACATCAATCCA	0.5293	AGTATATGGTGCGGGGCTTGCGG	65.9			GGGCCCATGCCGGGGCTGGATGG	69.5
ATGGATACATACTTCTGTTCTGG	63.4	CGATGTGGCAGCCGCCCTTTTGG	49.1	TTAGTCAAGGCGGTGGCTGCGGG	56	TTGCATGGTATCTCTCGCAT	0.5249	GCCGAGCTCGCGGCTAAGCTGGG	65.4			TATCCGTGGATGGCATTAGCAGG	69
TGCGAAGTCACACTCACATATGG	63.2	CTTTTGGCCGGAATATGCAAAGG	48.8	TGCGGAGTGTGAGGATGTTGCGG	55.7	AATTTCGGGTCAACGTAGCC	0.5094	AGCCGAGCTCGCGGCTAAGCTGG	65.3			CCTCCACTTGGACTCAGCGTTGG	68.9
CAAGGGTATCTTGGAATATCCGG	62.2	CGCCACCGTGATCGAGGAGCAGG	48.7	TAGTCAAGGCGGTGGCTGCGGGG	55.4	CGGCCACTACTCACTTCCCC	0.4827	TCTGATTCTTATAAGTGTACTGG	64.9			ATGGCATTAGCAGGTGCAAATGG	68.6
CCCTGTCACAAAATGGTCGAGGG	60.9	CTCCTGCTCCTCGATCACGGTGG	48.3	CGATCGTAGGGAGGGTAACGAGG	54.2	TGAATTTTGGTAGGACTCAT	0.4669	CGAGGTGAGGCGGTGGCTGTTGG	64.4			TGCCCAACGCTGAGTCCAAGTGG	68
GCACGGATTCAAAGGATTAGTGG	59.7	TTTGCATATTCCGGCCAAAAGGG	48	GATCGTAGGGAGGGTAACGAGGG	53.5	TGGTAGGACTCATCGGGCGG	0.4529	GCCGCGAGCTCGGCTCGCTCTGG	64.2			GCATGTATTCCATCCAGCCCCGG	67.8

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

CRISPRater

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
TCCCTGTCACAAAATGGTCGAGG	0.83	CGACGGCGGCGGAATTGTGGCGG	0.95	TTAGTCAAGGCGGTGGCTGCGGG	0.82	ATAGGGCATGACCTGCCCCTGGG	0.82	GTTCTGTCCGTACATGGCGGCGG	0.90	AACCGTGGCGGCCACGGCAGCGG	0.85	AAAGTTTACATGAAGCACTGGGG	0.84
ACCGTATTGCGACATGTCTGTGG	0.82	ATACGAAGAGCGGTCGCTCGAGG	0.89	GTACTCGTCGGGTGGTTGCGGGG	0.77	CATAGGGCATGACCTGCCCCTGG	0.79	TGCCGGAGCACAGCTGGCGGTGG	0.88	GCAGTCGCCGCAGATTATTATGG	0.84	TTTGTTGTGGTTGTTGTTGGTGG	0.83
GCACGGATTCAAAGGATTAGTGG	0.80	CTCCTGCTCCTCGATCACGGTGG	0.84	GTTGCGGGGGCCTATAGTCAGGG	0.77	GCCCCTGGGGAAGTGAGTAGTGG	0.78	ACAGTACCAACAGACTATGGCGG	0.87	CCACGGCAGCGGCTGCAGCTGGG	0.83	CTCCAGGGCGGGTTTTACGGAGG	0.82
CATGTCACCGTTTAGTGCCATGG	0.76	TGGCGACGGCGGCGGAATTGTGG	0.81	GATCGTAGGGAGGGTAACGAGGG	0.75	ATTGTAGTAGTCTGCATTGGAGG	0.75	CGAGGTGAGGCGGTGGCTGTTGG	0.84	TGACTTTTTAGCTTCAACCGTGG	0.80	ATTTGTCCAGGATTTTGTTGTGG	0.82
ACTAAACGGTGACATGTTCGAGG	0.75	GATCACGGTGGCGACGGCGGCGG	0.81	TGACTAACTTGGTACTCGTCGGG	0.73	TAGGGCATGACCTGCCCCTGGGG	0.73	TCTTGCCCTTCGCTGTTACGCGG	0.80	GGCCTGCTATGGCCGAGTGCAGG	0.79	TTTCGAGCCCTTCTCCAGGGCGG	0.81
TTGCTGCTAATCCACGCCATTGG	0.74	CATCATGAAGCAGTGGGTGGCGG	0.80	GTTAGTCAAGGCGGTGGCTGCGG	0.73	TGAGTAGTGGCCGTAATGAGCGG	0.73	AGCGTGAGGGTACCCTGCCAAGG	0.79	TCAGTTTGTGAATTCGATGTCGG	0.77	TATCTCTGGGTGACGTACTAGGG	0.81
AGTCTGTGATGGACATCCAAGGG	0.74	ACGGCGGATAGGATGAACGGTGG	0.79	GATGGTGAGCGTAATAAACATGG	0.73	CATCGGGCGGAGGGAATTTCGGG	0.72	TATGGCGGGGTACAGTAACATGG	0.79	GTCCTGCACTCGGCCATAGCAGG	0.76	CCTTTCAAAGTAGACTGCAGTGG	0.79
GGCGTGGATTAGCAGCAAATTGG	0.73	CGCCGTCGCCACCGTGATCGAGG	0.76	CGGAGTGTGAGGATGTTGCGGGG	0.73	TGTATTGTAGTAGTCTGCATTGG	0.72	GTACTGTTGGAACTGTGAATAGG	0.77	CTTTTTAGCTTCAACCGTGGCGG	0.76	GTTGGTGGTGTTGTTGGTGTTGG	0.78
GGTCGAGGGATGTCATTCCTTGG	0.73	AGTGGGTGGCGGCGGAGAGACGG	0.76	GCGGAGTGTGAGGATGTTGCGGG	0.72	TGGTAGGACTCATCGGGCGGAGG	0.72	AGACTCGAGTATATGGTGCGGGG	0.77	AATTCGATGTCGGCCTGCTATGG	0.72	ATTGGATCCGGAGTGCAGACAGG	0.78
CATGTCTGTGGCACCGGACTGGG	0.73	GTCCTCCTGCTCCTCGATCACGG	0.75	ATTATTCATAAGCCGATCGTAGG	0.72	AATTTCGGGTCAACGTAGCCCGG	0.70	GAACGGTTACGGGTACTCCTTGG	0.77	TCAGCAGCTGGAGTTCCATTTGG	0.71	AGTGGCGGCAAAGACCAGAATGG	0.78
TGTCCATCACAGACTGCTCAGGG	0.73	ATGCAAAGGCGCCTGGGCAGCGG	0.75	TTCATAAGCCGATCGTAGGGAGG	0.71	TGGAGGGCATGTATGAATTTTGG	0.70	ACGCGGACTTCTCCGCCCTCCGG	0.76	CCCAGCTGCAGCCGCTGCCGTGG	0.70	CCCGGCATGGGCCCTATCTCTGG	0.76
ACATGTCTGTGGCACCGGACTGG	0.71	GCGGATAGGATGAACGGTGGTGG	0.74	GACGAGTACCAAGTTAGTCAAGG	0.71	GAATTTTGGTAGGACTCATCGGG	0.69	GCCGCGAGCTCGGCTCGCTCTGG	0.75	AGCAGCGCCATAATAATCTGCGG	0.70	CAGGGCGGGTTTTACGGAGGCGG	0.76
CTTGGAATATCCGGTATATGTGG	0.71	CATGAAGCAGTGGGTGGCGGCGG	0.74	ACACCTCATACCCTGACTATAGG	0.71	GGCCACTACTCACTTCCCCAGGG	0.68	GCTGGGCCTCACGCAGACGCAGG	0.75	AGCCGCTGCCGTGGCCGCCACGG	0.69	CTTTGGCACCGTATCCATTCTGG	0.75
CAGCCCTGAGCAGTCTGTGATGG	0.71	CCAGGCGCCTTTGCATATTCCGG	0.71	GGTACTCGTCGGGTGGTTGCGGG	0.71	TGATGTTGCGGCATGATACATGG	0.68	GATGTCACAGAGAAGCCGGAGGG	0.75	CTGCGCTAACGTCCTGCACTCGG	0.69	CCTCGACCAGAAATTGGATCCGG	0.75
ATTGGTTCTCTATGGCACCAAGG	0.71	GCCCGGCAACGTTGCCGATGTGG	0.71	GTGAGCGTAATAAACATGGGGGG	0.70	TTTTGGTAGGACTCATCGGGCGG	0.68	TGTGAATAGGAGGCGAGGTGAGG	0.75	TATGGAGCCTATACACCTACTGG	0.68	CCTCCACTTGGACTCAGCGTTGG	0.74
ACTATCCGCAGACGCAGCCATGG	0.69	CTTTTGGCCGGAATATGCAAAGG	0.71	ATGGTGAGCGTAATAAACATGGG	0.70	TCATCGGGCGGAGGGAATTTCGG	0.67	TCTGTCCGTACATGGCGGCGGGG	0.75	GGAAGCACAGCCTTGCCAAATGG	0.68	CGCCTCCGTAAAACCCGCCCTGG	0.74
TAGTGCCATGGCTGCGTCTGCGG	0.69	GGATGTGGATAGGTTGTGCCCGG	0.70	GGTGAGCGTAATAAACATGGGGG	0.69	GATACCATGCAATATAGTCACGG	0.67	TGGCGGGGTACAGTAACATGGGG	0.75	GCTCCATACTGTTGTAGTGGGGG	0.68	CCACTGCAGTCTACTTTGAAAGG	0.74
GGATTAGCAGCAAATTGGAACGG	0.69	GTGGCGCTGTATACGAAGAGCGG	0.70	AACGAGGGTAACCTTGTTGATGG	0.69	GCCACTACTCACTTCCCCAGGGG	0.66	ACAGGGTGCCGGAGCACAGCTGG	0.74	AGGCTCCATACTGTTGTAGTGGG	0.67	CTTCTCCAGGGCGGGTTTTACGG	0.74
GCATCAGCTGGCATCCAACGTGG	0.68	GGGTGGCGGCGGAGAGACGGCGG	0.69	AAACATGGGGGGGCATACGAGGG	0.69	CATGCCGCAACATCAATCCATGG	0.66	TGATGTCACAGAGAAGCCGGAGG	0.74	GCCACGGCAGCGGCTGCAGCTGG	0.65	GATTTTGTTGTGGTTGTTGTTGG	0.73
CCCTGTCACAAAATGGTCGAGGG	0.68	CGATGTGGCAGCCGCCCTTTTGG	0.67	CTCGTTACCCTCCCTACGATCGG	0.69	ACCATGCAATATAGTCACGGTGG	0.65	AGAAGTCCGCGTAACAGCGAAGG	0.73	TAGGCTCCATACTGTTGTAGTGG	0.65	AATTTTCGAGCCCTTCTCCAGGG	0.73
CGGATATTCCAAGATACCCTTGG	0.68	TGCCACATCGGCAACGTTGCCGG	0.67	TGCGGAGTGTGAGGATGTTGCGG	0.69	CGGCCACTACTCACTTCCCCAGG	0.65	GTCCACCGCCAGCTGTGCTCCGG	0.73	AGCTGGAGTTCCATTTGGCAAGG	0.65	GTGGTGTTGTTGGTGTTGGACGG	0.73
TCCACGCCATTGGTTCTCTATGG	0.67	CTCCTCGATCACGGTGGCGACGG	0.66	CGATCGTAGGGAGGGTAACGAGG	0.67	CTACTCACTTCCCCAGGGGCAGG	0.65	GCGGTGGACGCAGAACTTGCCGG	0.72	GGCTCCATACTGTTGTAGTGGGG	0.64	GGTGACGTACTAGGGTCTGGAGG	0.73
TCCCTCGACCATTTTGTGACAGG	0.67	CTCGATCACGGTGGCGACGGCGG	0.66	TGAGCGTAATAAACATGGGGGGG	0.66	ATTTCGGGTCAACGTAGCCCGGG	0.65	ACCGTTCTGTCCGTACATGGCGG	0.72	GGCAAGGCTGTGCTTCCAGTAGG	0.63	ATTCCATCCAGCCCCGGCATGGG	0.72
CATTTTACCAAAGTATCAAGCGG	0.66	GTGGCAGCCGCCCTTTTGGCCGG	0.66	AGGGTATGAGGTGTAGTAGGAGG	0.66	AGGGTCTTACATTATGAACCCGG	0.64	AGTATATGGTGCGGGGCTTGCGG	0.71	TCTCCCCCACTACAACAGTATGG	0.62	TGTCCCTGTCCCATTTGTCCAGG	0.72
ATATGTGGAGACGGTGCCGTAGG	0.66	GTGGCGGCATCATGAAGCAGTGG	0.64	TAAACATGGGGGGGCATACGAGG	0.66	TTGCATGGTATCTCTCGCATAGG	0.64	TCTTCACCTGCGTCTGCGTGAGG	0.71	TGTGCTTCCAGTAGGTGTATAGG	0.62	GCATGTATTCCATCCAGCCCCGG	0.72
TGACATGTTCGAGGCCACGTTGG	0.66	CACCGTGATCGAGGAGCAGGAGG	0.64	TACTCGTCGGGTGGTTGCGGGGG	0.65	TGCATGGTATCTCTCGCATAGGG	0.63	TGTACCCCGCCATAGTCTGTTGG	0.71	GCGGCTGCAGCTGGGAGTCTTGG	0.61	GGGGCTGGATGGAATACATGCGG	0.72
TTGCGACATGTCTGTGGCACCGG	0.65	TGGCGGCATCATGAAGCAGTGGG	0.63	CTAACTTGGTACTCGTCGGGTGG	0.65	GGTAGGACTCATCGGGCGGAGGG	0.62	TGTACGGACAGAACGGTTACGGG	0.70	AGCTTCAACCGTGGCGGCCACGG	0.60	GATAGGGCCCATGCCGGGGCTGG	0.72
CAGTCTGTGATGGACATCCAAGG	0.64	CAAAAGGGCGGCTGCCACATCGG	0.63	TGGTACTCGTCGGGTGGTTGCGG	0.65	ATGCCGCAACATCAATCCATGGG	0.61	GCCGAGCTCGCGGCTAAGCTGGG	0.70	GAAGCTAAAAAGTCAGCAGCTGG	0.57	CAATTTTCGAGCCCTTCTCCAGG	0.72
ATGTCCATCACAGACTGCTCAGG	0.64	GTCGCTCGAGGCCGCTGCCCAGG	0.62	CAACATCCTCACACTCCGCACGG	0.64	GGGCATGTATGAATTTTGGTAGG	0.61	ACATCACCCAAGCCGTTCTTTGG	0.69			CCAGAGATAGGGCCCATGCCGGG	0.72
TGCGAAGTCACACTCACATATGG	0.64	GCGGCGGAGAGACGGCGGATAGG	0.62	GGTGTAGTAGGAGGACATGTTGG	0.64	TACAATACAACCGCTCATTACGG	0.59	GGGTGCCGGAGCACAGCTGGCGG	0.69			AAAGATATGTCCGTTGGATGGGG	0.71

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

CRISPR MultiTargeter

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
AGTCTGTGATGGACATCCAA	0.76	GCATATTCCGGCCAAAAGGG	0.92	CAACATCCTCACACTCCGCA	0.8	GCCACTACTCACTTCCCCAG	0.78	GATGTCACAGAGAAGCCGGA	0.96	GGCAAGGCTGTGCTTCCAGT	0.63	GCAAATGGATTACGGCGACG	0.76
GCATCAGCTGGCATCCAACG	0.69	GAGACGGCGGATAGGATGAA	0.53	TCATAAGCCGATCGTAGGGA	0.68	TGAGTAGTGGCCGTAATGAG	0.67	TCTTGCCCTTCGCTGTTACG	0.89	GGAAGCACAGCCTTGCCAAA	0.62	GTGGTGTTGTTGGTGTTGGA	0.65
ATATGTGGAGACGGTGCCGT	0.62	GATCACGGTGGCGACGGCGG	0.5	TTATTCATAAGCCGATCGTA	0.57	TTTTGGTAGGACTCATCGGG	0.65	GTTCTGTCCGTACATGGCGG	0.74	CTTTTTAGCTTCAACCGTGG	0.55	ATACCACAACTTTCTATCCG	0.64
CATTTTACCAAAGTATCAAG	0.59	CTTTTGGCCGGAATATGCAA	0.47	GTACTCGTCGGGTGGTTGCG	0.57	ATAGGGCATGACCTGCCCCT	0.61	AAGAAGATATACAAGACACA	0.73	TGACTTTTTAGCTTCAACCG	0.47	GCACCTGCTAATGCCATCCA	0.6
GACGCAGCCATGGCACTAAA	0.57	GTGGCGCTGTATACGAAGAG	0.44	GATCGTAGGGAGGGTAACGA	0.54	GGCCACTACTCACTTCCCCA	0.52	AGTACCAACAGACTATGGCG	0.68	AGCAGCGCCATAATAATCTG	0.47	GTAAAACCCGCCCTGGAGAA	0.6
GAATCCGTGCACTCAACCTA	0.41	CATGAAGCAGTGGGTGGCGG	0.41	GTGAGCGTAATAAACATGGG	0.54	GGGCATGTATGAATTTTGGT	0.5	AGATTTGGTTTCAGAACAAG	0.61	TCAGCAGCTGGAGTTCCATT	0.46	GGTGCCAAAGATATGTCCGT	0.54
GCCATAGAGAACCAATGGCG	0.41	CTTTGCATATTCCGGCCAAA	0.39	CGGAGTGTGAGGATGTTGCG	0.52	GATACCATGCAATATAGTCA	0.49	GCCGCCATGTACGGACAGAA	0.6	GGCTCCATACTGTTGTAGTG	0.46	GGTGACGTACTAGGGTCTGG	0.54
CATGTCACCGTTTAGTGCCA	0.38	CGGAATATGCAAAGGCGCCT	0.31	CTAACTTGGTACTCGTCGGG	0.46	TGATGTTGCGGCATGATACA	0.41	GAAGTCCGCGTAACAGCGAA	0.6	AACCGTGGCGGCCACGGCAG	0.4	CAGAGATAGGGCCCATGCCG	0.53
ATATCCGGTATATGTGGAGA	0.36	CTCCTGCTCCTCGATCACGG	0.3	TGAGCGTAATAAACATGGGG	0.43	ATTTCGGGTCAACGTAGCCC	0.41	GTTACGGGTACTCCTTGGCA	0.57	TCTCCCCCACTACAACAGTA	0.39	GCATGTATTCCATCCAGCCC	0.53
GCACGGATTCAAAGGATTAG	0.36	CGGCATCATGAAGCAGTGGG	0.28	GGTGTAGTAGGAGGACATGT	0.42	GGGTCTTACATTATGAACCC	0.38	GTACTGTTGGAACTGTGAAT	0.56	TATGGAGCCTATACACCTAC	0.3	TTCAAAGTAGACTGCAGTGG	0.5
CCCTGTCACAAAATGGTCGA	0.34	ATACGAAGAGCGGTCGCTCG	0.27	CTCGTTACCCTCCCTACGAT	0.4	GGTAGGACTCATCGGGCGGA	0.35	TCTGTCCGTACATGGCGGCG	0.52	TCAGTTTGTGAATTCGATGT	0.24	GCTTATCGCTTCCCTCCACT	0.48
GCCACAGACATGTCGCAATA	0.31	ACGGCGGATAGGATGAACGG	0.25	GAGTACCAAGTTAGTCAAGG	0.35	GCCCCTGGGGAAGTGAGTAG	0.31	TGGCGGGGTACAGTAACATG	0.5	GCAGTCGCCGCAGATTATTA	0.23	GGGGCTGGATGGAATACATG	0.48
CGGATATTCCAAGATACCCT	0.31	CATCATGAAGCAGTGGGTGG	0.21	GGTGAGCGTAATAAACATGG	0.31	ATGCCGCAACATCAATCCAT	0.28	AGACTCGAGTATATGGTGCG	0.47	GCTCCATACTGTTGTAGTGG	0.23	CAACGCTGAGTCCAAGTGGA	0.47
TGTCCATCACAGACTGCTCA	0.31	CGACGGCGGCGGAATTGTGG	0.2	TATTACGCTCACCATCAACA	0.3	ACCATGCAATATAGTCACGG	0.26	ACAGTACCAACAGACTATGG	0.46	AGCCGCTGCCGTGGCCGCCA	0.22	AATTTTCGAGCCCTTCTCCA	0.43
TCCCTGTCACAAAATGGTCG	0.29	GTGGCGGCATCATGAAGCAG	0.2	TTCATAAGCCGATCGTAGGG	0.3	CATGCCGCAACATCAATCCA	0.25	GAACGGTTACGGGTACTCCT	0.46	CCCAGCTGCAGCCGCTGCCG	0.19	TAGTACGTCACCCAGAGATA	0.41
GGTGCCGTAGGTTGAGTGCA	0.28	GGGTGGCGGCGGAGAGACGG	0.2	TTGACTAACTTGGTACTCGT	0.27	TTGCATGGTATCTCTCGCAT	0.25	GGAACTGTGAATAGGAGGCG	0.45	AGCTGGAGTTCCATTTGGCA	0.19	CTCCAGGGCGGGTTTTACGG	0.39
TTGCTGCTAATCCACGCCAT	0.27	TGGCGGCATCATGAAGCAGT	0.19	TACCAAGTTAGTCAAGGCGG	0.25	GAGCCCATGGATTGATGTTG	0.24	GCAGACTCGAGTATATGGTG	0.41	CCACGGCAGCGGCTGCAGCT	0.16	AGTGGCGGCAAAGACCAGAA	0.39
GGCACCGTCTCCACATATAC	0.26	GGATGTGGATAGGTTGTGCC	0.19	ATGGTGAGCGTAATAAACAT	0.25	TGTATTGTAGTAGTCTGCAT	0.24	ATGACGGACTGGGAGCGTGA	0.41	AGCTTCAACCGTGGCGGCCA	0.15	TCGAAAATTGTCCATCGTCG	0.39
TTGGTGCCATAGAGAACCAA	0.25	TTTGCATATTCCGGCCAAAA	0.16	AAACATGGGGGGGCATACGA	0.24	CTACTCACTTCCCCAGGGGC	0.23	TGATGTCACAGAGAAGCCGG	0.37	TGTGCTTCCAGTAGGTGTAT	0.12	AATCCTGGACAAATGGGACA	0.38
ATTGGTTCTCTATGGCACCA	0.25	CACCGTGATCGAGGAGCAGG	0.14	GTCAGGGTATGAGGTGTAGT	0.24	TGCATGGTATCTCTCGCATA	0.2	GCTGGGCCTCACGCAGACGC	0.3	GGCCTGCTATGGCCGAGTGC	0.1	GGATACCCTCGACCAGAAAT	0.37
CCCTCGACCATTTTGTGACA	0.25	CTCCTCGATCACGGTGGCGA	0.13	ACACCTCATACCCTGACTAT	0.23	ACCACCGTGACTATATTGCA	0.19	ATATACAAGACACAGGGTGC	0.29	CTGCGCTAACGTCCTGCACT	0.09	ATTTGTCCAGGATTTTGTTG	0.34
TGACATGTTCGAGGCCACGT	0.25	GCCACATCGGCAACGTTGCC	0.12	CGATCGTAGGGAGGGTAACG	0.23	TTGTAGTAGTCTGCATTGGA	0.16	GTTGTTACCAAAGAACGGCT	0.28	AATTCGATGTCGGCCTGCTA	0.08	TTTCGAGCCCTTCTCCAGGG	0.33
CAATTATCCCTGTCACAAAA	0.19	ATGCAAAGGCGCCTGGGCAG	0.1	TGGTGAGCGTAATAAACATG	0.23	AGGGTCTTACATTATGAACC	0.14	TGTACCCCGCCATAGTCTGT	0.28	AGGCTCCATACTGTTGTAGT	0.05	TTTTACGGAGGCGGTTCAGG	0.32
GGCGTGGATTAGCAGCAAAT	0.19	CCAGGCGCCTTTGCATATTC	0.09	TGATGACCGTGCGGAGTGTG	0.2	CGGCCACTACTCACTTCCCC	0.13	CAAGAGTTGTTACCAAAGAA	0.27	GAAGCTAAAAAGTCAGCAGC	0.04	CTAGTACGTCACCCAGAGAT	0.32
TGCGAAGTCACACTCACATA	0.19	GTCCTCCTGCTCCTCGATCA	0.08	GGGCCTATAGTCAGGGTATG	0.2	TGAATTTTGGTAGGACTCAT	0.13	GTAACCGTTCTGTCCGTACA	0.26	GTCCTGCACTCGGCCATAGC	0.04	CCTCGACCAGAAATTGGATC	0.32
TCCACGCCATTGGTTCTCTA	0.18	CTCGATCACGGTGGCGACGG	0.08	TGACTAACTTGGTACTCGTC	0.18	TGGTAGGACTCATCGGGCGG	0.11	CCAACAGTACCAACAGACTA	0.25	GCCACGGCAGCGGCTGCAGC	0.02	TGCCCAACGCTGAGTCCAAG	0.32
GGTTGAGTGCACGGATTCAA	0.18	GCGGATAGGATGAACGGTGG	0.08	TGCGGAGTGTGAGGATGTTG	0.14	CATCGGGCGGAGGGAATTTC	0.11	AGCTCAACAAAATGTTCCAA	0.24	TAGGCTCCATACTGTTGTAG	0.01	ATTGGATCCGGAGTGCAGAC	0.29
ACTATCCGCAGACGCAGCCA	0.17	CAAAAGGGCGGCTGCCACAT	0.08	GTTAGTCAAGGCGGTGGCTG	0.13	AATTTCGGGTCAACGTAGCC	0.1	ACATCACCCAAGCCGTTCTT	0.24	GCGGCTGCAGCTGGGAGTCT		TGCTGACTGGCCCCATCCAA	0.27
TTGTGCAACTACTAGTAACA	0.15	GTGGCAGCCGCCCTTTTGGC	0.08	GCGGAGTGTGAGGATGTTGC	0.13	CATAGGGCATGACCTGCCCC	0.09	ACCGTTCTGTCCGTACATGG	0.23			TGGAGTCATGTGGTCTTGGT	0.26
CTTGGAATATCCGGTATATG	0.15	AGTGGGTGGCGGCGGAGAGA	0.07	AACGAGGGTAACCTTGTTGA	0.13	TAGGGCATGACCTGCCCCTG	0.09	GTGACATCAGATGACGGACT	0.22			CCAACGCTGAGTCCAAGTGG	0.26

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

sgRNADesigner

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
GGTGCCGTAGGTTGAGTGCA	0.6395	ACGGCGGATAGGATGAACGG	0.7706	GAGTACCAAGTTAGTCAAGG	0.7716	ACCATGCAATATAGTCACGG	0.7355	AGTACCAACAGACTATGGCG	0.7344	AGCAGCGCCATAATAATCTG	0.6966	AAAGTTTACATGAAGCACTG	0.7788
ACTAAACGGTGACATGTTCG	0.6044	CTCCTGCTCCTCGATCACGG	0.7181	CGGAGTGTGAGGATGTTGCG	0.7468	GCCACTACTCACTTCCCCAG	0.734	AAGAAGATATACAAGACACA	0.7284	TGACTTTTTAGCTTCAACCG	0.6607	GGGGCTGGATGGAATACATG	0.7237
CATTTTACCAAAGTATCAAG	0.6137	TGGCGGCATCATGAAGCAGT	0.6768	GATCGTAGGGAGGGTAACGA	0.7216	GATACCATGCAATATAGTCA	0.6851	AGACTCGAGTATATGGTGCG	0.7068	TCAGTTTGTGAATTCGATGT	0.6084	GCAAATGGATTACGGCGACG	0.7108
TTGCTGCTAATCCACGCCAT	0.5495	GTGGCGCTGTATACGAAGAG	0.6698	GGTGAGCGTAATAAACATGG	0.679	TAGGGCATGACCTGCCCCTG	0.6771	TGGCGGGGTACAGTAACATG	0.7035	GCTCCATACTGTTGTAGTGG	0.6036	AGTGGCGGCAAAGACCAGAA	0.6984
GAATCCGTGCACTCAACCTA	0.5701	GTGGCGGCATCATGAAGCAG	0.6517	GACGAGTACCAAGTTAGTCA	0.6732	TTTTGGTAGGACTCATCGGG	0.6738	GCAGACTCGAGTATATGGTG	0.7025	CTTTTTAGCTTCAACCGTGG	0.6002	AATCCTGGACAAATGGGACA	0.696
ACGTGTCGATATACCCAGTC	0.5415	CGGCATCATGAAGCAGTGGG	0.633	TATTACGCTCACCATCAACA	0.6712	TGAGTAGTGGCCGTAATGAG	0.6589	ACCGTTCTGTCCGTACATGG	0.7015	CTGCGCTAACGTCCTGCACT	0.5795	CCTTTCAAAGTAGACTGCAG	0.6857
TTGGTGCCATAGAGAACCAA	0.6294	GCATATTCCGGCCAAAAGGG	0.6307	GGTGTAGTAGGAGGACATGT	0.6701	TTGTAGTAGTCTGCATTGGA	0.6465	TGATGTCACAGAGAAGCCGG	0.6965	AGCTGGAGTTCCATTTGGCA	0.563	GGTGACGTACTAGGGTCTGG	0.6788
CCCTGTCACAAAATGGTCGA	0.572	CACCGTGATCGAGGAGCAGG	0.6272	CAACATCCTCACACTCCGCA	0.6679	GGGTCTTACATTATGAACCC	0.6429	GATGTCACAGAGAAGCCGGA	0.6947	GGCAAGGCTGTGCTTCCAGT	0.5605	CATCCACGGATAGAAAGTTG	0.673
AGTCTGTGATGGACATCCAA	0.6969	GGGTGGCGGCGGAGAGACGG	0.6172	AAACATGGGGGGGCATACGA	0.6537	GAGCCCATGGATTGATGTTG	0.6353	ACAGTACCAACAGACTATGG	0.6846	AATTCGATGTCGGCCTGCTA	0.5554	CCAACGCTGAGTCCAAGTGG	0.6678
GAGTGTGACTTCGCATCAGC	0.5616	GAGACGGCGGATAGGATGAA	0.613	GTGAGCGTAATAAACATGGG	0.6536	ACCACCGTGACTATATTGCA	0.6343	GGAACTGTGAATAGGAGGCG	0.6839	GGCTCCATACTGTTGTAGTG	0.5512	TCGAAAATTGTCCATCGTCG	0.6513
TGTCCATCACAGACTGCTCA	0.6122	ATACGAAGAGCGGTCGCTCG	0.6106	AGGGTATGAGGTGTAGTAGG	0.6488	GGGCATGTATGAATTTTGGT	0.6221	AGCTCAACAAAATGTTCCAA	0.6756	TATGGAGCCTATACACCTAC	0.5313	ATACCACAACTTTCTATCCG	0.6497
GCACGGATTCAAAGGATTAG	0.52	CTTTTGGCCGGAATATGCAA	0.6059	TGAGCGTAATAAACATGGGG	0.639	GGCCACTACTCACTTCCCCA	0.5998	TCTGTCCGTACATGGCGGCG	0.6717	CCACGGCAGCGGCTGCAGCT	0.515	TTCAAAGTAGACTGCAGTGG	0.6449
TGACATGTTCGAGGCCACGT	0.6468	GCCCGGCAACGTTGCCGATG	0.5997	TAGTCAAGGCGGTGGCTGCG	0.6349	GCCCCTGGGGAAGTGAGTAG	0.5838	ATGACGGACTGGGAGCGTGA	0.6711	TCTCCCCCACTACAACAGTA	0.5032	CAACGCTGAGTCCAAGTGGA	0.6422
CGGATATTCCAAGATACCCT	0.5845	CGACGGCGGCGGAATTGTGG	0.5884	TGGTGAGCGTAATAAACATG	0.6254	TGCATGGTATCTCTCGCATA	0.5744	GCCGCCATGTACGGACAGAA	0.6695	GAAGCTAAAAAGTCAGCAGC	0.5002	TGCCCAACGCTGAGTCCAAG	0.6201
GCCACAGACATGTCGCAATA	0.4748	CTCCTCGATCACGGTGGCGA	0.5766	CTAACTTGGTACTCGTCGGG	0.6244	ATGCCGCAACATCAATCCAT	0.5621	GTTACGGGTACTCCTTGGCA	0.6652	GGCCTGCTATGGCCGAGTGC	0.498	GGTGCCAAAGATATGTCCGT	0.6193
ACCGTATTGCGACATGTCTG	0.6604	CTCGATCACGGTGGCGACGG	0.5559	TGATGACCGTGCGGAGTGTG	0.6243	GGTAGGACTCATCGGGCGGA	0.5519	GTAACCGTTCTGTCCGTACA	0.6551	GTCCTGCACTCGGCCATAGC	0.4917	GCACCTGCTAATGCCATCCA	0.6157
GCATCAGCTGGCATCCAACG	0.6449	GCGGATAGGATGAACGGTGG	0.5508	TCATAAGCCGATCGTAGGGA	0.6216	ATTTCGGGTCAACGTAGCCC	0.5473	TCTTGCCCTTCGCTGTTACG	0.6514	AGGCTCCATACTGTTGTAGT	0.48	CTCCAGGGCGGGTTTTACGG	0.6135
ATTGGTTCTCTATGGCACCA	0.6868	GATCACGGTGGCGACGGCGG	0.5369	TTATTCATAAGCCGATCGTA	0.62	AGGGTCTTACATTATGAACC	0.5467	GGGTGCCGGAGCACAGCTGG	0.6479	TAGGCTCCATACTGTTGTAG	0.4782	GGGTCATAAGCATCTTGTCG	0.6119
GGCACCGTCTCCACATATAC	0.505	CGCCGTCGCCACCGTGATCG	0.5368	CGATCGTAGGGAGGGTAACG	0.6174	ATTGTAGTAGTCTGCATTGG	0.5396	AGATTTGGTTTCAGAACAAG	0.6478	TCAGCAGCTGGAGTTCCATT	0.4779	AGCACGGCCTGTCTGCACTC	0.6092
GCCATAGAGAACCAATGGCG	0.6144	GCGGCGGAGAGACGGCGGAT	0.5234	ATGGTGAGCGTAATAAACAT	0.6122	CTACTCACTTCCCCAGGGGC	0.5358	TGTGAATAGGAGGCGAGGTG	0.6324	GGAAGCACAGCCTTGCCAAA	0.4761	GCATGTATTCCATCCAGCCC	0.6021
ATATCCGGTATATGTGGAGA	0.6437	CATCATGAAGCAGTGGGTGG	0.5214	TTCATAAGCCGATCGTAGGG	0.6034	ATAGGGCATGACCTGCCCCT	0.5344	TGTACCCCGCCATAGTCTGT	0.6265	TGTGCTTCCAGTAGGTGTAT	0.4727	TGAAAGGATTTGGAGTCATG	0.6014
CTTGGAATATCCGGTATATG	0.5536	CGGAATATGCAAAGGCGCCT	0.5201	TACCAAGTTAGTCAAGGCGG	0.5874	TGTATTGTAGTAGTCTGCAT	0.5318	GAACGGTTACGGGTACTCCT	0.6185	AGCTTCAACCGTGGCGGCCA	0.462	CACAACTTTCTATCCGTGGA	0.6005
TTGTGCAACTACTAGTAACA	0.5634	GTCCTCCTGCTCCTCGATCA	0.5193	GTTGCGGGGGCCTATAGTCA	0.5852	TGATGTTGCGGCATGATACA	0.5302	ATATACAAGACACAGGGTGC	0.611	CCCAGCTGCAGCCGCTGCCG	0.4611	TTTTACGGAGGCGGTTCAGG	0.6005
GGTCGAGGGATGTCATTCCT	0.5181	TGGCGACGGCGGCGGAATTG	0.4941	CTCGTTACCCTCCCTACGAT	0.5839	CATGCCGCAACATCAATCCA	0.5293	CTGTTGGAACTGTGAATAGG	0.6055	AACCGTGGCGGCCACGGCAG	0.4299	TTTCGAGCCCTTCTCCAGGG	0.5984
TGCGAAGTCACACTCACATA	0.4676	CATGAAGCAGTGGGTGGCGG	0.4838	GGGCCTATAGTCAGGGTATG	0.5791	TTGCATGGTATCTCTCGCAT	0.5249	GTGACATCAGATGACGGACT	0.603	AGCCGCTGCCGTGGCCGCCA	0.4054	CGGATCCAATTTCTGGTCGA	0.5982
TCCCTCGACCATTTTGTGAC	0.4961	GTGGCAGCCGCCCTTTTGGC	0.4802	ATTATTCATAAGCCGATCGT	0.5682	AATTTCGGGTCAACGTAGCC	0.5094	GAATAGGAGGCGAGGTGAGG	0.6024	GCAGTCGCCGCAGATTATTA	0.3366	TAGTACGTCACCCAGAGATA	0.5919
GACGCAGCCATGGCACTAAA	0.3816	CGCCACCGTGATCGAGGAGC	0.4662	TGCGGAGTGTGAGGATGTTG	0.5607	CGGCCACTACTCACTTCCCC	0.4827	GTCACAGAGAAGCCGGAGGG	0.6008	GCCACGGCAGCGGCTGCAGC	0.1928	AATTTTCGAGCCCTTCTCCA	0.5895
CTTGAGTCCGCTTGATACTT	0.3205	AGTGGGTGGCGGCGGAGAGA	0.4638	TTGACTAACTTGGTACTCGT	0.5383	TGAATTTTGGTAGGACTCAT	0.4669	CAAGAAGATATACAAGACAC	0.5945			ATTTGTCCAGGATTTTGTTG	0.5832
ATATGTGGAGACGGTGCCGT	0.6116	TTTGCATATTCCGGCCAAAA	0.4387	GTACTCGTCGGGTGGTTGCG	0.5308	TGGTAGGACTCATCGGGCGG	0.4529	GTTCTGTCCGTACATGGCGG	0.5889			GTGGTGTTGTTGGTGTTGGA	0.5739

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

sgRNAScorer

AbdA		AbdB	Antp		Dfd	Dll		Scr		Ubx
GCCATAGAGAACCAATGGCGTGG	1.15	-	GTGAGCGTAATAAACATGGGGGG	1.57	-	GTCACAGAGAAGCCGGAGGGCGG	1.72	AACCGTGGCGGCCACGGCAGCGG	1.05	-
GCATCAGCTGGCATCCAACGTGG	1.04		TGAGCGTAATAAACATGGGGGGG	1.50		TGATGTCACAGAGAAGCCGGAGG	1.69	CTTTTTAGCTTCAACCGTGGCGG	0.72	
TCCCTGTCACAAAATGGTCGAGG	0.87		GGTGAGCGTAATAAACATGGGGG	1.30		TGCCGGAGCACAGCTGGCGGTGG	1.58	AGCAGCGCCATAATAATCTGCGG	0.62	
GCACGGATTCAAAGGATTAGTGG	0.74		TTCATAAGCCGATCGTAGGGAGG	1.06		TCTGTCCGTACATGGCGGCGGGG	1.55	TCTCCCCCACTACAACAGTATGG	0.47	
ACTATCCGCAGACGCAGCCATGG	0.72		TAGTCAAGGCGGTGGCTGCGGGG	0.97		GGAACTGTGAATAGGAGGCGAGG	1.37	GAAGCTAAAAAGTCAGCAGCTGG	0.39	
ACTAAACGGTGACATGTTCGAGG	0.70		CTAACTTGGTACTCGTCGGGTGG	0.91		GGGTGCCGGAGCACAGCTGGCGG	1.37	TGACTTTTTAGCTTCAACCGTGG	0.33	
CATTTTACCAAAGTATCAAGCGG	0.35		GTTAGTCAAGGCGGTGGCTGCGG	0.86		AGTACCAACAGACTATGGCGGGG	1.23	GCCACGGCAGCGGCTGCAGCTGG	0.27	
TAGTGCCATGGCTGCGTCTGCGG	0.21		CGGAGTGTGAGGATGTTGCGGGG	0.85		GATGACGGACTGGGAGCGTGAGG	1.11	CCACGGCAGCGGCTGCAGCTGGG	0.23	
ACCGTATTGCGACATGTCTGTGG	0.21		TACCAAGTTAGTCAAGGCGGTGG	0.72		GATGTCACAGAGAAGCCGGAGGG	1.02	GGCTCCATACTGTTGTAGTGGGG	0.23	
ACGTGTCGATATACCCAGTCCGG	0.18		GAGTACCAAGTTAGTCAAGGCGG	0.70		TAGGAGGCGAGGTGAGGCGGTGG	0.98	AGCCGCTGCCGTGGCCGCCACGG	0.11	
ACATGTCTGTGGCACCGGACTGG	0.11		CGATCGTAGGGAGGGTAACGAGG	0.64		GAATAGGAGGCGAGGTGAGGCGG	0.91	GCTCCATACTGTTGTAGTGGGGG	-0.06	
TGACATGTTCGAGGCCACGTTGG	0.09		TAAACATGGGGGGGCATACGAGG	0.61		GTTCTGTCCGTACATGGCGGCGG	0.91	TAGGCTCCATACTGTTGTAGTGG	-0.07	
GCCACAGACATGTCGCAATACGG	-0.03		GTACTCGTCGGGTGGTTGCGGGG	0.41		TGGCGGGGTACAGTAACATGGGG	0.79	GGAAGCACAGCCTTGCCAAATGG	-0.13	
GGATTAGCAGCAAATTGGAACGG	-0.06		CAACATCCTCACACTCCGCACGG	0.41		AGACTCGAGTATATGGTGCGGGG	0.72	GGCCTGCTATGGCCGAGTGCAGG	-0.15	
TTGGTGCCATAGAGAACCAATGG	-0.06		GATCGTAGGGAGGGTAACGAGGG	0.38		GAAGTCCGCGTAACAGCGAAGGG	0.72	AGCTGGAGTTCCATTTGGCAAGG	-0.17	
GAATCCGTGCACTCAACCTACGG	-0.06		TGATGACCGTGCGGAGTGTGAGG	0.36		GCTGGGCCTCACGCAGACGCAGG	0.71	GTCCTGCACTCGGCCATAGCAGG	-0.24	
ATTGGTTCTCTATGGCACCAAGG	-0.07		TGGTGAGCGTAATAAACATGGGG	0.34		ACAGTACCAACAGACTATGGCGG	0.70	AGCTTCAACCGTGGCGGCCACGG	-0.29	
ATATCCGGTATATGTGGAGACGG	-0.11		GGGCCTATAGTCAGGGTATGAGG	0.33		TGTGAATAGGAGGCGAGGTGAGG	0.66	TATGGAGCCTATACACCTACTGG	-0.37	
ATATGTGGAGACGGTGCCGTAGG	-0.12		TACTCGTCGGGTGGTTGCGGGGG	0.23		ACCAGAGCGAGCCGAGCTCGCGG	0.56	AATTCGATGTCGGCCTGCTATGG	-0.42	
CAGTCTGTGATGGACATCCAAGG	-0.14		AAACATGGGGGGGCATACGAGGG	0.17		GTTGTTACCAAAGAACGGCTTGG	0.51	GCAGTCGCCGCAGATTATTATGG	-0.49	
GACGCAGCCATGGCACTAAACGG	-0.14		GGTACTCGTCGGGTGGTTGCGGG	0.15		GCAGACTCGAGTATATGGTGCGG	0.43	GCGGCTGCAGCTGGGAGTCTTGG	-0.54	
TGTCCATCACAGACTGCTCAGGG	-0.14		GATGGTGAGCGTAATAAACATGG	0.11		TCTTGCCCTTCGCTGTTACGCGG	0.42	GGCAAGGCTGTGCTTCCAGTAGG	-0.57	
GGCGTGGATTAGCAGCAAATTGG	-0.15		TATTACGCTCACCATCAACAAGG	0.10		ACCGTTCTGTCCGTACATGGCGG	0.41	TCAGCAGCTGGAGTTCCATTTGG	-0.59	
AGTCTGTGATGGACATCCAAGGG	-0.19		AGGGTATGAGGTGTAGTAGGAGG	0.10		GGGTACAGTAACATGGGGTACGG	0.41	TCAGTTTGTGAATTCGATGTCGG	-0.63	
CAGCCCTGAGCAGTCTGTGATGG	-0.20		TGCGGAGTGTGAGGATGTTGCGG	-0.02		CCAACAGTACCAACAGACTATGG	0.39	CTGCGCTAACGTCCTGCACTCGG	-0.83	
GAGTGTGACTTCGCATCAGCTGG	-0.22		GACGAGTACCAAGTTAGTCAAGG	-0.11		TCTTCACCTGCGTCTGCGTGAGG	0.39	TGTGCTTCCAGTAGGTGTATAGG	-1.10	
GGTGCCGTAGGTTGAGTGCACGG	-0.23		GTTGCGGGGGCCTATAGTCAGGG	-0.16		GCCGCGAGCTCGGCTCGCTCTGG	0.36	AGGCTCCATACTGTTGTAGTGGG	-1.13	
CTTGGAATATCCGGTATATGTGG	-0.26		ATTATTCATAAGCCGATCGTAGG	-0.16		ATGACGGACTGGGAGCGTGAGGG	0.36			
ATGTCCATCACAGACTGCTCAGG	-0.26		GGTGTAGTAGGAGGACATGTTGG	-0.25		AGATTTGGTTTCAGAACAAGCGG	0.30			
CATGTCTGTGGCACCGGACTGGG	-0.28		GCGGAGTGTGAGGATGTTGCGGG	-0.26		TATGGCGGGGTACAGTAACATGG	0.27			

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

SSC

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
CATTTTACCAAAGTATCAAG	0.88	ACGGCGGATAGGATGAACGG	1.27	TTGACTAACTTGGTACTCGT	0.01	TGGTAGGACTCATCGGGCGG	1.12	TGATGTCACAGAGAAGCCGG	1.32	AACCGTGGCGGCCACGGCAG	0.92	CCAACGCTGAGTCCAAGTGG	1.12
GCCATAGAGAACCAATGGCG	0.64	CTTTTGGCCGGAATATGCAA	1.08	TTAGTCAAGGCGGTGGCTGC	0.00	TTTTGGTAGGACTCATCGGG	0.93	TAGGAGGCGAGGTGAGGCGG	1.25	TGACTTTTTAGCTTCAACCG	0.87	CAGAGATAGGGCCCATGCCG	1.08
TTGGTGCCATAGAGAACCAA	0.61	GCATATTCCGGCCAAAAGGG	1.06	AACGAGGGTAACCTTGTTGA	0.00	GCCACTACTCACTTCCCCAG	0.76	TGCCGGAGCACAGCTGGCGG	1.06	AGCAGCGCCATAATAATCTG	0.62	CTCCAGGGCGGGTTTTACGG	1.06
GAATCCGTGCACTCAACCTA	0.56	GTGGCGGCATCATGAAGCAG	1.04	CAACATCCTCACACTCCGCA	-0.06	ATGCCGCAACATCAATCCAT	0.55	AGTACCAACAGACTATGGCG	0.96	CTTTTTAGCTTCAACCGTGG	0.56	CAGGGCGGGTTTTACGGAGG	0.76
GCATCAGCTGGCATCCAACG	0.53	CATCATGAAGCAGTGGGTGG	1.03	GGTTGCGGGGGCCTATAGTC	-0.11	GCCCCTGGGGAAGTGAGTAG	0.52	TGGCGGGGTACAGTAACATG	0.94	CCCAGCTGCAGCCGCTGCCG	0.52	TTCAAAGTAGACTGCAGTGG	0.70
AGTCTGTGATGGACATCCAA	0.49	GGGTGGCGGCGGAGAGACGG	0.99	CTCGTTACCCTCCCTACGAT	-0.12	ACCATGCAATATAGTCACGG	0.50	GTTCTGTCCGTACATGGCGG	0.89	GCTCCATACTGTTGTAGTGG	0.40	GCATGTATTCCATCCAGCCC	0.69
ACTATCCGCAGACGCAGCCA	0.39	ATGCAAAGGCGCCTGGGCAG	0.97	TAAACATGGGGGGGCATACG	-0.15	CATGCCGCAACATCAATCCA	0.47	ACCGTTCTGTCCGTACATGG	0.87	GAAGCTAAAAAGTCAGCAGC	0.30	CCTTTCAAAGTAGACTGCAG	0.66
TCCCTGTCACAAAATGGTCG	0.38	GATCACGGTGGCGACGGCGG	0.93	TTATTCATAAGCCGATCGTA	-0.16	TGAGTAGTGGCCGTAATGAG	0.46	GGGTGCCGGAGCACAGCTGG	0.85	AGCTGGAGTTCCATTTGGCA	0.30	TTTCGAGCCCTTCTCCAGGG	0.53
GAGTGTGACTTCGCATCAGC	0.32	CGACGGCGGCGGAATTGTGG	0.91	TAGTCAAGGCGGTGGCTGCG	-0.21	ATTTCGGGTCAACGTAGCCC	0.44	GCTGGGCCTCACGCAGACGC	0.83	TCAGTTTGTGAATTCGATGT	0.26	AAAGATATGTCCGTTGGATG	0.52
CCCTGTCACAAAATGGTCGA	0.29	CTCGATCACGGTGGCGACGG	0.78	GTACTCGTCGGGTGGTTGCG	-0.24	GGGTCTTACATTATGAACCC	0.38	AGAAGTCCGCGTAACAGCGA	0.78	AGCTTCAACCGTGGCGGCCA	0.19	CCTCCACTTGGACTCAGCGT	0.50
CATGTCACCGTTTAGTGCCA	0.28	CATGAAGCAGTGGGTGGCGG	0.76	TGGTACTCGTCGGGTGGTTG	-0.25	ATAGGGCATGACCTGCCCCT	0.37	ACAGTACCAACAGACTATGG	0.77	AATTCGATGTCGGCCTGCTA	0.17	GTAAAGTTTACATGAAGCAC	0.46
ATATCCGGTATATGTGGAGA	0.27	CACCGTGATCGAGGAGCAGG	0.72	GTCAGGGTATGAGGTGTAGT	-0.29	TTGTAGTAGTCTGCATTGGA	0.33	GTCACAGAGAAGCCGGAGGG	0.72	GGCCTGCTATGGCCGAGTGC	0.10	AATCCTGGACAAATGGGACA	0.41
ATATGTGGAGACGGTGCCGT	0.25	GTGGCGCTGTATACGAAGAG	0.69	GGTACTCGTCGGGTGGTTGC	-0.31	CATAGGGCATGACCTGCCCC	0.32	GAATAGGAGGCGAGGTGAGG	0.69	TAGGCTCCATACTGTTGTAG	0.08	TTTTACGGAGGCGGTTCAGG	0.39
ACTAAACGGTGACATGTTCG	0.16	CTCCTGCTCCTCGATCACGG	0.67	GCGGAGTGTGAGGATGTTGC	-0.33	CTACTCACTTCCCCAGGGGC	0.31	GGAACTGTGAATAGGAGGCG	0.62	GTCCTGCACTCGGCCATAGC	0.06	CGTAAAACCCGCCCTGGAGA	0.37
ATGTCCATCACAGACTGCTC	0.14	AGTGGGTGGCGGCGGAGAGA	0.62	TGATGACCGTGCGGAGTGTG	-0.35	GGCCACTACTCACTTCCCCA	0.24	CTGTTGGAACTGTGAATAGG	0.57	TATGGAGCCTATACACCTAC	0.04	GGTGACGTACTAGGGTCTGG	0.37
CCCTCGACCATTTTGTGACA	0.13	CGGCATCATGAAGCAGTGGG	0.62	TGACTAACTTGGTACTCGTC	-0.54	TAGGGCATGACCTGCCCCTG	0.24	CAAGAGTTGTTACCAAAGAA	0.57	CCACGGCAGCGGCTGCAGCT	0.03	GGGGCTGGATGGAATACATG	0.37
TTGTGCAACTACTAGTAACA	0.13	ATACGAAGAGCGGTCGCTCG	0.56	GGGCCTATAGTCAGGGTATG	-0.66	GAGCCCATGGATTGATGTTG	0.21	CAAGAAGATATACAAGACAC	0.55	AGCCGCTGCCGTGGCCGCCA	0.02	TAGTACGTCACCCAGAGATA	0.35
TGACATGTTCGAGGCCACGT	0.08	CTCCTCGATCACGGTGGCGA	0.55	AGCCACCGCCTTGACTAACT	-0.75	GGTAGGACTCATCGGGCGGA	0.11	GATGTCACAGAGAAGCCGGA	0.55	GGCAAGGCTGTGCTTCCAGT	0.02	GATAGGGCCCATGCCGGGGC	0.35
GGTTGAGTGCACGGATTCAA	0.06	GCGGATAGGATGAACGGTGG	0.53			GATACCATGCAATATAGTCA	0.10	TCTGTCCGTACATGGCGGCG	0.55	GCCACGGCAGCGGCTGCAGC	-0.07	AGTGGCGGCAAAGACCAGAA	0.33
CAGCCCTGAGCAGTCTGTGA	0.05	TGGCGGCATCATGAAGCAGT	0.42			TTGCATGGTATCTCTCGCAT	0.07	ACCAGAGCGAGCCGAGCTCG	0.55	GGAAGCACAGCCTTGCCAAA	-0.10	GGTGCCAAAGATATGTCCGT	0.30
TGCGAAGTCACACTCACATA	0.04	CGGAATATGCAAAGGCGCCT	0.24			TGTATTGTAGTAGTCTGCAT	0.07	TTCTCTGTGACATCAGATGA	0.49	AGGCTCCATACTGTTGTAGT	-0.13	GCAAATGGATTACGGCGACG	0.30
ATTGGTTCTCTATGGCACCA	0.03	CAAAAGGGCGGCTGCCACAT	0.19			AATTTCGGGTCAACGTAGCC	0.05	GCAGACTCGAGTATATGGTG	0.48	CTGCGCTAACGTCCTGCACT	-0.14	CCGGATCCAATTTCTGGTCG	0.29
TGTCCATCACAGACTGCTCA	0.02	GTCGCTCGAGGCCGCTGCCC	0.17			TGCATGGTATCTCTCGCATA	0.04	GTTACGGGTACTCCTTGGCA	0.42	GGCTCCATACTGTTGTAGTG	-0.14	TGCTGACTGGCCCCATCCAA	0.28
GCACGGATTCAAAGGATTAG	-0.02	GCGGCGGAGAGACGGCGGAT	0.16			GGGCATGTATGAATTTTGGT	-0.03	TGTGAATAGGAGGCGAGGTG	0.42	TGTGCTTCCAGTAGGTGTAT	-0.16	GTAAAACCCGCCCTGGAGAA	0.28
ACCGTATTGCGACATGTCTG	-0.05	CGCCACCGTGATCGAGGAGC	0.15			CGGCCACTACTCACTTCCCC	-0.04	GAAGTCCGCGTAACAGCGAA	0.39	TCAGCAGCTGGAGTTCCATT	-0.17	GCACCTGCTAATGCCATCCA	0.26
CAGTCTGTGATGGACATCCA	-0.06	GCCCGGCAACGTTGCCGATG	0.05			AGGGTCTTACATTATGAACC	-0.06	ATGACGGACTGGGAGCGTGA	0.38	GCAGTCGCCGCAGATTATTA	-0.28	ATTCCATCCAGCCCCGGCAT	0.24
CGGATATTCCAAGATACCCT	-0.08	GTGGCAGCCGCCCTTTTGGC	0.00			ATTGTAGTAGTCTGCATTGG	-0.09	AGATTTGGTTTCAGAACAAG	0.38	TCTCCCCCACTACAACAGTA	-0.31	GGGTCATAAGCATCTTGTCG	0.20
TTGCGACATGTCTGTGGCAC	-0.10	GGATGTGGATAGGTTGTGCC	-0.01			TGATGTTGCGGCATGATACA	-0.12	AGACTCGAGTATATGGTGCG	0.37			CGCCTCCGTAAAACCCGCCC	0.20

Table S4.1: Results of various sgRNA prediction tools for *Parhyale* genes previously targeted with CRISPR for knock-outs (continue)

WuCRISPR

AbdA		AbdB		Antp		Dfd		Dll		Scr		Ubx	
gcatcagctggcatccaacg	83	gtggcgctgtatacgaagag	85	gagtaccaagttagtcaagg	86	gccactactcacttccccag	85	gatgtcacagagaagccgga	92	ggaagcacagccttgccaaa	66	agtggcggcaaagaccagaa	80
agtctgtgatggacatccaa	81	acggcggataggatgaacgg	80	gatcgtagggagggtaacga	80	tgagtagtggccgtaatgag	82	agatttggtttcagaacaag	80			tgcccaacgctgagtccaag	76
ttggtgccatagagaaccaa	72	gcatattccggccaaaaggg	78	tattacgctcaccatcaaca	78	accatgcaatatagtcacgg	76	caagagttgttaccaaagaa	80			tttcgagcccttctccaggg	73
caattatccctgtcacaaaa	66	gagacggcggataggatgaa	75	agggtatgaggtgtagtagg	77	catgccgcaacatcaatcca	71	accgttctgtccgtacatgg	80			gcaaatggattacggcgacg	72
attggttctctatggcacca	65	caccgtgatcgaggagcagg	73	caacatcctcacactccgca	76	gataccatgcaatatagtca	68	acagtaccaacagactatgg	78			ggtgacgtactagggtctgg	72
ttgtgcaactactagtaaca	59	cggcatcatgaagcagtggg	71	gatggtgagcgtaataaaca	76	ggtaggactcatcgggcgga	66	agtaccaacagactatggcg	75			cctttcaaagtagactgcag	70
actaaacggtgacatgttcg	58	gtggcggcatcatgaagcag	67	cggagtgtgaggatgttgcg	75	tgatgttgcggcatgataca	64	agactcgagtatatggtgcg	72			ataccacaactttctatccg	68
tgtccatcacagactgctca	57	ctcctgctcctcgatcacgg	66	cgatcgtagggagggtaacg	74	ggccactactcacttccccaa	54	ctgttggaactgtgaatagg	71			ttcaaagtagactgcagtgg	65
gacgcagccatggcactaaa	56	agtgggtggcggcggagaga	61	ctaacttggtactcgtcggg	72	attgtagtagtctgcattgg	58	gtcacagagaagccggaggg	69			ccaacgctgagtccaagtgg	60
gccatagagaaccaatggcg	54	atacgaagagcggtcgctcg	58	ttcataagccgatcgtaggg	71			agctcaacaaaatgttccaa	68			tgctgactggccccatccaa	58
ggtgccgtaggttgagtgca	50	ctttgcatattccggccaaa	57	tagtcaaggcggtggctgcg	70			gccgccatgtacggacagaa	64			gcacctgctaatgccatcca	58
		tttgcatattccggccaaaa	52	tcataagccgatcgtaggga	60			atgacggactgggagcgtga	63			atggcattagcaggtgcaaa	55
		gggtggcggcggagagacgg	50	gacgagtaccaagttagtca	53			tctgtccgtacatggcggcg	63			gtaaaacccgccctggagaa	52
		cgacggcggcggaattgtgg	50					ggaactgtgaataggaggcg	62			gggtcataagcatcttgtcg	50
		catcatgaagcagtgggtgg	50					tgatgtcacagagaagccgg	62				
								gaataggaggcgaggtgagg	62				
								gaagtccgcgtaacagcgaa	58				
								gggtgccggagcacagctgg	54				
								tcttgcccttcgctgttacg	52				

5. Generating and screening CRE-reporters to label the sensory organs of *Parhyale* legs

Introduction

During animal development, multiple cell types arise from a single cell, the zygote. These cells acquire different functions and morphologies by expressing different sets of genes at the right time, in the right amount, and for an exact duration. The specific expression pattern of each gene is determined by regulatory DNA regions called *cis*-regulatory elements (CRE). The major types of functional elements included in CREs are promoters, enhancers, silencers, and insulators. Enhancers modulate gene expression by interacting with promoters, in which transcription is initiated. Since the identification of the first enhancer, a short sequence from the genome of SV40 virus, more than 40 years ago (Banerji et al., 1981), a lot of tools and techniques have been generated to identify and to validate the enhancers in different organisms (Levine, 2010). The identification of CREs draws heavy attention not only because it provides insights on how genes are regulated, but also because CREs are used to express marker genes (CRE reporters) or other transgenes, which are valuable tools to visualize and manipulate cells and tissues.

Identification of enhancers in the genome

Enhancers are typically 200bp to 1kb in length and they are generally located in intra- and inter-genic non-coding regions (Levine, 2010). There are different approaches developed to identify CREs, including enhancers, using features of these elements that are related to their function. One approach uses chromatin immunoprecipitation (ChIP) – either followed by DNA microarray hybridisation (ChIP-on-chip) or by deep sequencing (ChIP-seq) – to identify enhancers based on molecular markers, such as histone modifications or transcription factors, that are enriched in those regions ((Sandmann et al., 2007; Visel et al., 2009), figure 5.1b). A second approach is based on the fact that active enhancers have a relatively open chromatin structure, which makes the enhancer DNA accessible to transcription factors. This means that assessing DNA accessibility may predict the location enhancers (reviewed in (Clapier and Cairns, 2009)). In recent years, several methods have been developed to isolate the accessible regions of chromatin, to sequence them, and to map them on the genome. The most common techniques are DNAse-seq (G. E. Crawford et al., 2006), FAIRE-seq (Giresi et al., 2007), MNase-seq (Schones et al., 2008), and ATACseq (Buenrostro et al., 2013), (figure 5.1a).

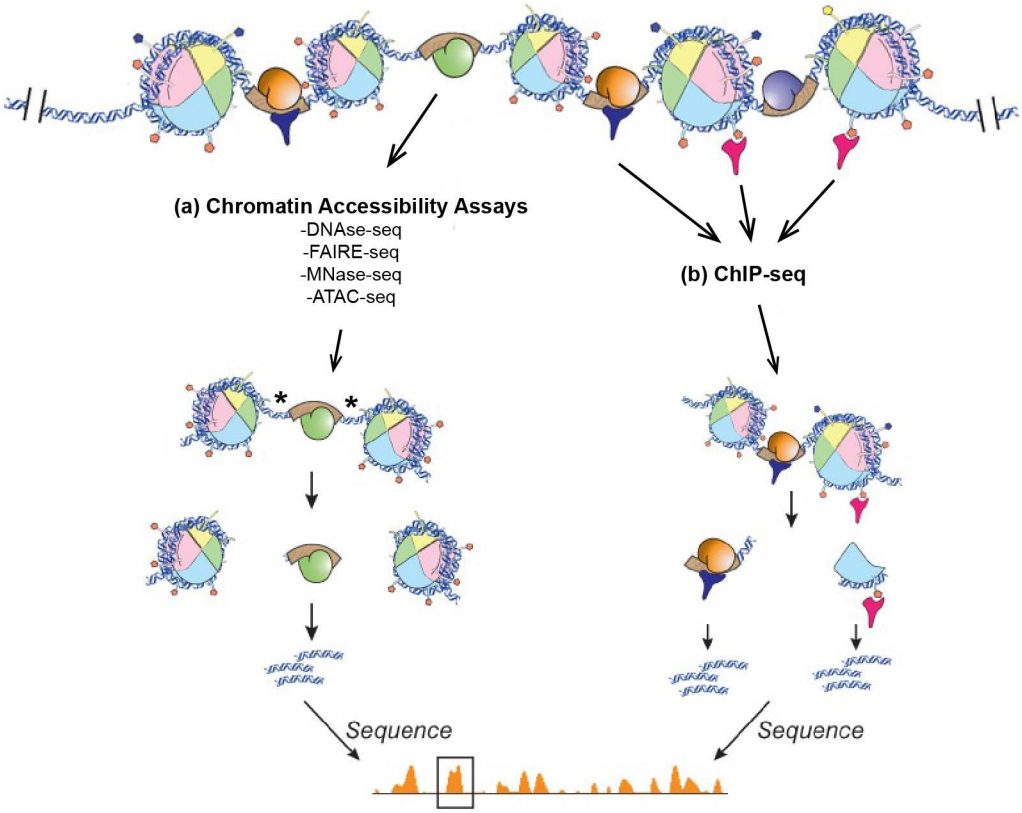


Figure 5.1: Illustration of experimental approaches for CRE discovery.
(a) CREs can be predicted through assays for accessible chromatin. **(b)** An additional method used for CRE discovery is chromatin immunoprecipitation (ChIP)-seq directed against specific histone modifications (pink) or transcription factors (brown). For both chromatin accessibility and ChIP-seq assays, predicted CRE regions identified by next-generation sequencing (adapted from (Suryamohan and Halfon, 2014)).

A third approach is to map the physical interactions of distant genomic regions using 3C methods, which can provide information on distant enhancer-promoter interactions (reviewed in (Denker and de Laat, 2016)).

Assay for transposase-accessible chromatin using sequencing (ATACseq)

The regulation of chromatin accessibility is a significant part of epigenetic regulation (Klemm et al., 2019). ATACseq is an innovative technique utilizing a hyperactive transposase of the Tn5 transposon to assess chromatin accessibility throughout the genome. In this method, the samples of tissues or cells are lysed, and the isolated and permeabilized nuclei are treated with Tn5 transposase which is pre-loaded with tagged DNA fragments. The transposase inserts the pre-loaded DNA fragments into chromatin regions that are accessible to the transposase (figure 5.2a); this process is called ‘tagmentation’ as the chromatin is tagged and fragmented simultaneously. Next, the tagmented DNA is purified and amplified with barcoded primers for next-generation sequencing (figure 5.2b). The sequencing reads are mapped back to the genome and

the abundance of reads corresponding to each genomic region is assessed (figure 5.2c, (Buenrostro et al., 2013; S et al., 2014)).

ATACseq has several advantages over the other techniques that measure chromatin accessibility. First, the samples can be prepared within a day; its simplified and rapid methodology leads to highly reproducible results. Second, the amount of starting material required is at least 1000-fold lower compared with other methods, which makes it the method of choice when sample collection is challenging. And finally, ATACseq is compatible with paired-end sequencing technology, in which the libraries are sequenced from both ends, allowing more accurate mapping to repetitive sequences in the genome (reviewed in (Sun et al., 2019)).

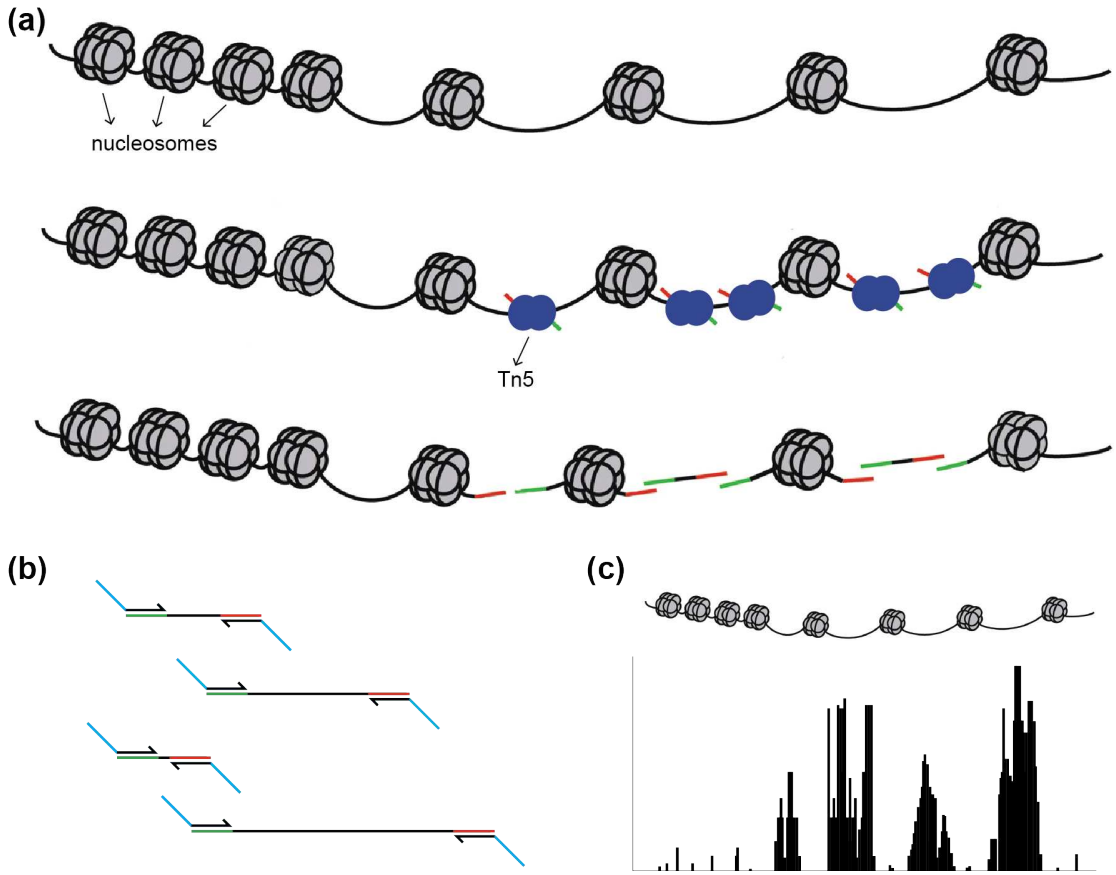


Figure 5.2: Illustration of the ATACseq method to identify accessible chromatin.
(a) Chromatin accessibility depends on the density of nucleosomes. The hyperactive transposase Tn5 (dark blue) tags the accessible chromatin by inserting pre-loaded tagged DNA fragments. The frequency of tagging depends on the accessibility of the chromatin. **(b)** The fragmented chromatin is then purified, amplified with barcoded primers, and sequenced. **(c)** The reads are mapped back to the genome and the accessibility of each locus is assessed by quantifying the relative abundance of reads across the genome (adapted from (Sun et al., 2019)).

Methods for the functional characterization of enhancers

The methods outlined above provide information that can be used to define putative CREs. However, these candidates need to be functionally validated. For instance, although accessible chromatin is likely to harbor *cis*-regulatory elements, not all accessible chromatin corresponds to CREs. Transcription factor binding sites identified with ChIP-seq have sometimes been shown to be inactive when tested for enhancer activities *in vivo* (Kvon et al., 2012). And chromatin accessibility based approaches do not provide any information on the functional properties of the identified CREs. Likewise, inconsistencies between fluorescent *in situ* hybridization experiments and 3C based methods indicates that these methods may give false-positive results (Williamson et al., 2014).

The development of transposable element based transgenesis in *Drosophila* (GM and AC, 1982) was the start of a new era to study enhancers. The pioneering study of the *even-skipped* locus of *Drosophila* demonstrated that gene regulation involves multiple separable enhancers, each carrying distinct activator and repressor complexes (Doyle et al., 1989). Since this pioneering work, transgenesis techniques have been developed in several animal species and transgenesis with CRE reporters is still used widely to study the function of enhancers *in vivo* (reviewed in (Kvon, 2015)).

The most common transgenic strategies to capture enhancer activities *in vivo* are enhancer trap screens and the integration of enhancer reporter constructs into the genome, either in a random location or in a specific site. Either of these approaches requires a reporter construct, consisting of a minimal promoter adjacent to a reporter gene, and a DNA vector. In enhancer trapping, this construct is inserted into the genome in a random manner via transposable element mediated transgenesis. The reporter expression is driven by CREs present at the insertion locus (figure 5.3a), so each insertion site has the potential of 'trapping' the activity of a different enhancer. In the second approach, a putative enhancer sequence is inserted in front of the core promoter-reporter cassette, to make an enhancer-reporter. This enhancer reporter can be integrated into the genome randomly, in which case it may also interact with enhancers around the insertion site (figure 5.3b, reviewed in (Arnone and Davidson, 1997)). In this approach, screening multiple insertion sites can help to distinguish the activity of the transgene enhancer from that of endogenous enhancers around the site of insertion. Alternatively, the enhancer-reporter can be inserted into a specific position in the genome with the help of site-specific integration methods, such as the φC31

integrase system (Thorpe and Smith, 1998). By pre-screening the gene insertion sites, endogenous enhancer activity can be avoided (figure 5.3c).

Generating enhancer reporters is widely-used for identifying enhancers and characterizing their activity. Apart from that, these reporters open a window to observe and manipulate the cells they are active in. For example, they can be used to express fluorophores to label those cells, or different effector proteins to manipulate them.

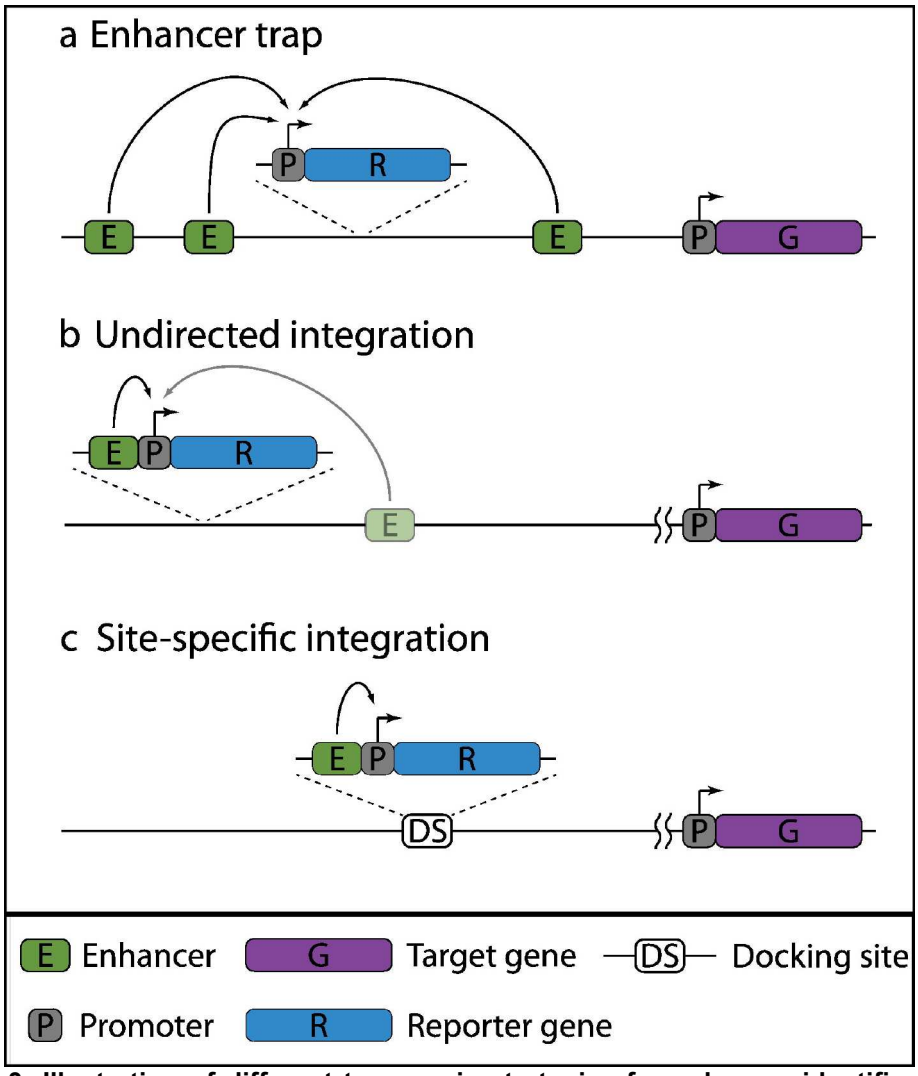


Figure 5.3: Illustration of different transgenic strategies for enhancer identification and characterization.
(a) *Enhancer-traps* use a transposon-based reporter construct containing a minimal promoter followed by a reporter gene (blue box) integrated into the genome. The reporter gene transcription is activated by enhancers (green boxes) surrounding the integration site. The resulting reporter expression pattern often recapitulates the expression pattern of a target gene (the gene which is normally regulated by these enhancers; purple). **(b)** *Undirected integration* of enhancer–reporter DNA into the genome results in the reporter transcription pattern that typically corresponds to the activity of the cloned enhancer, and may also be trapping the activity of endogenous enhancers around the insertion site. The chromatin environment or other enhancers surrounding the insertion site (faded green boxes) can affect reporter activity resulting in additional ectopic domains of expression. **(c)** *Site-specific integration* of the enhancer–reporter DNA into a "docking site" (DS) located in the genome (adapted from (Kvon, 2015)).

Identifying CREs in *Parhyale hawaiensis*

There are six functional CRE-reporters in *Parhyale*: two made with CREs from Opsin genes which are active in photoreceptors (Ramos et al., 2019), two with CREs from hsp70 family genes (one heat-inducible and one active in muscles; (Pavlopoulos and Averof, 2005)), and two synthetic CREs, made with multimerized Pax6 and Pax6+Sox binding sites (named 3xP3 and DC5, respectively), which are active in the nervous system (Konstantinides and Averof, 2014; Pavlopoulos and Averof, 2005). All of these constructs were integrated randomly in the genome of *Parhyale* using *Minos* transposable element mediated transgenesis, which is well established in *Parhyale* (Pavlopoulos and Averof, 2005).

In this study, I aim to generate cell-type-specific reporter lines to label the cells of peripheral sensory organs on *Parhyale* legs, via CRE-reporter cloning. To achieve this, I first identified marker genes that are specifically expressed in these cells and cloned the upstream sequences of the genomic loci of these genes. Then I generated reporter constructs in which these sequences were placed upstream of fluorescent marker genes, within transposon vectors, microinjected these constructs in embryos and screened these for reporter expression. Simultaneously, I performed ATACseq to identify putative CREs surrounding the promoter of these genes. The ATACseq results were only available towards the end of the project, so were not available while designing the reporter constructs. Here I present the reporter constructs together with the ATACseq results.

Results and discussion

ATACseq on *Parhyale* embryos and limbs

I prepared three ATACseq libraries: from a whole stage S24 embryo (sample 1), from all the legs of 4 stage S24 embryos (sample 2), and from a single T4 limb of an adult male (sample 3). Stage 24 embryos were chosen to avoid problems caused by cuticle deposition (see below), which occurs in later embryonic stages. To prepare the libraries, I dissected and prepared each tissue sample by lysing, pelleting, resuspending, and eventually tagmenting with Tn5 transposase (for the detailed protocol, see appendix A – Materials and Methods). Then, I prepared the libraries by purifying the DNA and amplifying it by PCR using barcoded primers. Inspecting the size distribution of these libraries with TapeStation displayed the expected laddering pattern of sub-, mono-, and di-nucleosomal fragments of ~200, ~350, and ~550 bp length, respectively, for the embryonic samples (Halstead et al., 2020). However, the adult limb library was highly enriched with sub-nucleosomal fragments, which can be an indication of an improper tagmentation process (figure 5.4). The three libraries were sequenced, yielding ~26M, ~19M, and ~34M single-end reads per library. 70-90% of the reads from the two embryonic libraries were successfully mapped to genome assembly phaw5.0, and the resulting datasets had a high proportion of reads falling in identified peaks (FRiP score 0.3-0.4) (see figure 5.4). In contrast, the reads from sample 3 had very low mapping performance, as well as, a low number of peaks, and a low FRiP score (figure 5.4). These results show that the two embryonic ATACseq datasets are of high quality, but the adult limb dataset is of poor quality. I also confirmed the quality of the datasets by visual inspection of the peak distribution in selected loci, using a genome browser (see below), e.g. confirming the presence peaks around transcription start sites (TSS).

Considering that the same protocol was used to prepare all three ATACseq libraries, the problem with the adult limb dataset is most likely stemming from problems in tissue preparation. During the preparation of adult arthropod tissues, removal of the cuticle is crucial because it dramatically decreases the quality of the results (Mathilde Paris, personal communication). This is the most likely reason why the quality of the third library was compromised.

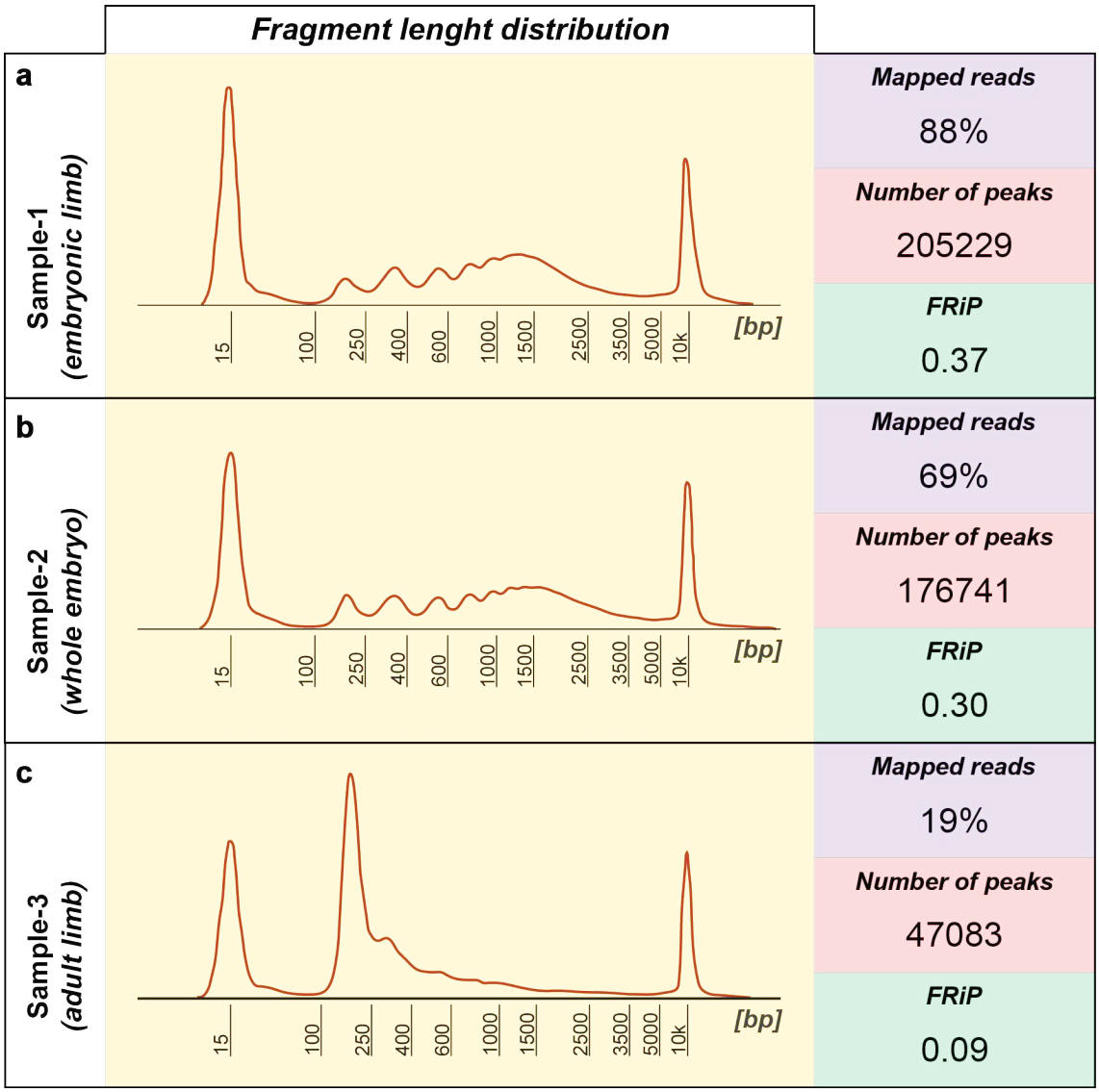


Figure 5.4: Summary of the ATACseq libraries.
Panels illustrate the cumulative distribution of fragment lengths assessed with Tapestation (yellow panels), percentage of the reads successfully mapped to the genome (purple panels), total number of peaks (red panels), and the FRiP scores (green panels) of **(a)** embryonic limb, **(b)** whole embryo, and **(c)** adult limb ATACseq datasets.

Transgenesis in *Parhyale* is a well-established technique and an efficient way to test the *cis*-regulatory element (CRE) activity (Pavlopoulos and Averof, 2005). The established method requires the insertion of the transgene of interest in a *Minos* transposable element vector, which carries a transgenesis marker. The established transgenesis markers, *3xP3-DsRed* and *3xP3-EGFP*, are expressed in a small group of neurons which is visible in late embryos as well as adults. However, the expression is sometimes weak and not easy to score in transgenic animals. This becomes limiting, especially when large numbers of embryos need to be screened. For this reason, I prepared a new *Minos* vector that carries *PhOpsin1-eGFP* as a marker. The *Minos(PhOpsin1-eGFP)* vector drives high levels of eGFP expression in photoreceptors (Ramos et al., 2019), which enables rapid and more accurate

screening. In order to test the activity of putative CREs, I cloned gene fragments of several marker genes, containing the 5' UTR, promoter and upstream sequences, in front of the eGFP coding sequence, and transferred this reporter cassette into the *Minos(PhOpsin1-eGFP)* vector (figure 5.5). These CRE reporter constructs were co-injected with the Minos transposase mRNA, and the injected mosaic embryos and hatchlings were screened for expression of the marker and the CRE reporter transgene. For each tested construct, I have screened a minimum of 50 embryos expressing the transgenesis marker.

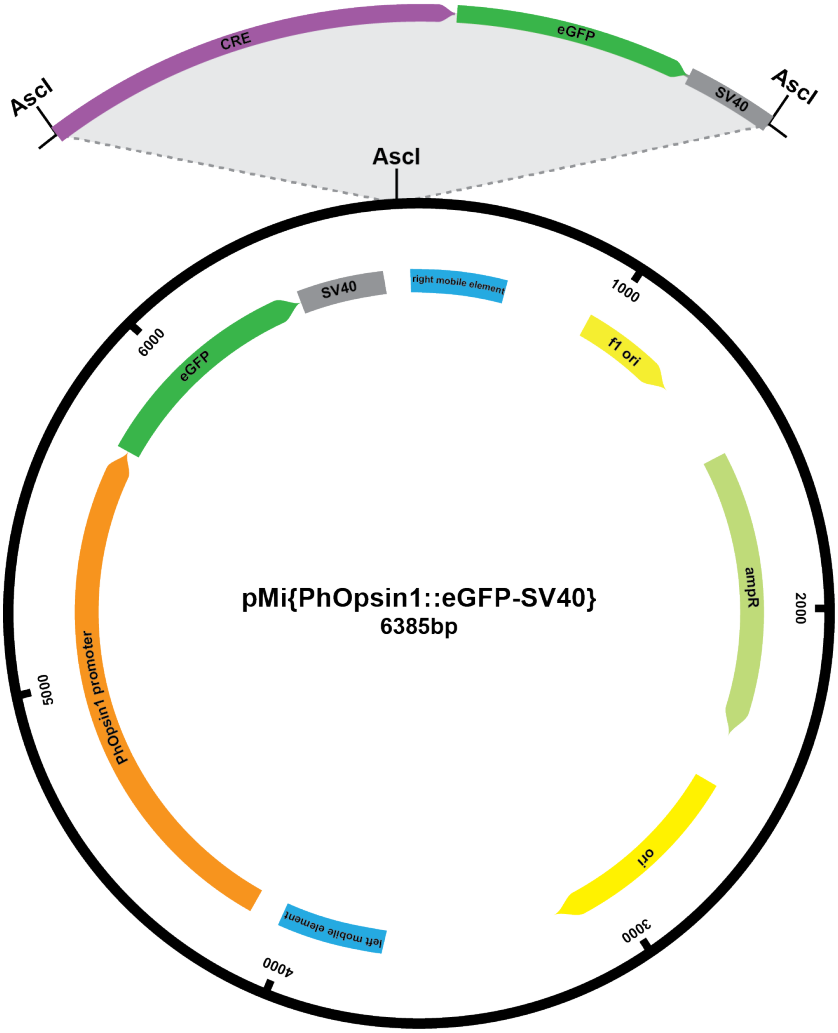


Figure 5.5: Illustration of *Minos* CRE reporter construct.
The *pMinos(phOpsin1-eGFP)* plasmid contains the PhOpsin1::eGFP construct as a transgenesis marker. The tested CRE is cloned upstream of the eGFP coding sequence. The tested cassette is cloned into the vector using AscI restriction sites. Once the *Minos* plasmid is co-injected into the embryos with *Minos* transposase mRNA, the sequences flanked by the mobile element arms (blue bars) are integrated into the genome. SV40 represents the SV40 early polyadenylation sequence.

Selecting putative sensory organ markers and testing putative CRE reporters

I took a candidate gene approach to choose marker genes, based on genes known to be expressed in the sensillum of *Drosophila*. I identified the *Parhyale* orthologs,

excluding genes with very large introns in their 5'UTR for two reasons. First, these introns are likely to contain CREs (Cenik et al., 2010), but are too large to clone in our transgenesis vectors. Second, since there is no functionally characterized minimal core promoter in *Parhyale*, I aimed to include the endogenous core promoter of each in the cloned CRE-reporters by including sequences upstream of the TSS. Transcriptome assemblies do not reveal the precise location of the TSS, I therefore always cloned the entire region upstream of the start codon, which ensures cloning the core promoter in genes that have small 5'UTRs without introns.

I had already identified several putative marker genes, as potential targets for CRISPR-mediated knock-ins (see chapter 4), however, most of these genes were not suitable candidates for promoter-reporters either because their sequence was incomplete in our *Parhyale* genome assembly or because there is a big intron in their 5'UTR. For this reason, I decided to expand my list of potential candidates (Table 5.1).

Table 5.1: The list of *Parhyale* transcripts targeted in this study.

gene	*source*	*transcript ID*	
brp	r5.1 (Kao, *unpublished*)	phaw_50.283862bG132	
syt4	Hunt+Rosata (Hunt *et al.* 2019)	PH.k21.comp3783_seq2	
VAChT	r5.1 (Kao, *unpublished*)	phaw_50.283866G200	
futsch	r3.1 (Kao *et al.* 2016)	multiple transcripts	incomplete, no 5' sequence of the mRNA
elav	r5.1 (Kao, *unpublished*)	phaw_50.283869cG696	big intron in the 5'UTR
cpo	r3.1 (Kao *et al.* 2016)	c100023_g1_i1	incomplete, no 5' sequence of the genomic locus
SNARE-1	r3.1 (Kao *et al.* 2016)	c20695_g1_i1	incomplete, no 5' sequence of the mRNA
SNARE-2	r3.1 (Kao *et al.* 2016)	c7854_g1_i1	incomplete, no 5' sequence of the mRNA
syt1-a	r3.1 (Kao *et al.* 2016)	c50879_g1_i1	big intron in the 5'UTR
syt1-b	r3.1 (Kao *et al.* 2016)	c225981_g1_i1	incomplete genome, no 5' of the gene
repo	r3.1 (Kao *et al.* 2016)	c37156_g1_i2	~15kb intron in the 5'UTR
gcm	r5.1 (Kao, *unpublished*)	phaw_50.283869bG256	
gs-1	r5.1 (Kao, *unpublished*)	phaw_50.283869dG38	~10kb intron in the 5'UTR
gs2	r5.1 (Kao, *unpublished*)	phaw_50.283869dG36	~9kb intron in the 5'UTR
ash1	r5.1 (Kao, *unpublished*)	phaw_50.283875aG454	
ash2	r5.1 (Kao, *unpublished*)	phaw_50.283875aG452	
ash3	r5.1 (Kao, *unpublished*)	phaw_50.283875aG460	
sens	r3.1 (Kao *et al.* 2016)	c230020_g1_i1	
pros	r3.1 (Kao *et al.* 2016)	c54410_g1_i1	
cut	r5.1 (Kao *et al.* 2016)	phaw_50.283865aG58	

Initially, I focused on neural markers and identified ten additional genes of interest. Seven of them were not suitable for CRE cloning, for the reasons mentioned above (Table 5.1), which left me with three candidates: *brp*, *chAT*, and *syt4*. I previously identified the *Parhyale brp* orthologue and revealed its expression pattern in embryos (see chapter 4). In order to prepare the CRE-reporter, I cloned two fragments, 1.5 kb and 3.8 kb upstream of the start codon, including the 133 bp 5'UTR, promoter region and upstream sequences (figure 5.6). Transgenic animals carrying either of

these constructs showed no eGFP expression besides the expression of the transgenesis marker.

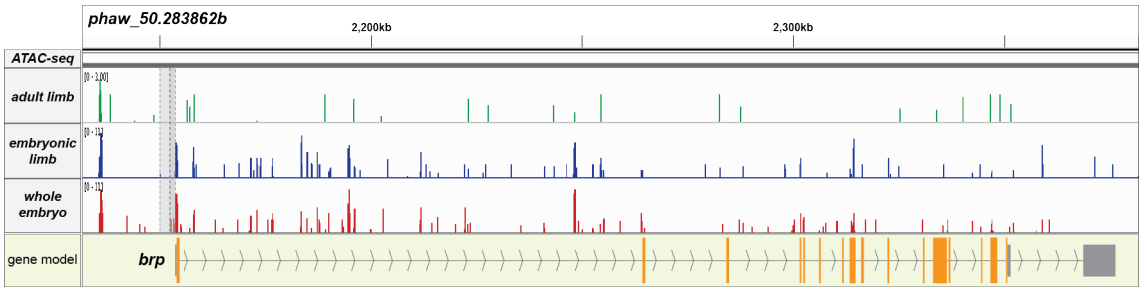


Figure 5.6: The overview of the *brp* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing putative CREs in the *brp* genomic locus. The large peaks ~20 kb upstream of brp locus mark the TSS of another gene expressed in the opposite direction. The gene model of *brp*, with introns (arrowheads), exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

My second candidate was the synaptotagmin family member *syt4*. Synaptotagmins are membrane trafficking genes with neuron restricted and ubiquitously expressed members (Südhof, 2002). *Drosophila* Syt4 is specifically expressed in neurons and localizes to the presynaptic axon terminals (Littleton et al., 1999). I identified the *Parhyale syt4* ortholog via reciprocal BLAST and tried to amplify the genomic fragments upstream of the start codon. I managed to clone a ~2 kb fragment containing the 139 bp long 5'UTR, and ~1,8 kb upstream of the putative TSS (figure 5.7). This fragment was not sufficient to drive neuronal eGFP expression in screened transgenic animals. There were no strong ATACseq peaks on the *syt4* genomic locus, which suggests that these constructs may lack the promoter of *syt4*.

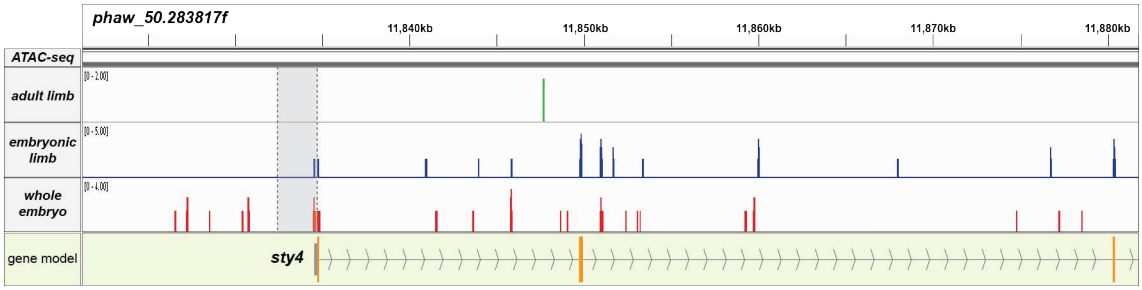


Figure 5.7: Overview of the *syt4* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *syt4* genomic locus. The gene model of *syt4*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

Choline O-acetyltransfarase (chAT) and *Vesicular acetylcholine transporter (VAChT)* are putative marker genes for cholinergic neurons. These genes are linked and have a particular organization that is widely conserved in the animal kingdom: *VAChT* is nested in the first intron of *chAT*. The two genes are alternatively spliced

from the same pre-mRNA and mature mRNAs share the first exon (Mathews et al., 2015). ~7 kb fragments upstream of the *chAT* start codon were enough to drive reporter expression both in drosophila (Salvaterra and Kitamoto, 2001) and in the mouse (Naciff et al., 1999). Since the crustacean sensory neurons are cholinergic (Barker et al., 1972), *chAT* and *VAChT* are promising candidates for marking those neurons.

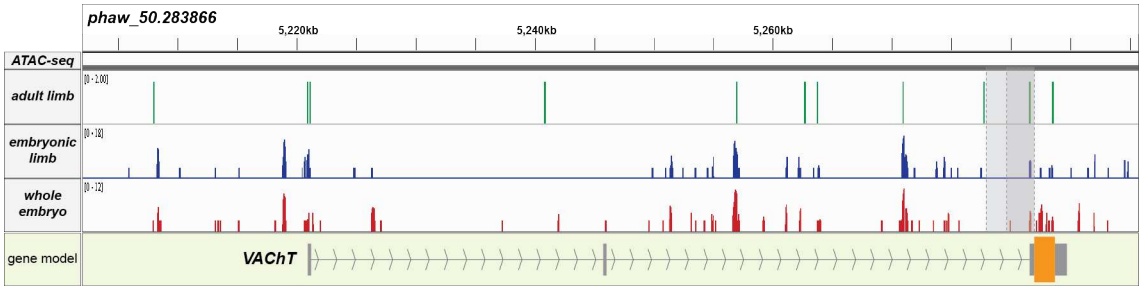


Figure 5.8: Overview of the *VAChT* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *VAChT* genomic locus. The gene model of *VAChT*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

I failed to find a *chAT* ortholog in the *Parhyale* genome and transcriptomes, but I found a *VAChT* ortholog. Since these two genes are expected to share their first exon, I cloned 2.3 kb and 4 kb fragments upstream of the start codon of *VAChT*. Neither of these constructs drives reporter expression in transgenic animals and I could not manage to clone a larger fragment. Subsequent analyses with the updated genome and transcriptome datasets revealed that I did not have the full picture of the *VAChT* genomic locus and that my fragments did not include the TSS of *VAChT*. There are two large introns in the 5'UTR of *VAChT* gene, which I was not aware of, and I had actually cloned a part of the second intron and the beginning of the third exon. The actual upstream of the *VAChT* gene is marked by several ATACseq peaks, which may be bearing the regulatory regions of this gene (figure 5.8).

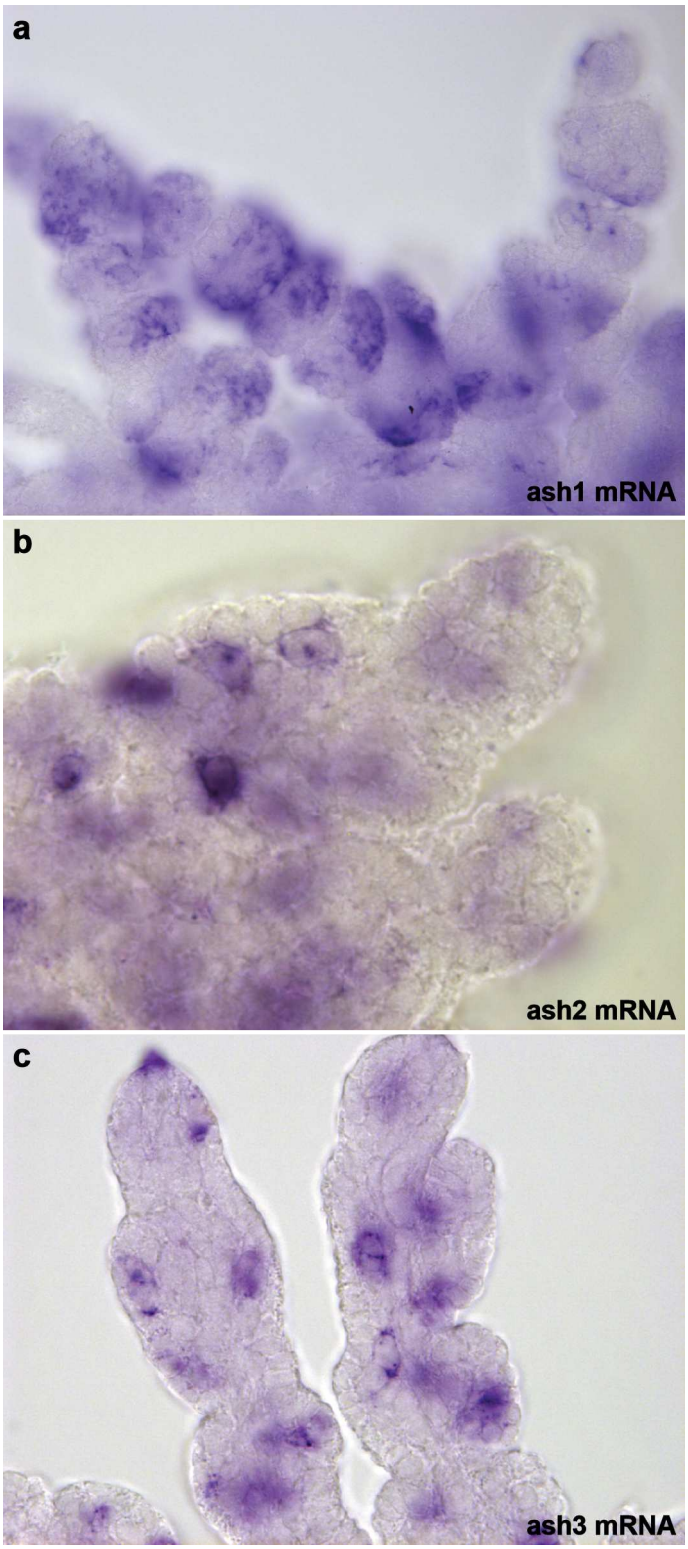


Figure 5.9: Expression pattern of *Parhyale* ash genes in the embryonic legs.
RNA *in situ* hybridization on embryos revealed that **(a)** ash1 transcript is present in large clusters of cells in the developing legs, whereas **(b)** *ash2* and **(c)** *ash3* expression is localized in smaller clusters and in isolated cells, which might be developing sensory organ cells.

Following these unsuccessful attempts to clone a neuron-specific CRE, I turned my attention towards the other sensilla cell types. None of the putative markers for glial cells were suitable for CRE-reporter cloning (Table 5.1), therefore I did not try to clone

a glial promoter. More recently, as more complete genomic and transcriptomic assemblies have become available for *Parhyale*, I revisited these genes using these updated resources. The sequence of the genomic locus of *gcm*, for which I previously had only a partial sequence corresponding to the 3' of the mRNA, is now available. The 5' of the *gcm* transcript has no peaks in the ATACseq datasets but this could be just because the 5' of the mRNA was truncated due to an assembly error, which is common. Around 2 kb upstream of the *gcm* gene there is an ATACseq peak which could mark TSS and the promoter, and another peak ~3 kb further upstream which could mark an additional CRE (supp. figure 5.1). Therefore, a ~6 kb fragment starting from the start codon may include the CREs of the *gcm* gene. Considering that *gcm* was recently shown to be expressed in the crayfish nervous system (Junkunlo et al., 2020), it is a good candidate for future trials.

I continued with marker genes for developing sensilla precursors rather than terminally differentiated cells and identified seven target genes: three *achaete*/*scute* homologs (ash), *sens*, *neur*, *cut*, and *pros*. Among the three *Parhyale* ash genes, *ash2* appears to be the ortholog of *ase* based on a previous classification (Ayyar et al., 2010). RNA in situ hybridization (RNAish) experiments revealed that *ash1* gene is expressed in clusters of cells in embryonic limbs, whereas *ash2* and *ash3* expressions are localized in smaller clusters of cells and in some parts in isolated cells (figure 5.9). I decided not to focus on *ash1* because it is broadly expressed; later analysis revealed that there are no clear ATACseq peaks around the 5' region of this gene (supp. figure 5.2a). *Ash3* is a good candidate based on its expression pattern, but due to an incomplete mRNA sequence, I could not identify the start codon. *ash3* gene locus has several ATACseq peaks (supp. Figure 5.2a), therefore this gene could be a good candidate in the future once the gene annotation is completed. *ash2* was also a promising candidate since its expression is restricted to small cell clusters in the limb which may correspond to developing sensory organs. I managed to clone ~4.2 kb upstream of the start codon of *ash2*, which includes a 90 bp long 5'UTR and this DNA segment covers the largest peak in the ATACseq profiles (figure 5.10). The corresponding reporter constructs did not drive any eGFP expression in the injected embryos and hatchlings. There are ATACseq peaks further upstream of the gene, which might correspond to additional CREs, but I did not pursue cloning larger fragments.

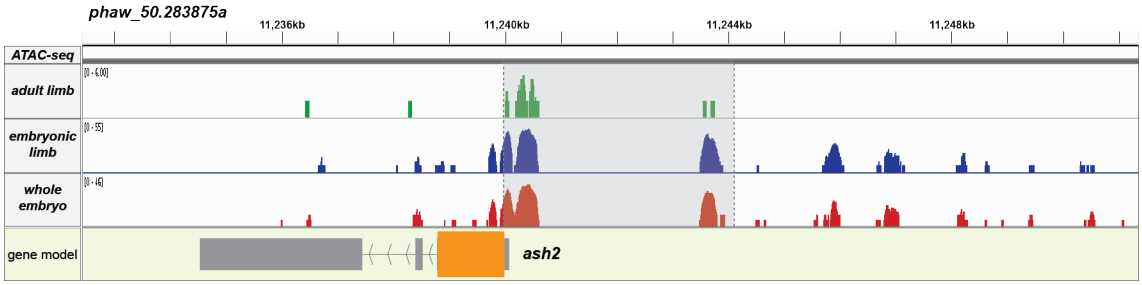


Figure 5.10: Overview of the *ash2* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *ash2* genomic locus. The gene model of *ash2*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

I previously explored the role of *sens* and *cut* in sensilla development as well as their expression patterns in developing *Parhyale* limbs (see chapter 4). I cloned 4.8 kb upstream of the start codon of *sens*, including 375 bp of 5'UTR and the only ATACseq peak near the 5' end of this locus (figure 5.11). This construct did not drive any reporter expression in the injected embryos. For the *cut* gene, I cloned a 3.9 kb DNA fragment upstream of the start codon (figure 5.12). This region was not sufficient to drive reporter expression when tested. However, the latest genome and transcriptome assemblies reveal that the *cut* transcript I used was incomplete and there is at least one more intron and exon upstream of the region I cloned. This explains why the reporter was not functional, it neither contains the TSS of the *cut* gene nor a core promoter.

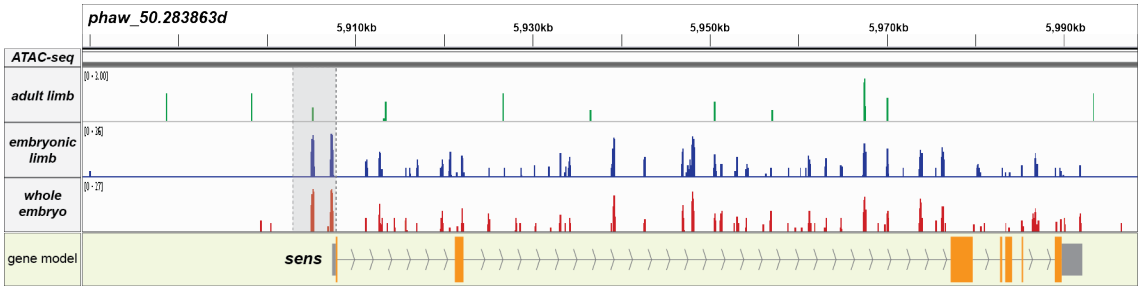


Figure 5.11: Overview of the *sens* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *sens* genomic locus. The gene model of *sens*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

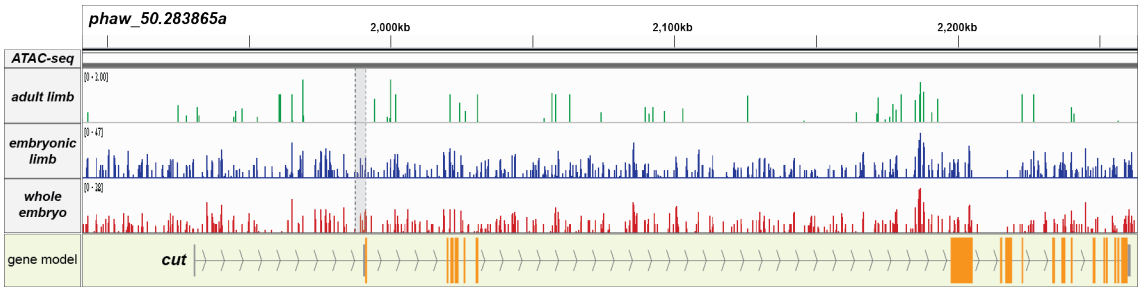


Figure 5.12: Overview of the *cut* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *cut* genomic locus. The gene model of *cut*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

Prospero (*pros*) is expressed in pIIb cells in *Drosophila*, the progenitors of sensory neurons during sensory organ development, and has similar expression in other arthropods (Klann and Stollewerk, 2017; Manning and Doe, 1999; Reddy and Rodrigues, 1999; Stollewerk and Seyfarth, 2008). RNAish revealed that the *Parhyale pros* is expressed in clusters of cells in the developing legs, which might correspond to the sensory neurons and/or to additional cells of the sensory organs (figure 5.13a). I cloned a ~5.5 kb DNA fragment upstream of the start codon, including the 282 bp 5'UTR (figure 5.13b), in front of eGFP and injected this construct. I could not detect any eGFP expression in embryos carrying this transgene. This construct includes some of the ATACseq peaks present in this region, but there are additional peaks located ~6 kb upstream of the TSS. I did not succeed in amplifying this extended upstream region but it can be pursued in the future.

My last candidate, *neur*, is part of the notch-delta signaling pathway. In *Drosophila* all the cells of a sensillum are generated by stereotypical divisions of a single cell, the sensory organ precursor, and *neur* is expressed in this cell (Yeh et al., 2000). The notch-delta signalling pathway is also crucial for sensilla development in other arthropods (Gold et al., 2009; Ungerer et al., 2012). I identified the *Parhyale neur* ortholog and revealed its expression pattern in developing limbs via RNAish (figure 5.14a). There are small clusters of cells expressing *neur* in the developing *Parhyale* limb and, considering that many sensory organs are organized in clusters in *Parhyale* limbs, these cells may correspond to the sensory organ cells. I cloned 5kb upstream of the start codon, including the ~700 bp 5'UTR and some of the ATACseq peaks in that region (figure 5.14b). As This construct did not show eGFP expression after injection. Cloning an additional 5kb fragment would include some other ATACseq peak regions, but I did not pursue this further.

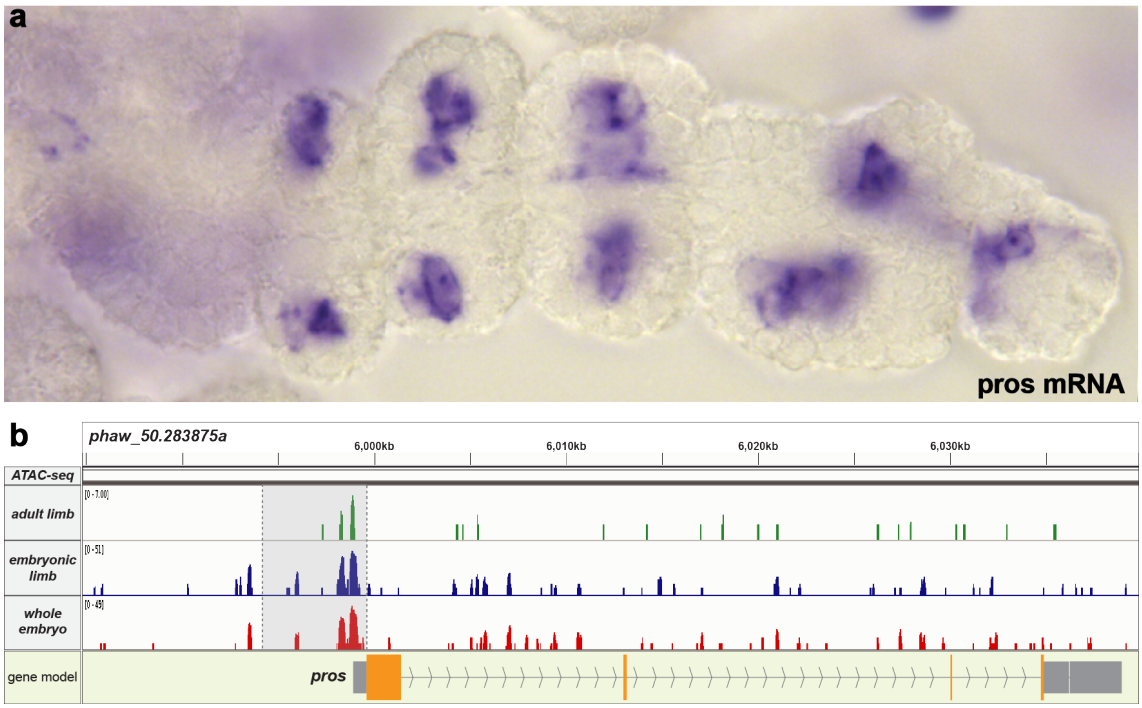


Figure 5.13: The expression pattern and overview of the genomic locus of *pros* gene. **(a)** RNA-*ish* on embryonic limbs shows that the *pros* transcript is expressed in clusters of 2-3 cells. **(b)** The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *pros* genomic locus. The gene model of *pros*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

The DC5 reporter and enhancer traps

Aside from cloning CREs to generate cell-type-specific reporters, which was very unfruitful in my hands, I also probed an existing enhancer-reporter construct, carrying the DC5 regulatory element. DC5 is an enhancer of a chick crystallin gene which is bound by the Pax2/Sox6 transcription factor heterodimer (Kamachi et al., 2001). The enhancer has also been shown to be active in flies (Blanco et al., 2005). The construct has also been tested in *Parhyale*, where it was observed to drive expression in the central nervous system {Konstantinides:2016wl}. In that study, the expression pattern was evaluated with a cytoplasmic DsRed, which revealed expression in nerve fibers in the legs of adult *Parhyale*. Analyzing the expression pattern of DC5 using a nuclear-localized reporter (histone H2B fused with eGFP) displayed that this element is also active in some nuclei outside of the CNS (Frederike Alwes, unpublished, figure 5.15): in young hatchlings, a single nucleus, located just below a microseta, is marked in the carpus of most thoracic legs, and there are several nuclei labeled in the antennules and the antennae. Putting these two observations together, it seems likely that these DC5-positive cells in the limbs correspond to a specific type of sensory neuron.

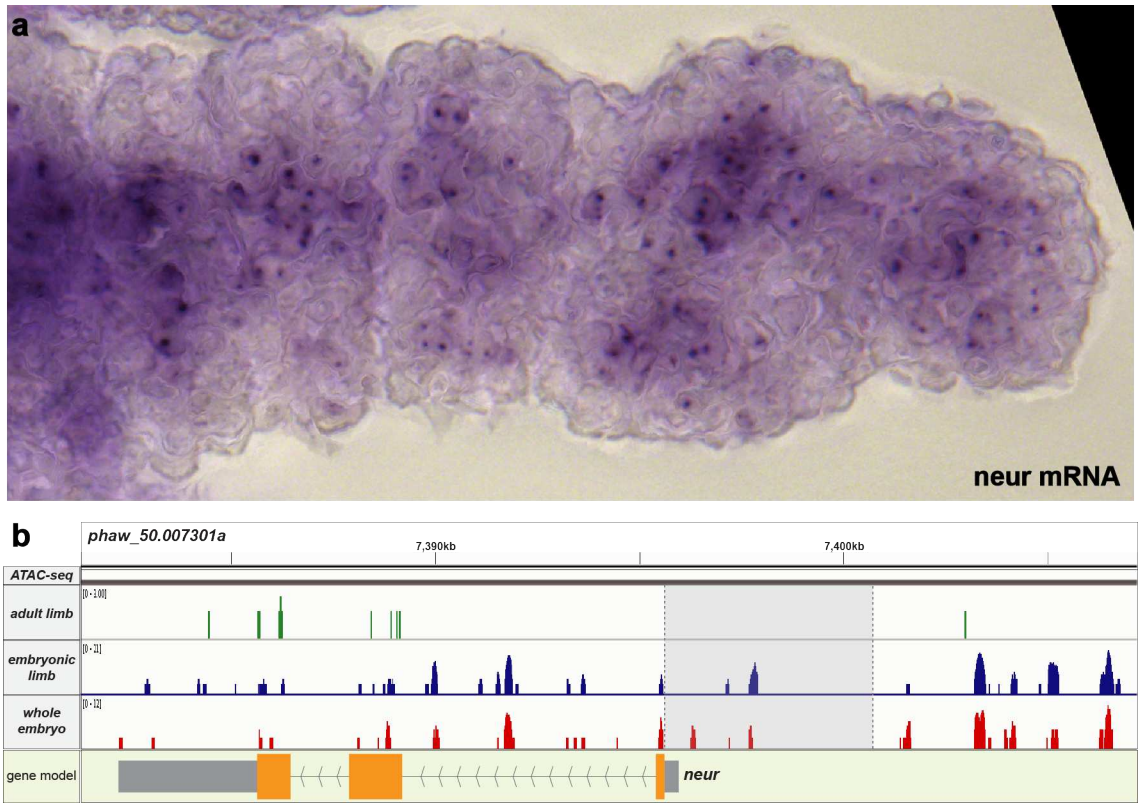


Figure 5.14: Neur gene expression pattern and overview of the genomic locus.
(a) RNA-ish on embryonic limbs shows that the neur transcript is expressed in some cells. (b) The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the neur genomic locus. The gene model of neur, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel. The cloned regions contained in the CRE reporter constructs are shaded with gray on the ATACseq panels.

This information about the DC5:H2B-EGFP reporter construct was very promising but the only data I had was a single image of a transgenic juvenile. Since we no longer had this transgenic line in the lab, I re-made this line by injecting the DC5-H2B::eGFP construct. In the transgenics, I generated, the expression in the antennules, antennae, and CNS was observed in all transgenic animals, but the expression in the single nucleus in the leg was present only in a fraction. I crossed the leg-positive DC5-H2B::eGFP animals with the PhHS-H2B::mRFPRuby animals in order to generate a stable line carrying the DC5 marker and the ubiquitous H2B::mRFPRuby fluorescence that we routinely use for cell tracking of nuclei in regenerating legs (see chapter 2). My aim was to explore whether this single nucleus belongs to a sensory cell, via antibody stainings, and if so, to investigate the behavior of this cell during limb regeneration and to identify its progenitors. Unfortunately, we had some issues with our animal stocks in the spring of 2019, which led to the loss of many animals. Both the G0 leg-positive animals and their progeny from the crosses died during that time. As discussed below, I think this work should be pursued in the future.

While screening DC5:H2B-eGFP injected animals, I observed that one individual, a female, had several nuclei labeled on each limb. A closer look revealed that the labeled nuclei were located within or just below the epidermal layer, and several of them had an elongated and curved shape (a croissant shape). The localization of these nuclei is evocative of their ectodermal origin, and the shape of these nuclei is similar to that of the accessory cells of the insect sensilla (figure 5.16). However, the labeled nuclei in the limbs were probably too many to be all associated with sensory organs.

Minos-mediated transgenesis inserts the transgene in random locations in the genome and each individual is likely to carry multiple independent insertions in its genome (Pavlopoulos and Averof, 2005). In this particular animal, a transgene is most likely to be trapping an endogenous enhancer activity at a specific insertion site. Since this putative enhancer trap – named *croissant* – could serve as a marker for sensory organs, I crossed this animal with PhHS-H2B::mRFPRuby animals to generate a stable line carrying the *croissant* insertion and the ubiquitous H2B::mRFPRuby marker used for cell tracking.

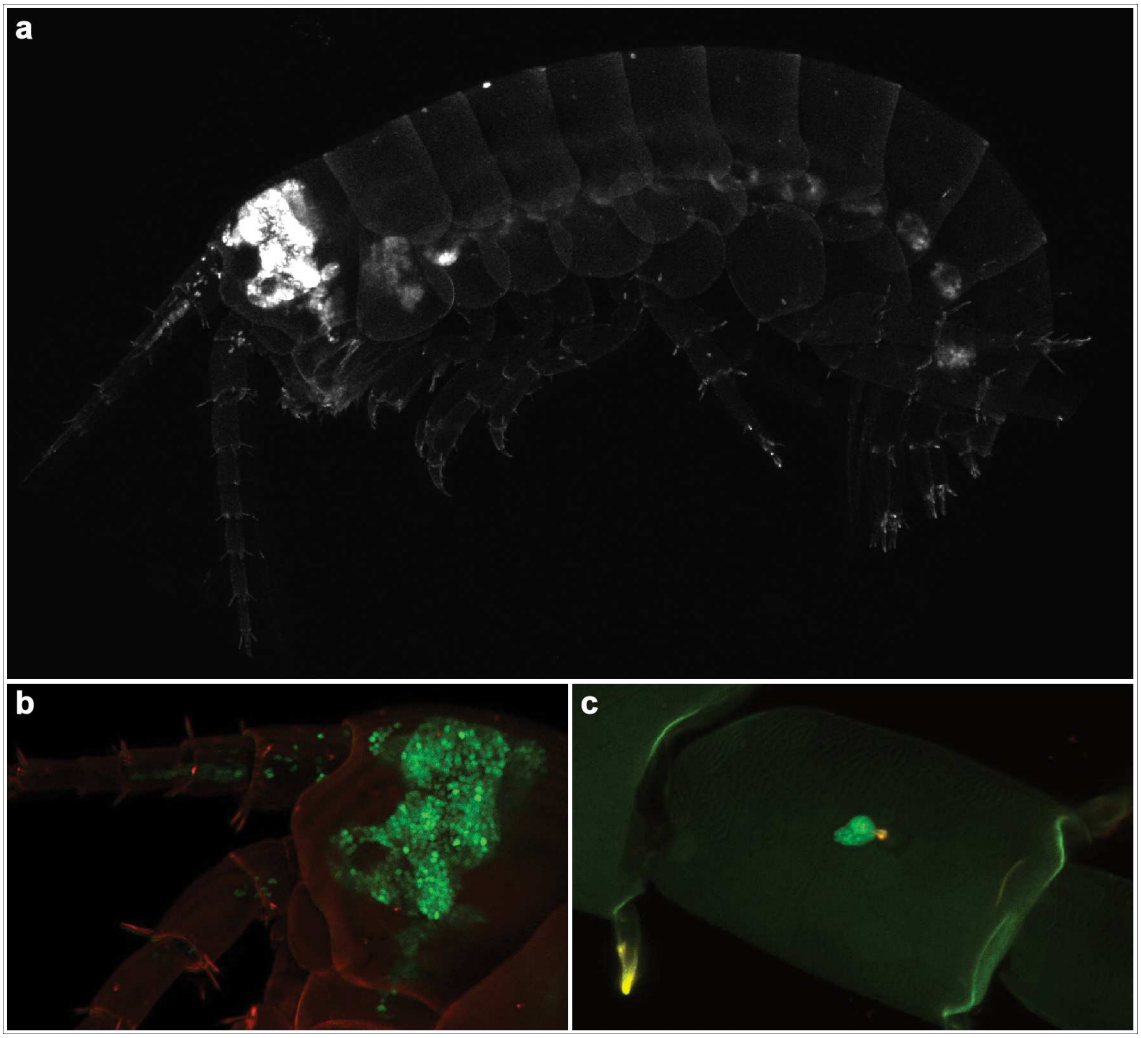


Figure 5.15: Expression pattern of DC5:H2B-eGFP reporter.
(a) A juvenile carrying the DC5:H2B-eGFP animal with many nuclei labeled in the head and the ventral nerve cord, as well as **(b)** the antennae, antennules, **(c)** and a single nucleus in the carpus of a T5 leg (image from Frederike Alwes).

I screened the progeny of this cross for eGFP expression in the limbs. I observed that the reporter is expressed in different patterns in the limbs of different individuals in the G1 and G2 generations (figure 5.17). The most likely explanation for this variation is that the founder G0 female had a mosaic germline, carrying different insertions that trap different enhancers, and different progeny inherited different subsets of these insertions.

I screened all of the adult animals in this line for the eGFP expression in their limbs and observed a range of different expression patterns (figure 5.17). There were some individuals with similar expression patterns, for example, individuals #1, #3, #6, #20, #23, #26, and #33, which have similar expressions in ventral rows of cells. My aim is now to establish new lines starting from couples displaying similar expression patterns. These enhancer trap lines could be used in the future to identify the labeled cell populations and to track their progenitors and cell fates during regeneration.

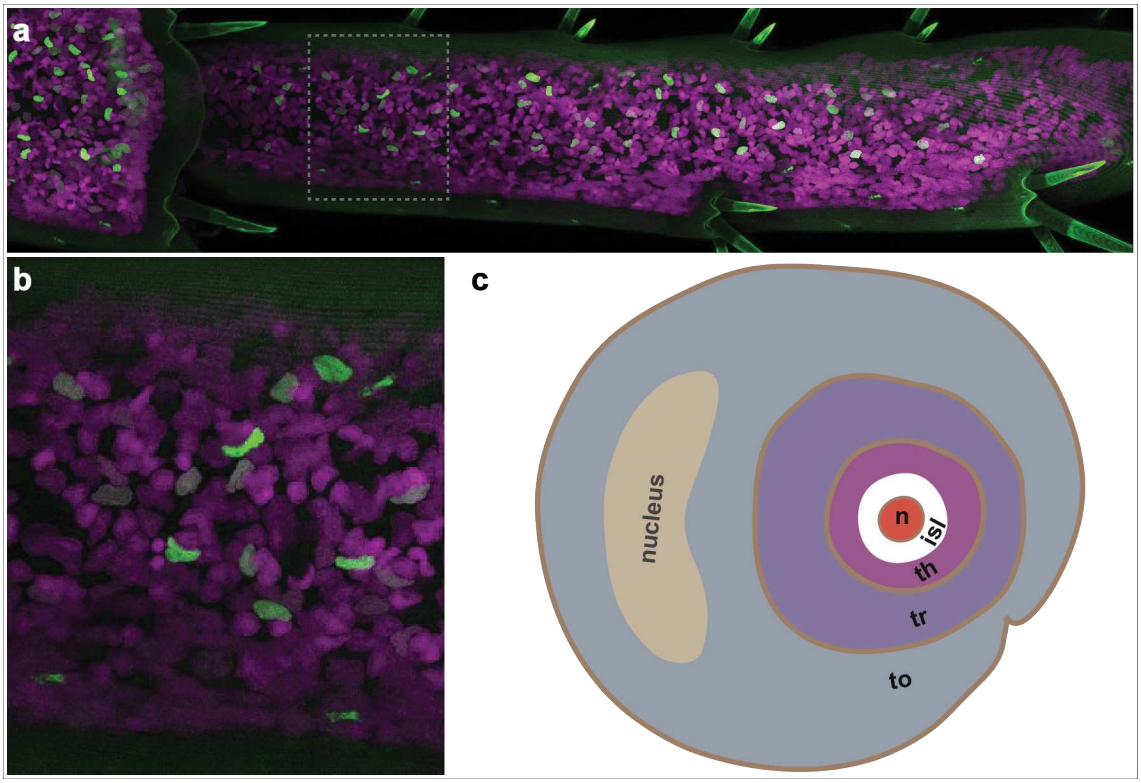


Figure 5.16: Expression pattern of the *croissant* putative enhancer trap reporter.
(a,b) The propodus of the T6 limb of the G0 croissant female has several eGFP-expressing nuclei some of which have a croissant shape. **(c)** Illustration of the cross-section of an insect external sensillum displaying the shape of the tormogen cell nucleus. to: tormogen cell, tr: tricogen cell, th: thecogen cell, isl: internal sensillum lymph space, n: neuron (adapted from Hartenstein, 2005).

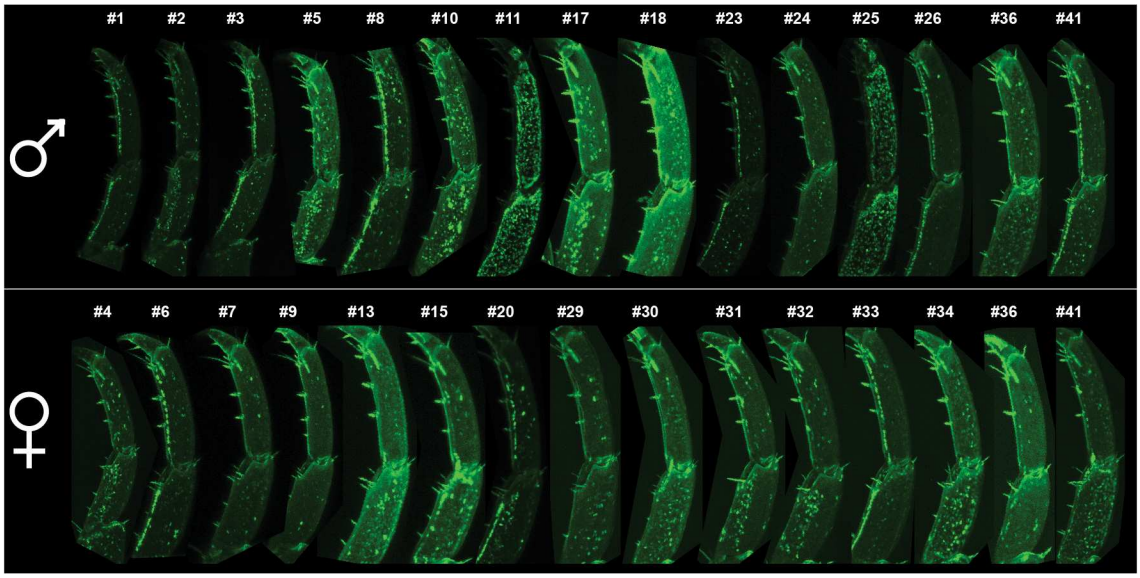


Figure 5.17: Expression pattern of the "croissant" enhancer trap in the G1 and G2 animals.
The expression pattern of the eGFP is variable in the T4 or T5 limbs of both male and female adults. The distal tip of the limbs is up, the ventral side is left.

Conclusions

In this part of my study, I tried to generate promoter reporters to label sensory organ cells in the *Parhyale* limbs. First, I identified putative marker genes for different cells of the sensory organs and then I attempted to clone the *cis*-regulatory elements that determine the expression of these genes. In order to prepare CRE-reporters, I cloned the upstream sequences of eight marker genes. To include the relevant CREs, I tried to clone the larger fragments possible, since I did not yet have any resources for predicting the CREs. In my hands, it was very challenging to clone large genomic fragments. Using high-fidelity DNA polymerases (*Pfu*, *Phusion*, and *Q5*) I could efficiently amplify large DNA fragments from plasmid DNA, but this was very inefficient on genomic DNA samples, most likely due to the complexity of the templates and impurities present in the DNA preparations. For this reason, I turned to *Taq* DNA polymerase, which is efficient for cloning fragments smaller than ~2.5 kb. As a result, the genomic regions to be tested were amplified as 0.8 to 2.5 kb fragments and then concatenated. Cloning the genomic regions in multiple fragments is not ideal for two reasons: first, it adds additional steps to the process, and second, it requires additional PCR primer pairs, which were not always functional, possibly due to polymorphisms in the *Parhyale* genome (Kao et al., 2016).

Only towards the end of this project, I managed to find a long-range DNA polymerase mix, LongAmp Taq DNA polymerase (NEB), working efficiently and reliably with *Parhyale* genomic DNA template. I used it to clone ~5 kb genomic fragments for the *pros* and *neur* reporters. This enzyme mix can be used in future studies to clone longer genomic DNA fragments.

For a more targeted strategy to locate putative cis-regulatory elements, I performed ATACseq on developing *Parhyale* embryos, embryonic legs, and adult legs. The ATACseq protocol was successful on embryos and I generated two good quality datasets, one from whole embryos and another from embryonic legs, which are the first ATACseq datasets obtained from *Parhyale* and will be a useful resource for the community beyond this study. The ATACseq dataset I generated from adult limbs was of poor quality, which suggests that a modified protocol is required for adult tissues. My colleague Emilia Skafida recently optimized both bulk and single-cell ATACseq protocols on adult limbs and generated good quality datasets. Analyzing the ATACseq datasets identified some ATACseq peaks that were not included in my reporter constructs and could be tested in the future.

The DC5 transgenic line could be used in the future to study the regeneration of a specific sensory organ on *Parhyale* limbs. Unfortunately, this line was lost

accidentally and will need to be generated again by re-injecting the construct, or by re-screening the enhancer trap line to identify if there are individuals only exhibiting the DC5 pattern. Once this line is established again, the identity of the labeled nuclei should be investigated. Antibody stainings with an acetylated tubulin antibody to label axons would identify whether these nuclei belong to the neurons. Two questions could be addressed with live imaging and cell tracking of DC5 animals: whether the marked cells contribute to the blastema (by amputating limbs just distal to the eGFP positive cells), and which are the progenitors for regenerating these cells (by amputating the region in which these cells reside).

The enhancer traps that I generated could also be used to mark different populations of cells and identify their progenitors. To carry out this analysis, transgenic lines with specific, reproducible patterns will need to be established from the parental line, which carries multiple transgene insertions, and the identity of the eGFP positive cells should be established. Stainings using antibodies labeling axons, or different sensory organ cells can be used to examine whether these cells are related to sensory organs (see Chapter 6 for details).

Supplementary Information

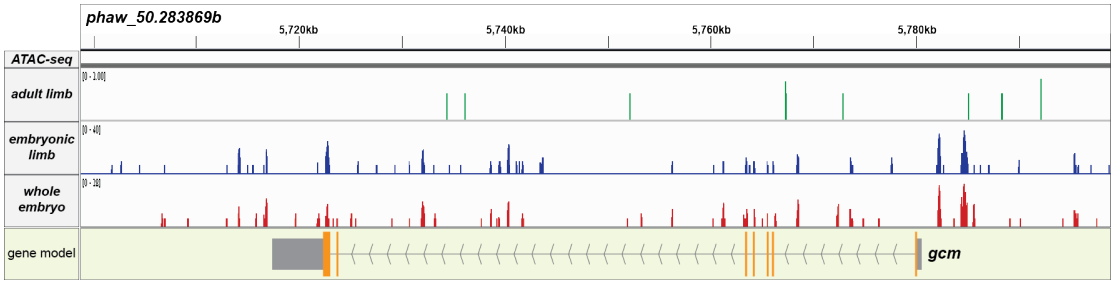


Figure S5.1: Overview of the *gcm* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the *gcm* genomic locus. The gene model of *gcm*, with the introns, exons (orange bars), and UTR (gray bars), is displayed at the bottom panel.

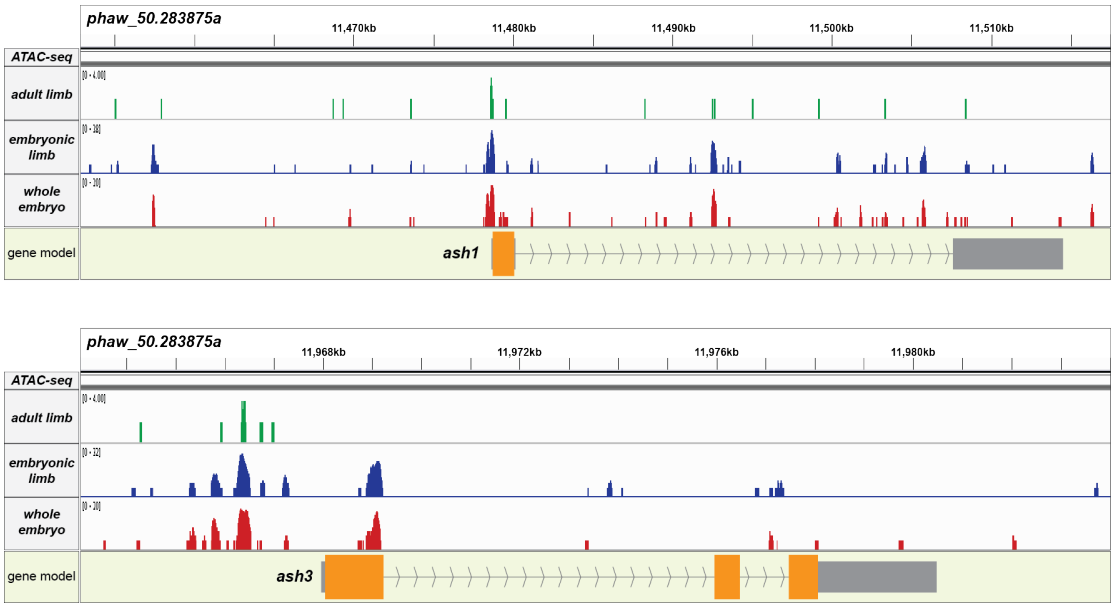


Figure S5.2: Overview of the *ash1* and *ash3* genomic locus with putative CREs.
The peaks in the ATACseq datasets of adult limb (green), embryonic limb (blue), and whole embryo (red) showing the putative CREs in the ash1 and ash3 genomic loci. The gene models, with the introns, exons (orange bars), and UTR (gray bars), are displayed at the bottom panels.

6. Combining live imaging with antibody staining to track the progenitors of *Parhyale* limb regeneration

Introduction

Previously, I tried to generate transgenic animals with cell type-specific markers using CRISPR-mediated knock-in and CRE-reporter approaches, however, both transgenic approaches were unfruitful in my hands. As an alternative, I decided to try to identify the fate of cells in regenerated limbs post-imaging, via fixing and staining of the legs to visualize cell-type-specific markers. For the success of this approach, the images of the stained limbs should be comparable to that of live recordings of the regeneration. There are two major techniques commonly used for fluorescent labeling different cell types on fixed tissue samples: fluorescent RNA *in situ* hybridization (RNA-*fish*), in which specific transcripts are labeled, and immunohistofluorescence (IHF) that is mainly used to label the proteins. Both techniques are labeling the tissue with fluorescent dyes, which can be imaged on a confocal microscope to generate 3D images of the stained samples, which should be comparable with the live image recordings of regenerating legs. RNA-*fish* operates via two complementary strands of nucleic acid annealing to each other to form a hybrid duplex (Felsenfeld and Miles, 1967). The transcription of a gene is marked by hybridizing a labeled complementary RNA probe to an mRNA. There are various RNA-*fish* protocols, but they all share the same principle and core steps. First, there are pre-hybridization steps to prepare the samples for hybridization: fixation and permeabilization, *e.g.* using chemical fixation and detergent washes or nonspecific protease treatments. The hybridization step calls for high temperatures and the use of formamide to prevent nonspecific hybridization. And finally, the hybridized mRNA-probe hybrid is visualized, often through an antibody (reviewed in (Young et al., 2020)). As a result of the cumulative effects of different RNA-*fish* steps, the morphology of the tissues can deteriorate considerably. I decided to focus on the IHF methodology since it is generally faster, more gentle to the tissue, and therefore more promising for labeling cells that can also be identified in the live imaging recordings.

The IHF method relies on the use of antibodies that recognise specifically a molecule of interest. To generate specific antibodies, a target macromolecule is delivered to animals by injection. Once an immune response is developed, B cells generate specific antibodies for the injected molecule, which can be harvested in the blood. Each type (clone) of B cells can produce a single type of antibody, recognizing a

small part of the antigen called an epitope and immunized animals usually activate multiple B cell clones, producing antibodies that recognize different epitopes of an antigen. Therefore, antibodies harvested from the blood serum are called 'polyclonal', containing antibodies produced by multiple B cell clones. If the B cells from immunized animals are isolated and fused with myeloma cells, an immortalized antibody-producing cell line, a hybridoma, can be produced. This way antibodies derived from single B cell clones, called monoclonal antibodies, can be produced in large quantities.

Once antibodies have been produced, they can be applied to cells and tissues to bind and label the corresponding antigens. In the direct IHF technique, the antibodies are conjugated with fluorophore molecules before binding. Alternatively, after binding with the antigen, these antibodies can be bound and labeled by fluorophore-tagged secondary antibodies, a method called indirect IHF. Both of these methods are tagging the antigen, but the latter method is more commonly used for two reasons. First, there is a wide range of commercially available secondary antibodies to recognize antibodies raised in various species. Therefore, just developing a specific antibody and acquiring a suitable secondary antibody is enough to perform IHF, which makes the process simpler. Second, multiple secondary antibodies can bind to each primary antibody molecule producing a signal amplification, which makes the detection easier.

The methodology of staining tissues with antibodies

IHF protocols are adjusted to the type of tissues and the antibodies used. Despite the range of protocols available, there is a set of steps that are common to most, including pre-staining (fixation and blocking) and antibody staining steps. Additionally, some tissues require a supplementary permeabilization step after fixation. The fixation step is essential for preventing tissue degradation during and after the staining protocol. The most common fixative used for IHF is paraformaldehyde (PFA), which is a cross-linking fixative that forms covalent links between macromolecules. This ensures the molecular components of the tissues are fixed in place and enzymatic degradation is minimized (Eltoum et al., 2001). The duration of PFA treatment, as well as concentration and the temperature, should be adjusted to avoid over-fixing the tissue, which makes antigens inaccessible for the antibodies. Some of the antigens can become accessible again with an additional antigen retrieval step, but some antigens are resistant to this unmasking (Howat and Wilson, 2014). Therefore, optimizing the fixation step and removing the fixative completely with several washes after fixation, is essential for a successful IHF.

Depending on the localization of their targets, antibodies should penetrate through the cuticle, extracellular matrix, cell, and nuclear membranes. Considering the antibodies are large molecules, most tissues require to be permeabilized to facilitate antibody penetration. For microsectioned or dissected soft tissues, the addition of a detergent at low concentration, *e.g.* 0.1% Triton X-100, into the washing and staining solutions is sufficient. However, some tissues require longer and stronger detergent treatments, or their integument needs to be permeabilized enzymatically or mechanically. For example, in a whole-mount *Drosophila* antenna IHF protocol, 3% Triton X-100 is used to permeabilize the tissue (Saina and Benton, 2013). A whole-mount *Drosophila* embryo staining protocol combines methanol treatment, and enzymatic digestion with chitinase, as permeabilization steps {Manning:2016da}. There are also protocols using sonication to crack the cuticle of *Drosophila* larvae (Patel, 1994) and cuticularised insect limbs (Ehrhardt et al., 2015). These treatments are permeabilizing the samples by introducing microscopic holes or cracks to the cuticle and denaturing/degrading some macromolecules, both of which alter tissue morphology. Therefore the permeabilization step should be adjusted carefully for each type of sample, and the chemical or enzymatic agents, if used, should be washed away before proceeding with the blocking step.

Primary and secondary antibodies can sometimes bind to molecules other than their intended epitopes. In order to prevent this unspecific binding, the tissue is pre-blocked using protein-rich solutions. Blocking solutions often contain normal goat serum (NGS) and/or bovine serum albumin (BSA). There are some studies showing that a blocking step does not decrease the nonspecific antibody binding significantly (Buchwalow et al., 2011), yet most IHF protocols call for it.

Once the tissue is prepared as described above, the samples are ready to be stained. The samples are incubated with the antibodies, generally in the blocking solution, to reduce nonspecific binding and to prevent antibody degradation. The duration of the antibody incubation is adjusted based on the tissue to be stained. It is crucial to wash away the unbound antibodies from the tissue to avoid nonspecific signals before imaging the sample.

Antibody staining in *Parhyale*

Antibody staining on *Parhyale* embryos, before cuticle deposition, is a well established technique (Rehm et al., 2009) and has been used in several studies (Alwes et al., 2016; Pavlopoulos et al., 2009; Serano et al., 2016). On cuticularized limbs, it has been performed only in fragmented legs or in cryosections with long antibody incubations

(Konstantinides and Averof, 2014). But fragmented legs can be stained properly only by dissecting individual podomeres into several pieces. Previous attempts to stain larger leg fragments by permeabilizing the cuticle enzymatically, using chitinase, or by chitinase in combination with a decalcification step in an EDTA solution, had limited success (Chiara Sinigaglia, personal communication). Therefore, we are still lacking a robust method to label adult *Parhyale* limbs.

Besides the problem of limited permeability, IHF approaches in *Parhyale* are also limited by the relative lack of available antibodies to stain *Parhyale* tissues. In this study, I aimed to establish a protocol for antibody staining in regenerating *Parhyale* limbs to identify the fate of the cells in regenerated legs. Next, I aimed to develop antibodies that would allow me to label specific cell types. Finally, I aim to use these antibodies on live-imaged regenerated legs to identify the progenitors of different cell types via cell tracking.

Results and Discussion

Developing an antibody staining protocol for regenerating legs

There are two important points to consider while developing an IHF protocol for regenerating limbs. First, the morphology of the limbs should be conserved well during the protocol, so that I should be able to register the individual nuclei in the images of antibody stained leg to live images. Second, the protocol should be robust and work for every specimen. The live-imaged regenerated limb specimens are highly precious – it takes on average 10 days to live image a regenerating limb – thus I should be able to stain them very reliably.

The characteristics of the specimen, such as dimensions and permeability, are the key factors to consider while optimizing IHF steps. Therefore, the tissue preparation step has a serious effect on the success of the protocol. I have already standardized the amputations for live imaging (see chapter 3 for details), and I aimed also to standardize the tissue dissection -post imaging for antibody stainings. With the type of amputations I am performing for live imaging, the regenerated limb develops in the leftover stump of the carpus podomere, meaning that dissecting the carpus is sufficient to stain all of the regenerating tissue. Therefore, I cut the legs at the proximal side of the carpus, just distal to the podomere junction, to maximize the antibody penetration by maximizing the length of the cut. I also make an incision on the melanized scab covering the wound plane, using a sharp tungsten needle. This way, the antibody can reach the regenerating leg from two sides.

Following the protocol of antibody staining on *Parhyale* embryos (Rehm et al., 2009), I decided to use 3.6% formaldehyde in artificial seawater as a fixative. In my experience, the optimal fixation time is 20 minutes, which makes the tissue stable enough to maintain its morphology throughout the staining procedure, without reducing the accessibility of the antigens. Formaldehyde fixation, unlike acetone or methanol, does not permeabilize the tissues (Howat and Wilson, 2014; Jamur and Oliver, 2009), which means that, generally, an additional permeabilization step is required.

Sonication has been shown to be effective for permeabilizing soft and non-calcified cuticles (Ehrhardt et al., 2015; Manning and Doe, 2017; Patel, 1994). However, my previous trials on intact *Parhyale* limbs revealed that sonication fractures the cuticle rather randomly. Only around half of the limbs get their cuticles permeabilized at random positions. And in most of the limbs, cuticle breaks were not providing enough permeabilization for antibodies to stain entire podomeres. Therefore, I tried different detergent combinations and varying concentrations for permeabilization.

0.1% of Triton-X100 combined with 0.1% of either digitonin or sodium deoxycholate (DOC) were the best combinations to permeabilize the tissue without altering the morphology. Since digitonin is toxic, I decided to continue with Triton-X100 and DOC. To shorten the protocol, I combined permeabilization and blocking steps, and did not notice any drawbacks. Finally, I optimized the antibody staining steps. Initially, I used 7 days of primary and another 7 days of secondary antibody stainings, and gradually shortened these steps. In the end, reducing the antibody staining steps to 3 days each does not compromise the efficiency of the protocol. Shortening those times further does not work well for some antibodies.

I used two monoclonal antibodies, mouse monoclonal 6-11B-1 and rabbit monoclonal RM318, both raised against acetylated tubulin, to improve the protocols as mentioned above. Next, I tested other antibodies likely to be working on *Parhyale* tissues, including antibodies raised against tyrosinated tubulin, Lamin Dm0, Pax3/7, and Glutamine synthetase, to test the robustness of the protocol (see material and methods for details). I amputated T4 and T5 limbs of wild-type animals, dissected and fixed the limbs 7 days post-amputation (dpa), and performed antibody stainings with the above-mentioned antibodies (figure 6.1). All of the stainings worked very well, which shows that the protocol is robust and can be used to label cells of live imaged legs.

Registration of nuclei between live imaging and IHF stainings

It is fundamental to identify the same nuclei both in antibody staining images and in the live recordings to be able to track the progenitors of cell types identified with IHF. However, there are some obstacles that make nuclei registration challenging. First, during live imaging the nuclei are labeled with H2B::mRFPRuby (see chapter 3 for details) but after antibody staining, they are labeled with DAPI. This results in some minor differences in nuclear shape and brightness in the images. Second, the antibody staining protocol causes some mild changes in the limb morphology. The tissue slightly shrinks; this effect is rather uniform throughout the limb tissue, therefore it does not alter the positions of the nuclei in the podomeres but the relative positions of regenerated podomeres may change. The optical properties of the limbs are different once they are mounted after antibody staining, and this can lead to differences in imaging, especially in the axial (z) dimension (see chapter 3 for details). Finally, slight differences in mounting orientation could affect my ability to register the nuclei in live and stained samples.

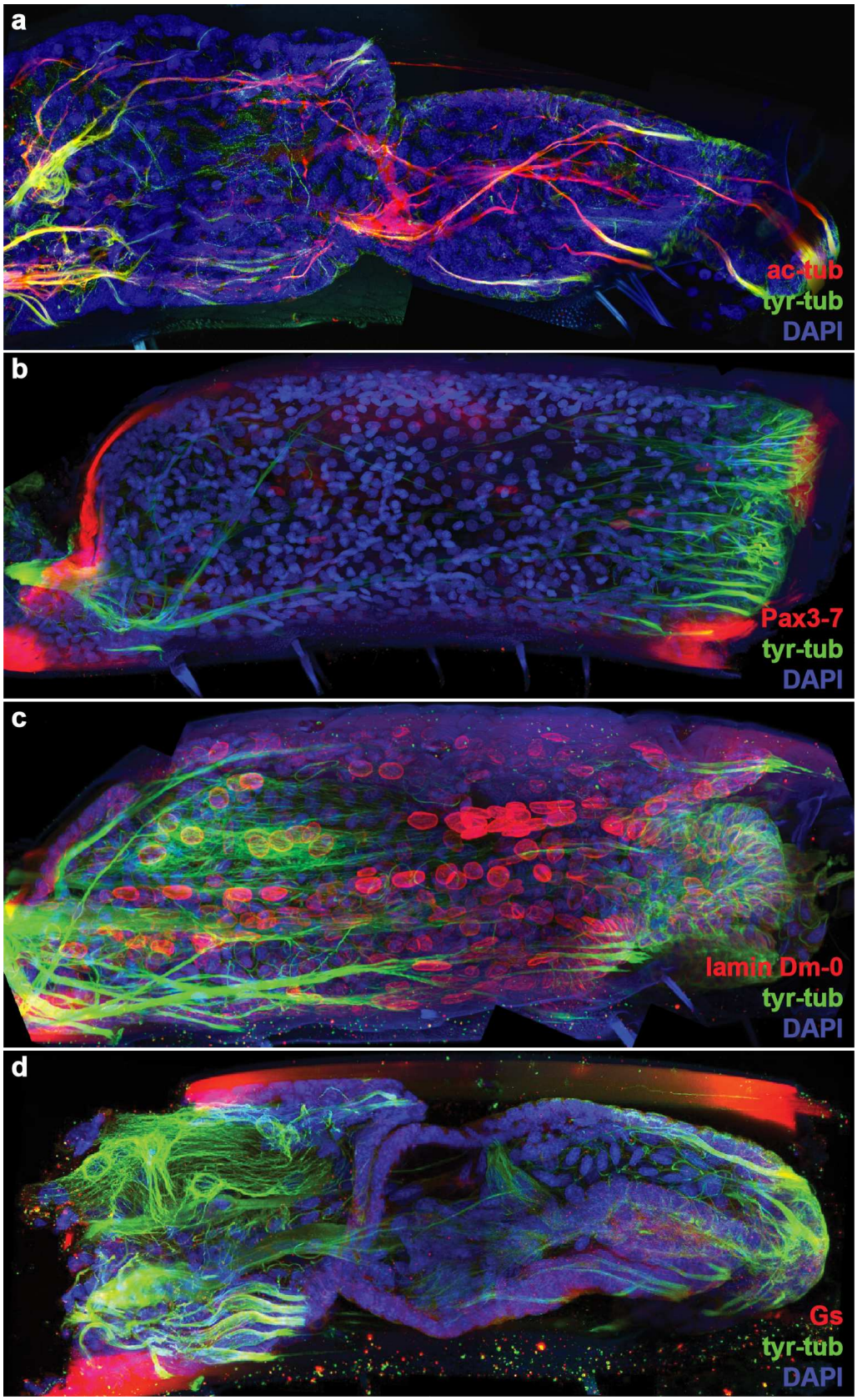


Figure 6.1: IHF on *Parhyale* legs 7 days post amputation, with different antibodies.
(a) The neuronal axons are labeled both with acetylated tubulin and tyrosinated tubulin antibodies. The tyrosinated tubulin antibody labels also additional features. (b) Pax3/7 antibody labels weakly the nuclei of satellite-like cells. Note that in this sample leg regeneration has not started yet. (c) lamin Dm-0 antibody labels the nuclear envelope, and (d) Glutamine synthetase antibody stains the cytoplasm of almost all of the cells.

In spite of these problems, I have consistently been able to recognize the nuclei marked by antibody stainings with their counterparts in the live recordings, with high confidence (figure 6.2). The change in the nuclear morphology and the signal intensity is not too large to prevent me from recognizing them. The limited movement of some podomeres during antibody staining (note the dactylus rotated clockwise in figure 6.2b) makes it challenging to register the nuclei on the podomere junctions, but not the nuclei that lie within each podomere. The biggest challenge is the difference in the axial flattening of the images resulting from the aberrations caused by the refractive index (*n*) mismatch. When there is a difference between the *n* of immersion medium of the objective and the *n* of the mounting medium, the distance of the object from the cover slip is misrepresented in the acquired image. The strength of this axial aberration depends on the extent of the refractive index mismatch, and it is more pronounced as the distance from the coverslip increases (Diaspro et al., 2002).

My live imaging experiments are performed with a dry objective and the *n* of air is 1, whereas the *n* of live cells are ~1.4 (Liang et al., 2007) and there is sea water and surgical glue between the tissue and the cover slip, and it's hard to the *n* of this combination. After the antibody staining, the limbs and the mounting medium have similar refractive indexes (*n*=1.45). Therefore the aberrations are different when the same limb is imaged under these two different circumstances. It can be seen clearly on the lateral sides of the leg in figure 6.2 (upper and lower parts of the images), where the visibility of the nuclei is different in the images of the live and stained limb. The aberrations are more visible in this part of the leg because the tissue is located in deeper sections due to the cuticle curvature. Yet, despite these problems, some of the nuclei are very easy to recognize in both images due to their specific shape and localization (circles in figure 6.2). By using these nuclei as landmarks, it is possible to register almost all of the nuclei in both images.

Overall, these results show that the antibody protocol is working reliably on regenerating limbs and can be used to label cells after live imaging. The next step is to develop antibodies specifically labeling the sensory organ cells of *Parhyale*.

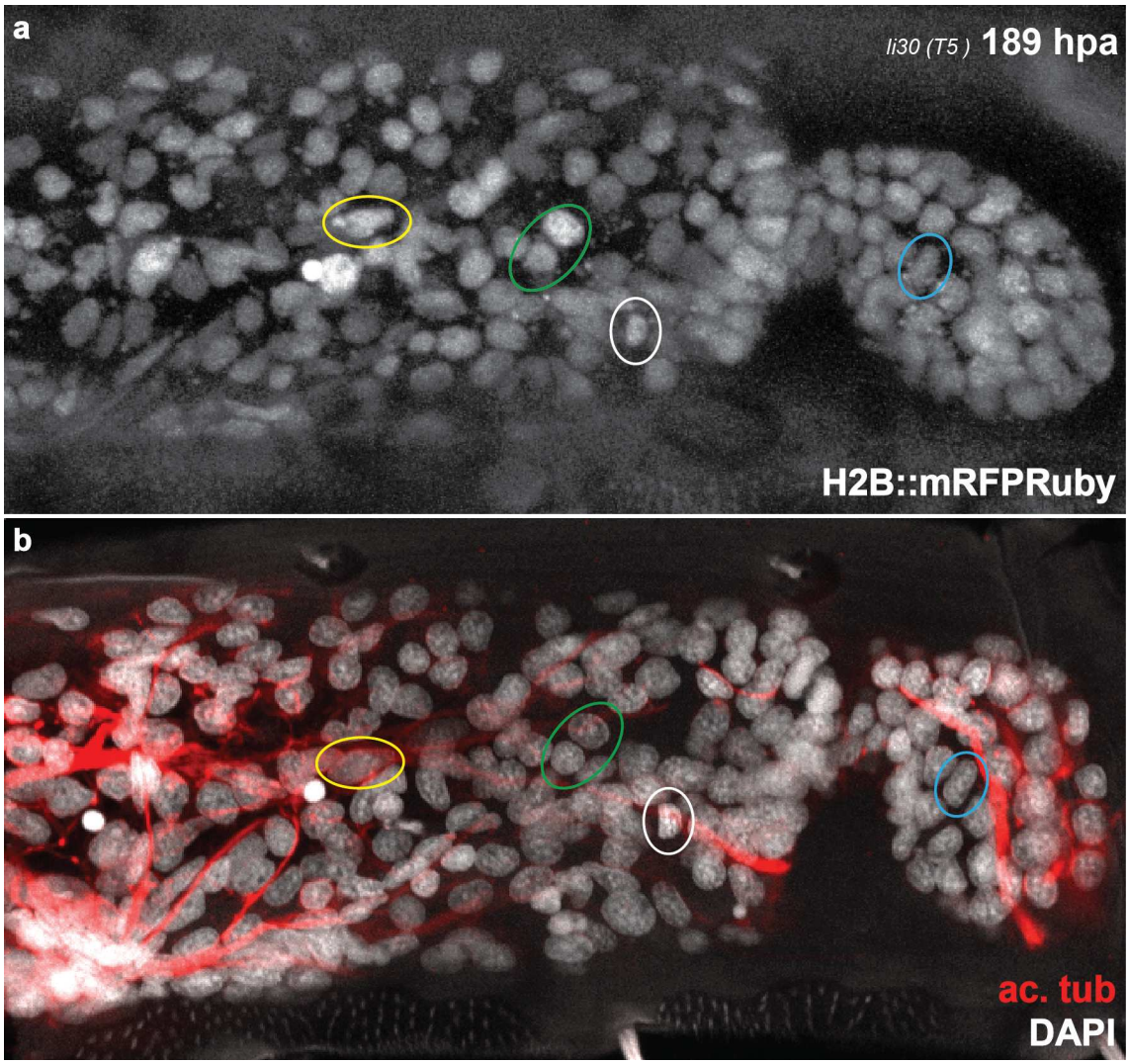


Figure 6.2: An antibody stained T5 leg after live imaging
An amputated T5 leg was imaged live until the final stages of regeneration and fixed before molting. Axons were labelled by IHF using monoclonal antibody 6-11B-1, which recognizes acetylated tubulin. The same nuclei can be identified in (a) the last frame of the live recording and (b) in the same specimen stained by IHF (e.g. nuclei marked by coloured circles).

Putative markers for recognizing sensory organ cell types using antibodies

By the time I had optimized the antibody staining protocol, my colleague Alba Almazán had developed a single nuclei RNA sequencing (snRNAseq) method for adult *Parhyale* legs and generated a good quality dataset. Based on their transcriptomes, limb nuclei can be clustered into 16 groups, which are likely to represent different cell types (figure 6.3a). Among these, we have so far identified cell clusters corresponding to epidermal, muscle, neuronal and blood cells, based on conserved genes whose expression is associated with these cell types (Alba Almazán, unpublished). In order to identify the cell types that constitute the sensory organs, besides neurons, I performed a literature search for the genes expressed in the sensory organs of arthropods (mainly *D. melanogaster*) and identified the expression of their *Parhyale* ortholog in the snRNA-seq dataset (Table 6.1).

Table 6.1: The genes screened in this study.

	gene	transcript ID
socket cell	*hairless*	no *Parhyale* ortholog
	tropomodulin	mikado.phaw_50.283355aG638
	supressor of hairless	mikado.phaw_50.283826G114
	singed	mikado.phaw_50.000203bG828
	SoxF	mikado.phaw_50.283828dG54
shaft cell	*artichoke*	mikado.phaw_50.283874aG488
	spineless	mikado.phaw_50.000135aG10
	shaven	mikado.phaw_50.282639aG316
sheath cell	*lmp*	mikado.phaw_50.283867cG584
	hamlet-1	mikado.phaw_50.283868bG396
	hamlet-2	mikado.phaw_50.283869cG784
	hamlet-3	mikado.phaw_50.283864cG30
	hamlet-4	mikado.phaw_50.283869cG786
	prospero	mikado.phaw_50.283875aG24 mikado.phaw_50.283875aG244
scolapale cell	*Trpγ*	not included in the r5.1 assembly
	contactin	mikado.phaw_50.282396bG28
	CAP	mikado.phaw_50.283873bG144
	neuroglian-1	mikado.phaw_50.282260aG238
	neuroglian-2	*mikado.phaw_50.283821G226*
	ubr3	*mikado.phaw_50.000081aG180*
	neurexin-IV-1	*mikado.phaw_50.000214eG76*
	neurexin-IV-1	*mikado.phaw_50.000214eG78*
	nervana	*mikado.phaw_50.282543aG130*
	Calx-1	*mikado.phaw_50.282861eG256*
	Calx-2	*mikado.phaw_50.283872cG352*
	crumbs	*mikado.phaw_50.283850aG514*

	gene	transcript ID
dendritic cap cell	*painless*	*mikado.phaw_50.282654bG1710*
	cornetto	*no Parhyale ortholog*
	distroglycan-1	*mikado.phaw_50.283861bG912*
	distroglycan-2	*mikado.phaw_50.283861bG908*
ligament cell	*POU-1*	*mikado.phaw_50.283869dG190*
	POU-2	*mikado.phaw_50.282396bG10*
	POU-3	*mikado.phaw_50.283826G1238*
	POU-4	*mikado.phaw_50.282861eG500*
	repo	*mikado.phaw_50.282997aG292*
	sr	*mikado.phaw_50.000135dG14*
tendon cell	*held out wings*	*mikado.phaw_50.282260aG736*
	short stop-1	*mikado.phaw_50.283870G888*
	short stop-2	*mikado.phaw_50.283870G892*
	myspheroid-1	*mikado.phaw_50.282275bG140*
	myspheroid-2	*mikado.phaw_50.283865bG202*
	myspheroid-3	*mikado.phaw_50.283743aG402*
	APC-1	*mikado.phaw_50.007301aG1512*
	APC-2	*mikado.phaw_50.007301aG1516*
	lrt	*mikado.phaw_50.283850aG1106*
	eb1	*mikado.phaw_50.283862aG156*
	by	*mikado.phaw_50.283872cG578*
	trombospondin-1	*mikado.phaw_50.283826G2062*
	trombospondin-2	*mikado.phaw_50.283826G2056* *mikado.phaw_50.283826G2064*
	fmr1	*mikado.phaw_50.282861cG222* *mikado.phaw_50.282861cG220*

SoxF is a transcription factor expressed in the socket (tormogen) cells of *Drosophila* mechanoreceptors (Miller et al., 2009); in the house spider *P. tepidariorum*, SoxF is expressed in small clusters of cells in developing limbs, which may also represent the developing sensory organs (Baudouin-Gonzalez et al., 2020). The *Parhyale* SoxF ortholog is expressed mainly by a single cluster of cells . Among the differentially expressed genes in this cluster I also found the orthologs of sensory organ related genes *emc* and *nompA*. Therefore, this cluster may represent the socket cells. I decided to develop antibodies against the *Parhyale* SoxF and Emc-1 proteins.

Pros is expressed in pIIb cells, the mother cell of sensory neurons and sheath or scolapale cells in *Drosophila* sensilla. Later in development, *pros* expression persists in the sensory neurons but ceases in the sheath cells (Manning and Doe, 1999; Reddy and Rodrigues, 1999). *Pros* is also expressed by developing peripheral sensory organs in spiders (Stollewerk and Seyfarth, 2008) and water fleas (Klann and Stollewerk, 2017). The *Parhyale pros* ortholog is expressed in two clusters. The neuronal clusters were already identified, therefore *pros*-positive two clusters could

represent the sheath cells of different types of sensory organs. Alternatively, only one of the clusters could be related to sensory organs and *pros* might be also expressed in an additional cell type.

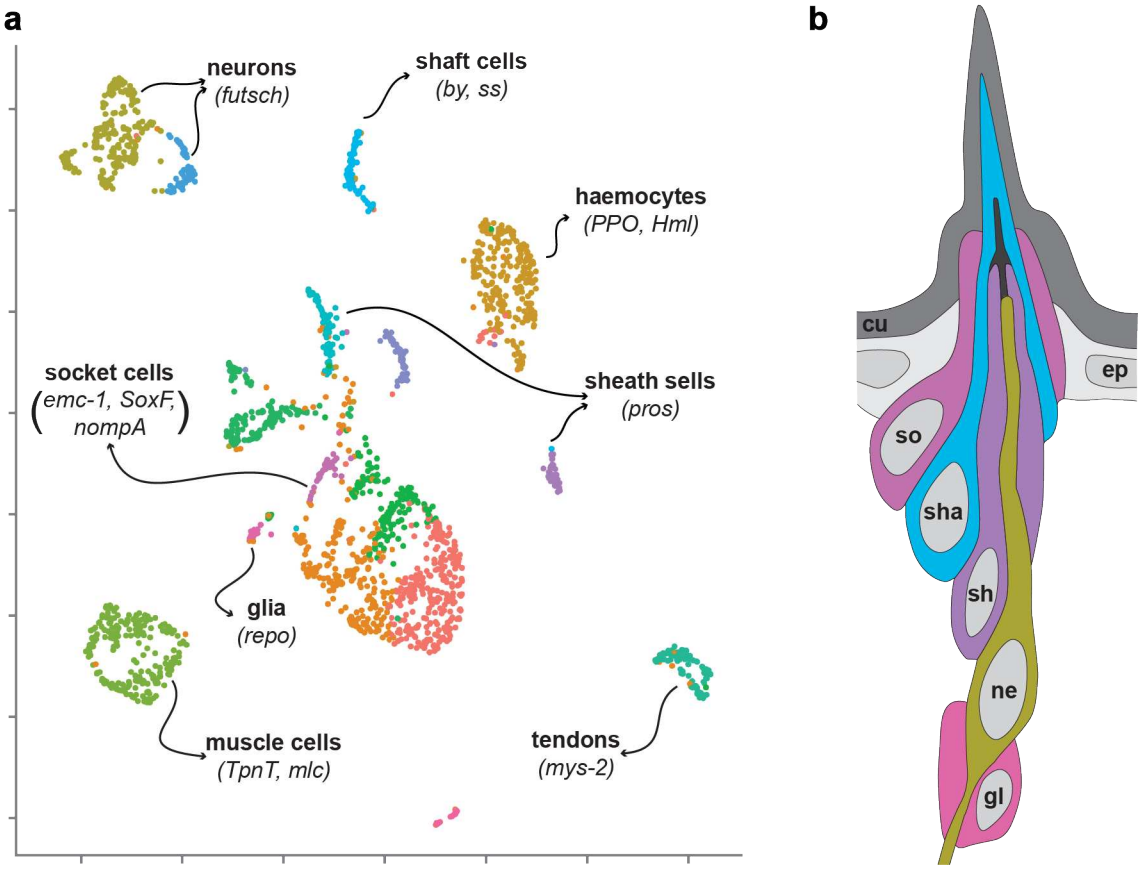


Figure 6.3: Identifying the clusters corresponding to components of *Parhyale* sensory organs based on putative markers in the snRNAseq dataset.
(a) Nonlinear dimensionality reduction of the snRNAseq dataset with UMAP, displaying 16 clusters of nuclei (data and image from Alba Almazán). Clusters corresponding to putative sensory organ cells, haemocytes, and tendon cells are labeled. (b) Illustration of the cellular organization of a crayfish peripheral mechanosensory organ (based on the original work of (Kouyama and Shimozawa, 1982), illustration modified from (Hartenstein, 2005)) cu, cuticle; ep, epidermis; so, socket cell; sha, shaft cell; sh, sheat cell; ne, neuron; gl, glia.

In a recent study, an antibody was developed against Pros in the butterfly *Papilio xuthus* (Perry et al., 2016), using a part of the Pros protein that is highly conserved between the butterfly and *Parhyale* (figure 6.4a). I tested this PxPros antibody by IHF on *Parhyale* embryos and regenerating legs (figure 6.4b-c). Several nuclei are stained in both samples and some of the labelled cells are clearly associated with nerve fibers. This suggests that at least one of the *pros*-positive clusters in the snRNAseq dataset is related to sensory organs. The available quantity of the PxPros antibody is limited, therefore I decided to raise a specific antibody targeting *Parhyale* Pros.

Finally, *repo*, a putative marker for glial cells (see chapter 4 for details), is expressed in a single cluster of cells. In *Drosophila*, some of the sensilla cell lineages

include a glial cell (see chapter 1, reviewed in (Hartenstein, 2005)). In *Parhyale* limbs, an antibody marking *repo*-expressing glial cells could include glia associated either with the sensory organs and neurons, and/or with motor neurons. Double staining with other sensory organ markers should allow me to distinguish those two types.

a

```
PxPros      PMHLRKAKLMFFWVRYPSSAVLKMYFPDIKFNKNNTAQLVKWFSNFREFYYIQMEKYARQ   60
PhPros      PMHLRKAKLMFFYVRYPSSSVLKTFFPDIKFNKNNTAQLVKWFSNFREFYYIQMEKHARQ   791
            ************:******:*** :*******************************:***

PxPros      AISEGLKAADDLHVAGDSELYRVLNLHYNRNNHIEVPPNFRYVVEQTLREFFRAIQGGKD   120
PhPros      ALSEGVKNPDDLVVTMDSELIRVLNLHYNRNNHLEIPSSFRFVTEQTLKEFFKSIQDGKD   851
            *:***:*  *** *: **** ************:*:* .**:*.****:***::**.***

PxPros      TEQSWKKSIYKVISRLDDPVPEYFKSPNFLEQLE    154
PhPros      SESSWKKAIYKVIARLDETVPDYFKSPTFLEQLE    885
            :*.****:*****:***: **:*****.******
```

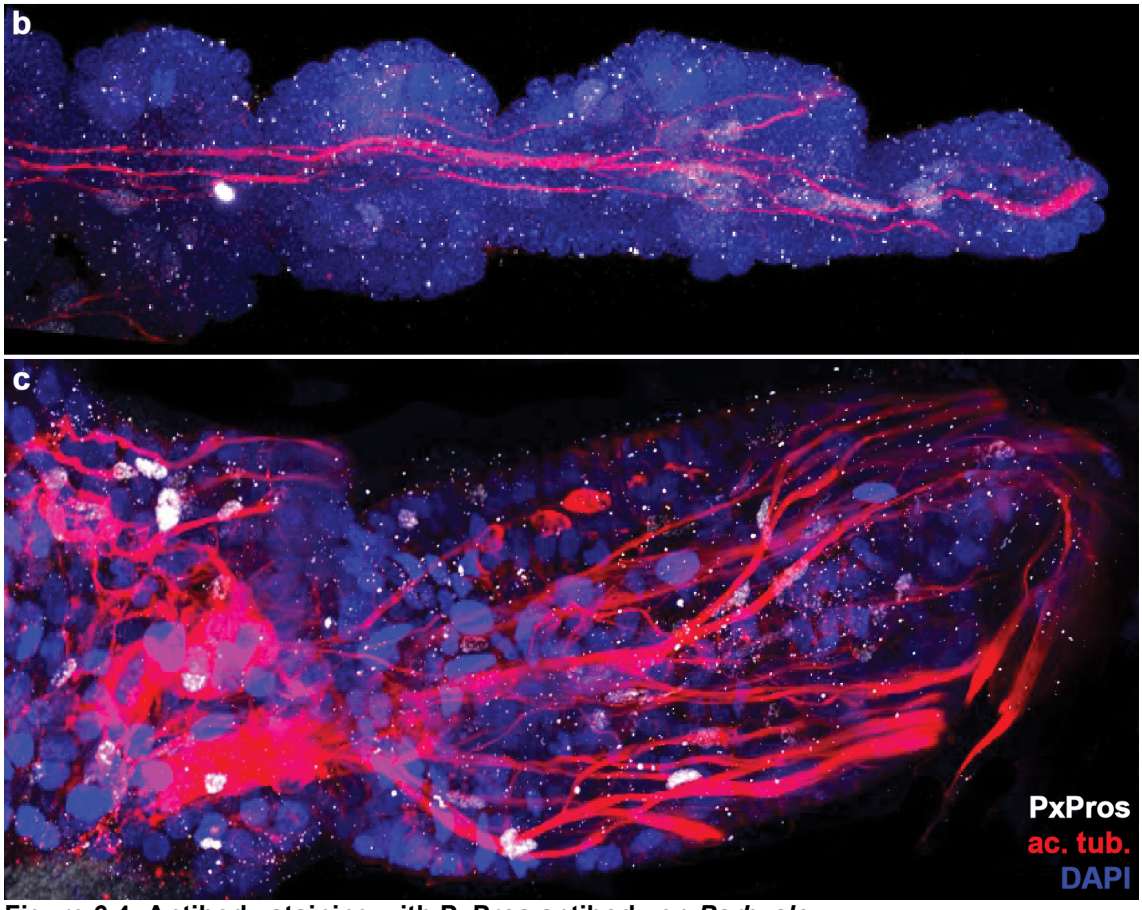


Figure 6.4: Antibody staining with PxPros antibody on *Parhyale*.
(a) Alignment of *Papilio xuthus* and *Parhyale hawaiensis* Pros sequences, in the C-terminal region that was used to raise the PxPros antibody, showing their high degree of sequence conservation. IHF with PxPros antibody on (a) stage S22 embryonic and (b) regenerating adult legs of *Parhyale* labels several nuclei, some of which are associated with axons labeled with an antibody for acetylated tubulin.

In addition to these markers, I also chose to raise antibodies against Mys-2, a *Parhyale* orthologue of the Myospheroid (Mys) protein. In *Drosophila*, *mys* is expressed at high levels in tendon cells (Jannuzi et al., 2002; Subramanian et al., 2007). In *Parhyale*, mys-2 is enriched in a single cluster, which could represent tendon cells.

Arthropod tendons (muscle attachment cells) derive from the epidermis. In the context of my project, it would be interesting to investigate whether different epidermal derivatives such as sensory organs and tendons derive from common epidermal precursors.

Protein expression and immunization

To express and purify the selected target proteins, I used a bacterial expression system that relies on the *trp* operon (Koerner et al., 1991). The coding sequence of the target protein was cloned into the plasmid pATH11, in frame with a *TrpE* coding sequence and downstream of the regulatory region of the *trp* operon. There are two important features of this system: first, under the *trp* regulatory region, the transcription of the TrpE fusion protein can be induced in the absence of tryptophan in the culture medium, and higher levels of transcription can be induced by treating the bacteria with a tryptophan analog, indoleacrylic acid (IAA). Second, the fusion with TrpE directs the expressed fusion protein into inclusion bodies, which are insoluble and can be efficiently pelleted by centrifugation once the cells are lysed. Thus, the protein of interest can be recovered easily and in large quantities.

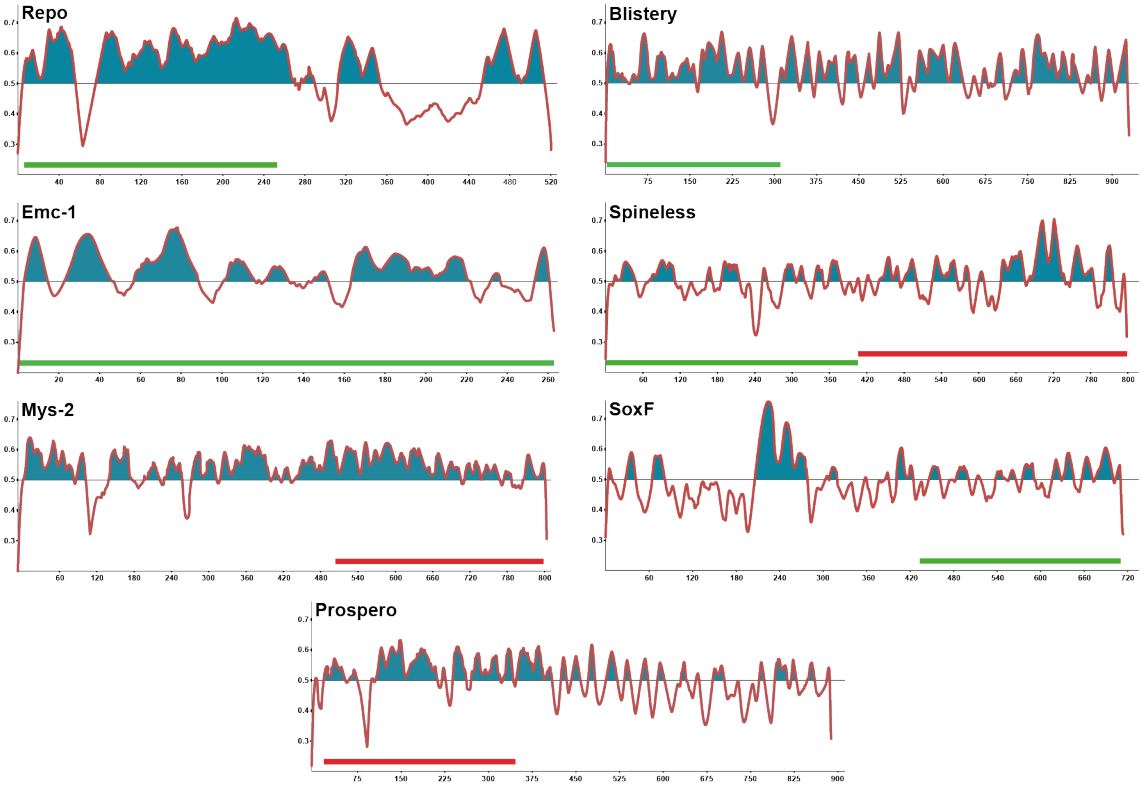


Figure 6.5: Predicted antigenicity profiles of the putative marker proteins.
The antigenicity score (y-axis) over the entire length of the target protein sequences (x-axis) is plotted based on the BepiPred2.0 algorithm. The regions with higher antigenicity than the default threshold are marked with blue. The green bar labels the successfully produced peptides for antibody generation, and the red bars mark the regions that failed to be expressed in *E.coli*.

In my initial trials, I did not succeed in inducing the expression of fusion proteins larger than 75 kD (TrpE is itself 39 kD). Therefore, the size my target peptides should be no larger than 35-40 kD.

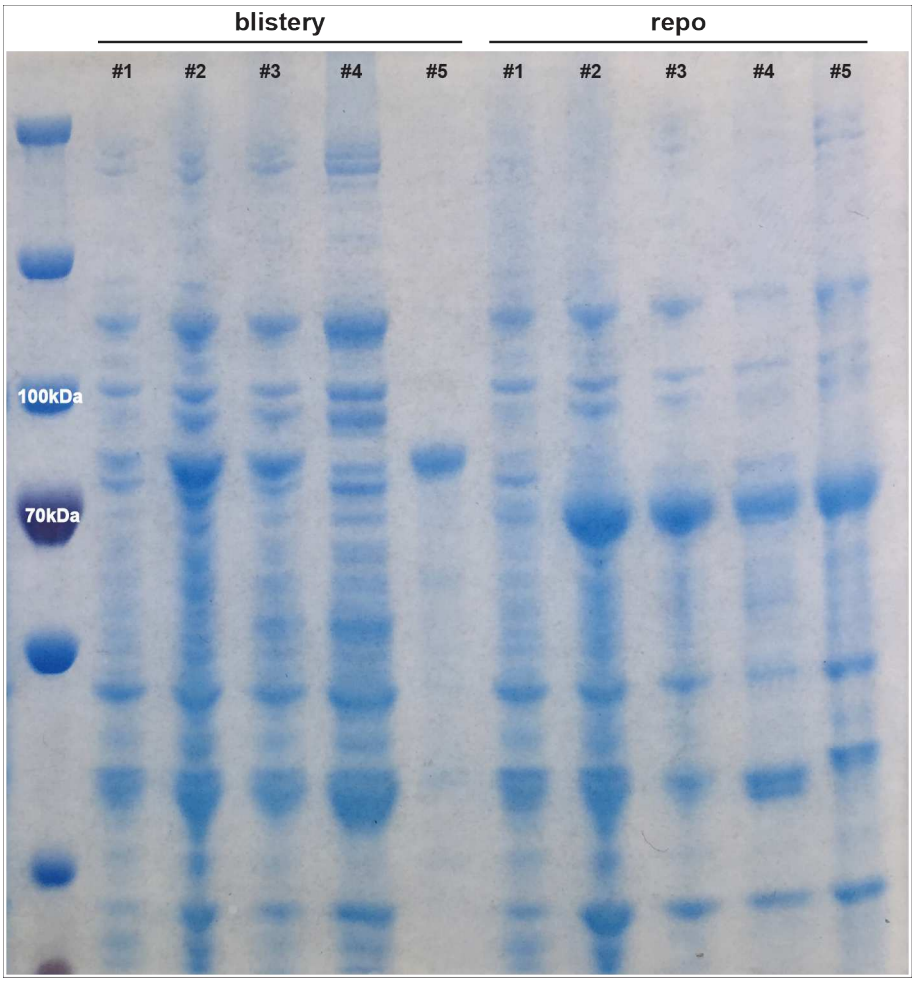


Figure 6.6: Protein samples visualized on an SDS-PAGE at different stages of production and purification.
Protein extracts from bacteria expressing Blistery and Repo, before (lane #1) and 4 hours after (lane #2) induction; cell lysate before ultracentrifugation (lane #3), the supernatant (lane #4), and the pellet dissolved in 8M urea (lane #5). In all lanes, the extracts were mixed with SDS loading buffer and separated by SDS-PAGE. The gel is stained with Coomassie blue to visualize the proteins. The Blistery-TrpE fusion protein (~72kDa) is highly expressed after induction and purified well from the rest of the bacterial proteins via centrifugation. The Repo-TrpE fusion protein (~65 kDa) is also highly expressed upon induction but the protein is recovered at much lower purity.

Since most of my target proteins were larger than the desired size, I decided to clone only a part of their coding sequence. In order to identify the most antigenic parts of these proteins, I used BepiPred-2.0, a machine learning-based algorithm trained with protein fragments of known antigenicity (Jespersen et al., 2017). I used this tool to determine potentially antigenic regions in my target proteins, which I cloned in expression vectors (indicated by red and green bars in Figure 6.5). For Emc-1 and Ss, I cloned the entire coding sequence, either in a single piece (Emc-1) or in two

fragments (Ss). I cloned the coding sequences of the selected fragments into the pATH11 plasmid, in frame with a *TrpE* coding sequence.

I transformed each expression plasmid individually into the *E.coli* TOP10 strain, inoculated liquid cultures with single transformants, induced protein production with IAA, harvested the bacteria, and purified the inclusion bodies. I tested three different protocols to produce and purify proteins (Hoey, 2001), (Kevin Jones and Kevin Moses, personal communication), but none of them worked very reliably in my hands. I tried to optimize various steps and generated a new protocol that is a combination of the original three. The efficiency of expression and purification were variable, depending on the protein being expressed (as illustrated in figure 6.6 for Blistery and Repo fragments, which were processed in parallel). Among the eight target fragments, I purified five. I did not succeed in producing the target fragments Mys-2, Pros, and the C terminus half of Ss (red bars in figure 6.5), due to low expression levels in *E. coli* (undetectable in SDS-PAGE gels). I continued with the antibody production with the five antigens I purified.

Immunizations were carried out by David's Biotechnologie GmbH, with each of the above-mentioned protein preparations, shipped in 6M urea. Mice, rabbit, and chicken were immunized, to generate antibodies that could be detected with different secondary antibodies, for double IHF stainings (see table 6.2 in Materials and methods). For mice and rabbits, a 63 days protocol with 5 rounds of immunization was used. The serum response of the animals was measured with ELISA on day 35, and on day 56 a small amount of serum was harvested and tested by IHF (see below). Due to the CoviD19 restrictions, the IHF tests took longer than expected and some of the animals were kept beyond 63 days. Final bleeds were collected on days 63 to 140. For chickens, immunizations were carried out over 36 days but with 4 rounds of injection. Antibodies were purified from 10 eggs per chicken, collected on 46 to 63 days from the first injection.

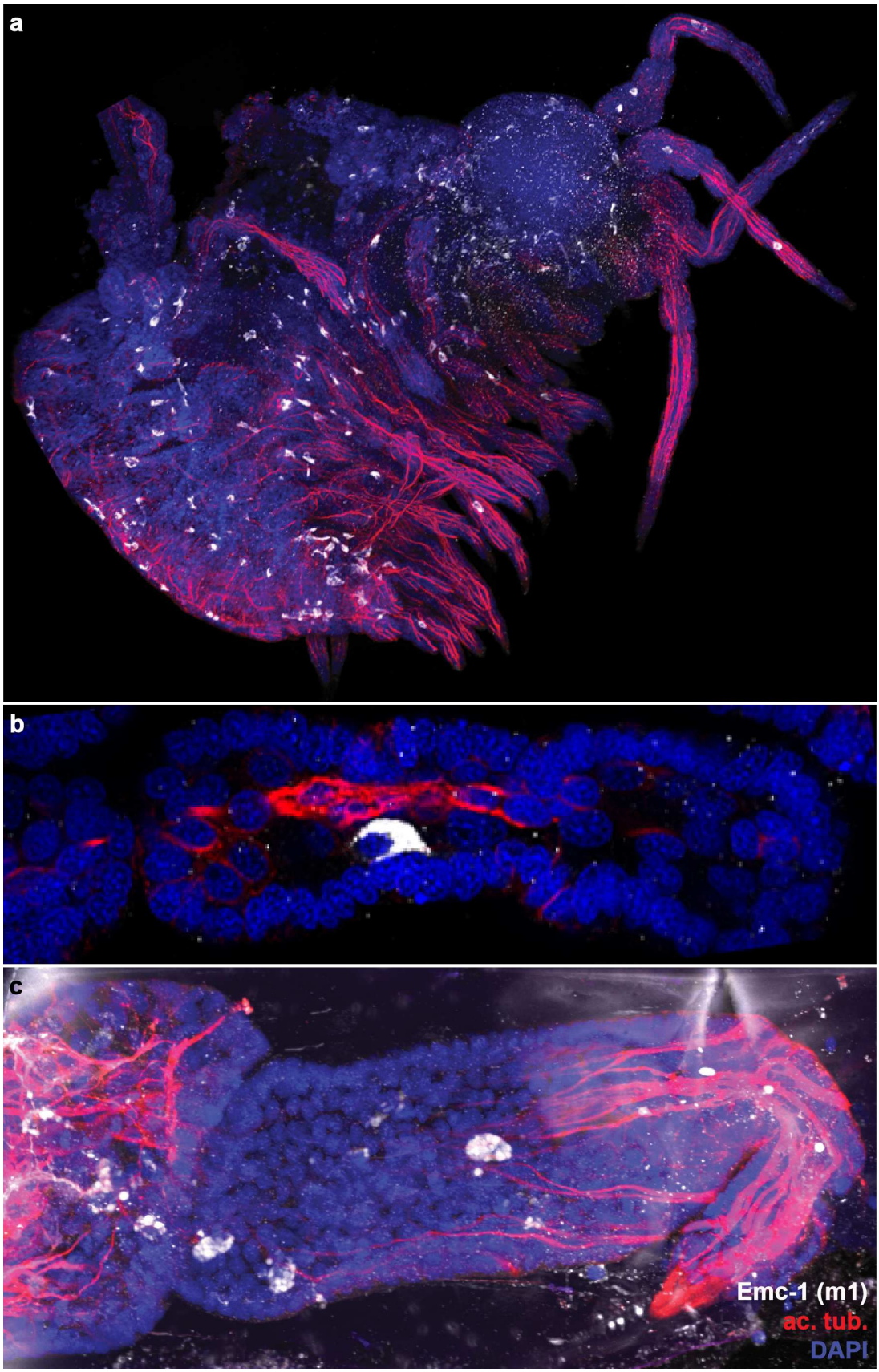


Figure 6.7: Antibody staining with an Emc-1 antiserum
(a) IHF with the Emc-1 antiserum from mouse #1 on a stage S25 embryo, labeling dispersed cells, including some on the limbs. (b) Close-up view of an embryonic limb shows the Emc-1 antiserum is labelling the cytoplasm. (c) IHF on regenerating legs, where we find several cells labeled, but none that appear to be associated with the epidermis or axons.

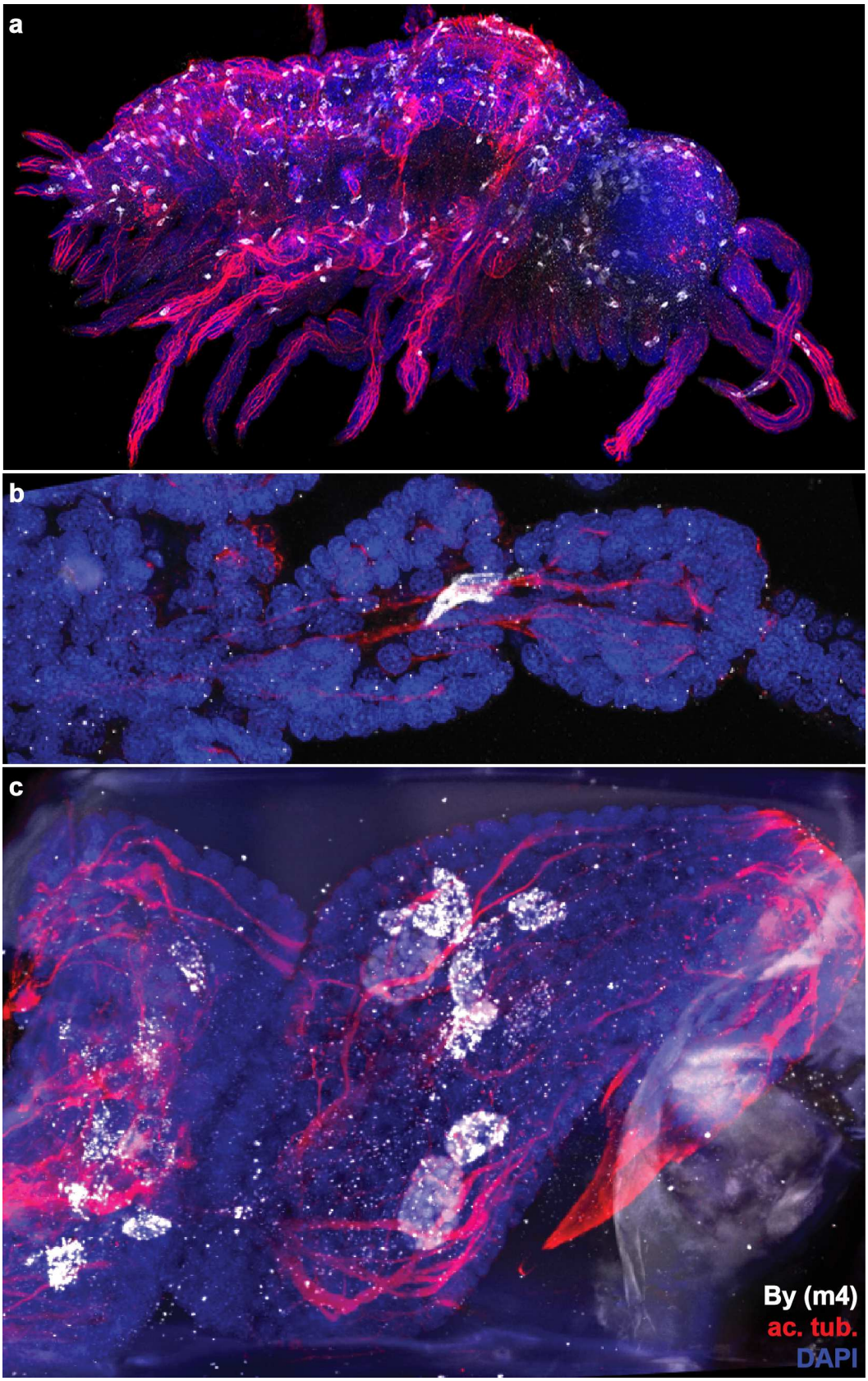


Figure 6.8: Antibody staining with a By antiserum
(a) IHF with the By antiserum from mouse #4 on a stage S25 embryo, labeling dispersed cells, including in the head, the ventral nerve cord, and the limbs. (b) Close-up view of an embryonic leg in which the cytoplasm of a cell closely located to axons is stained. (c) IHF on regenerating legs, where several large cells are labeled; none of these cells appear to be associated with the epidermis or axons.

Testing the antibodies raised against *Parhyale* proteins by IHF

IHF stainings on late embryos using the sera raised against Emc-1, label dispersed cells that are broadly distributed in the body, including a few cells in the antennae and legs (figure 6.7a-b); the localization seems to be cytoplasmic. All of the mouse sera raised against Emc-1 gave similar staining; the chicken antibodies raised against Emc-1 did not give a staining. IHF on regenerating limbs, using any of the Emc-1 sera labeled some cells that appeared to have granular structures in their cytoplasm (figure 6.7c). The number and the distribution of these granular cells varies from sample to sample, but in general they are located in the deeper sections of the leg and do not appear to be associated with neurons.

IHF stainings in late embryos using the sera raised against By also stain the cytoplasm of dispersed cells (Figure 6.8a-b). On regenerating legs, these antisera label some large cells located internally in the leg, away from the epidermis, and the staining appears in granular structures in the cytoplasm of these cells. (Figure 6.8c). The localization of stained cells, as well as their numbers, varies between samples. In some samples, there are cells labeled outside of the regenerating limb, between the cuticle and the regenerating epidermis, suggesting that these cells could represent macrophages. All of the four mice antisera gave similar staining patterns, but neither rabbit nor chicken antibodies showed any staining.

Repo antisera label several cells within the brain and the ventral nerve cord of late embryos, as well as multiple cells labeled in each limb (figure 6.9a-b). In both embryos and regenerating legs (figure 6.9c), the staining is exclusively nuclear; in both cases the labeled nuclei are associated with the axons staining with an antibody for acetylated tubulin. All 4 mouse antisera resulted in similar staining in embryos and adult limbs. Based on these results, I believe that Repo antisera is labelling glial cell nuclei. I have co-stained a regenerated leg with Repo and PxPros antibodies, along with acetylated tubulin to see if there are labelled cells associated with each other and/or with the same axons. The stainings are non-overlapping, and the labelled cells do not appear to be in physical contact. Yet, there are multiple axons, where both repo and Pros positive nuclei are associated with it.The chicken and rabbit sera raised against Repo did not label any cells.

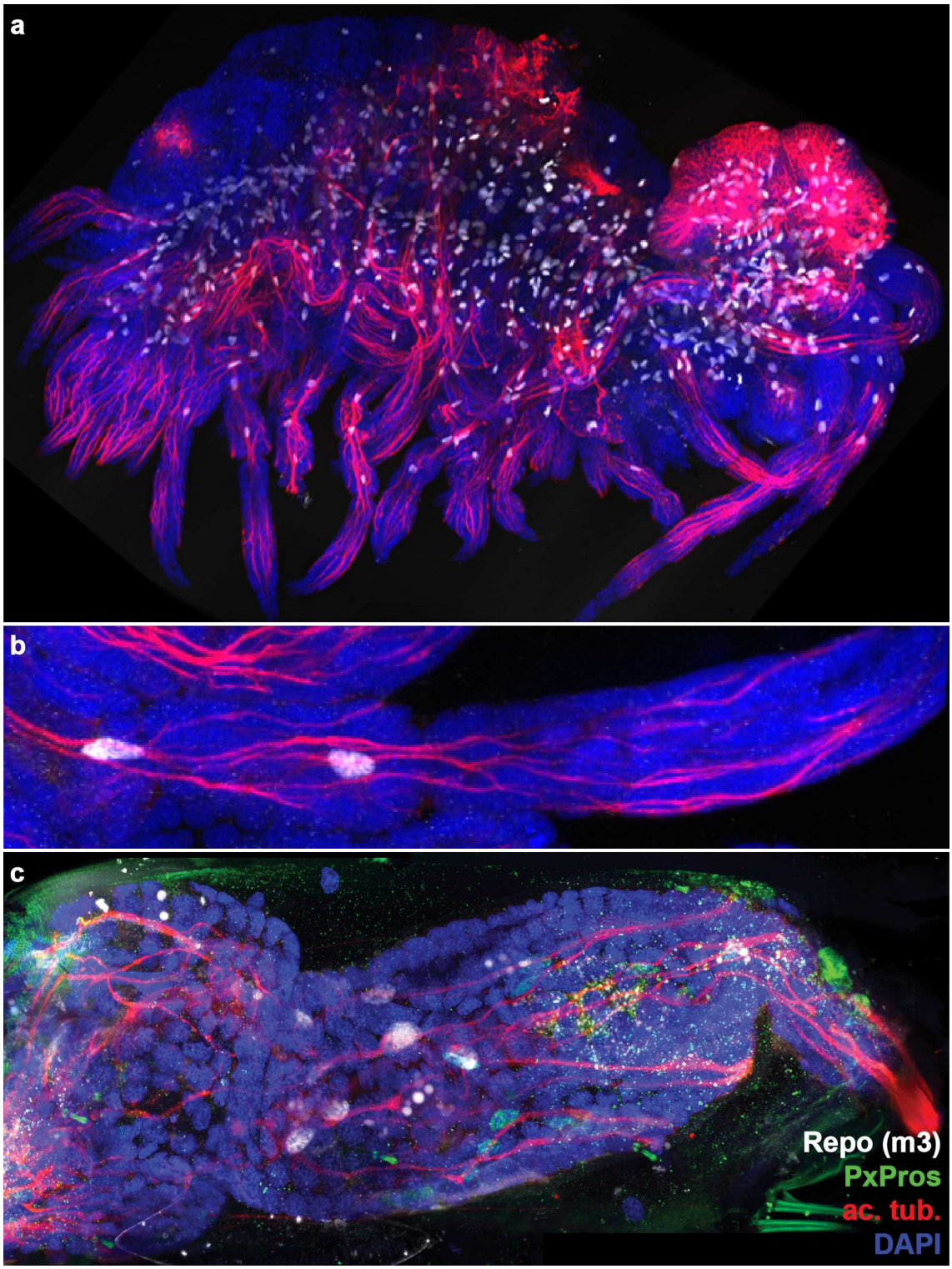


Figure 6.9: Antibody stainings with a Repo antiserum
(a) IHF with Repo antiserum from mouse #3 on a stage S25 embryo, labeling dispersed nuclei, including in the head, the ventral nerve cord, and the limbs, in close association with axons. (b) Magnified view of an embryonic leg clearly shows two Repo stained nuclei that are closely associated with axons, (c) On regenerating legs, labelled nuclei are not close to the epidermis but they are always associated with axons. Repo-positive and Pros-positive nuclei are not closely associated with each other but they are sometimes associated with the same axons.

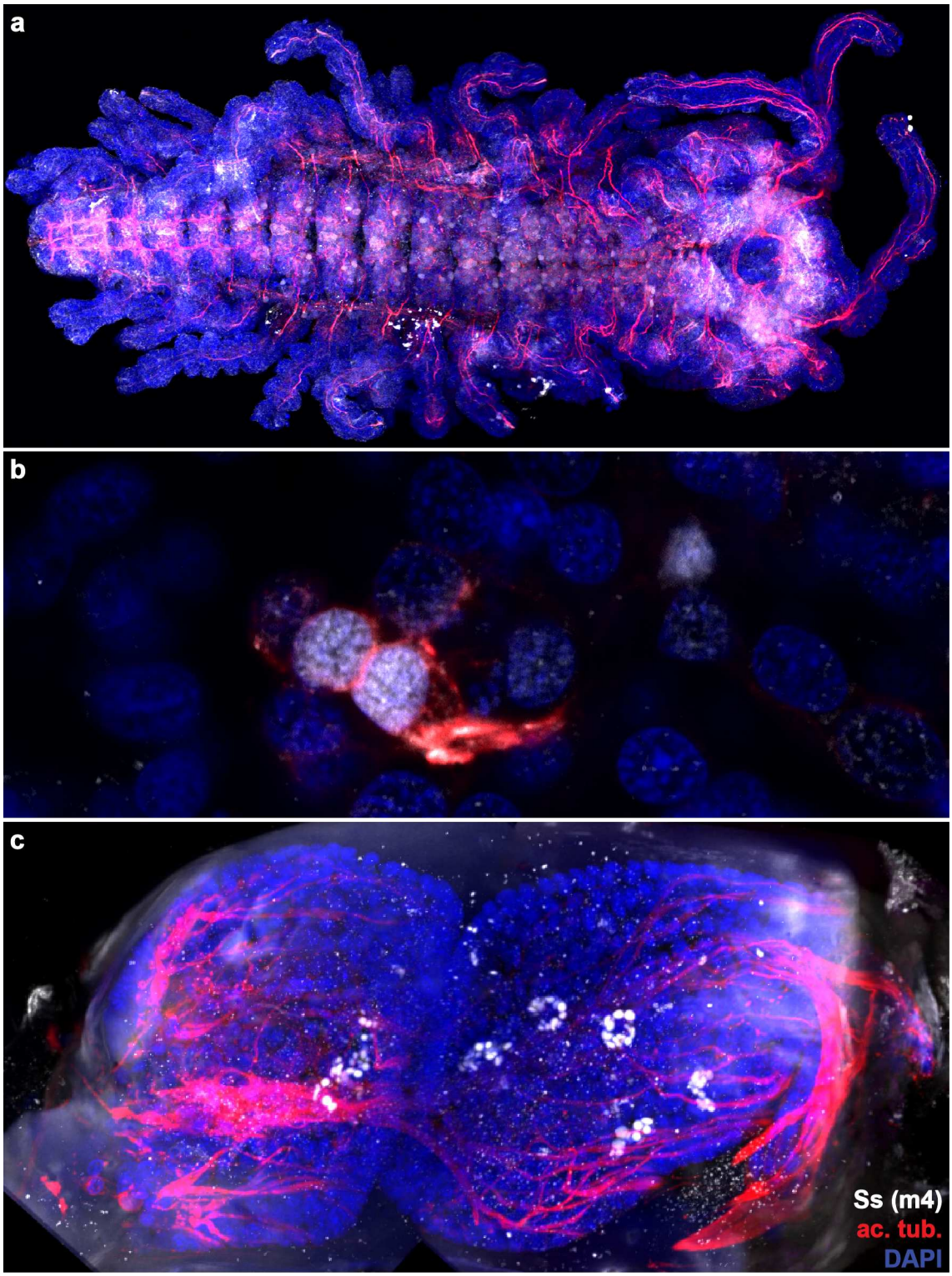


Figure 6.10: Antibody stainings with Ss antisera.
(a) IHF with an antiserum raised against Ss (mouse #4) on a stage S22 embryo, labeling several nuclei in the head and the ventral nerve cord. (b) A close-up view of the proximal part of an embryonic leg showing two stained nuclei associated with developing axons. (c) On regenerating legs several cells are stained, which are neither close to the epidermis nor associated with axons.

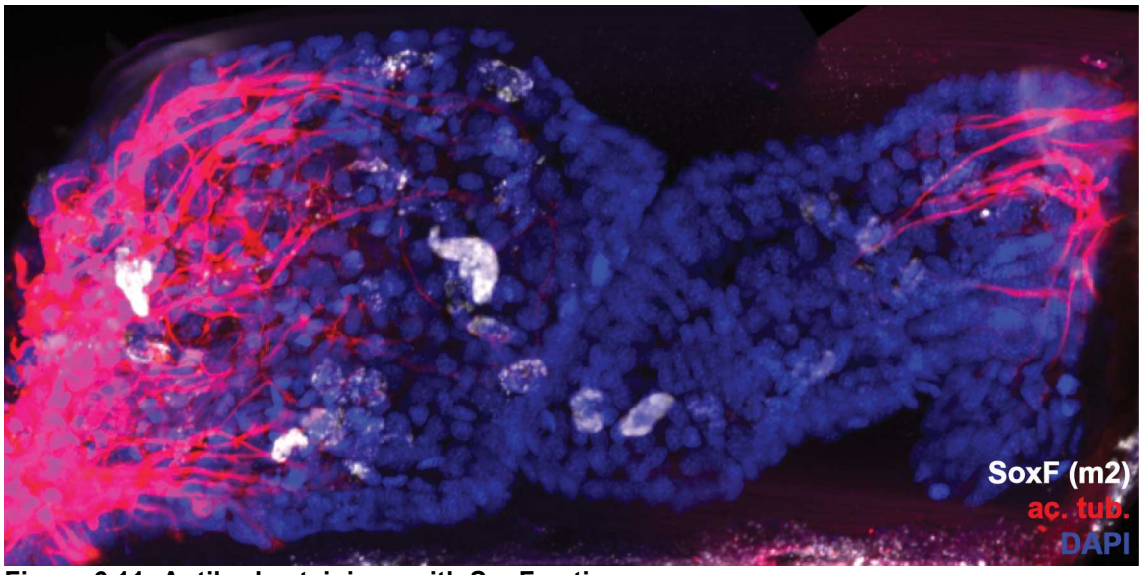


Figure 6.11: Antibody stainings with SoxF antisera.
IHF with an SoxF antiserum (mouse #2) on a regenerating leg labels several cells in the cytoplasm. The stained cells are generally located in the deeper sections of the leg and do not seem to be associated with the axons, labeled with an acetylated tubulin antibody, or with the epidermis.

In late embryos, most of the cells labeled with Ss antisera appear to be associated with the central nervous system (the VNS and the brain), and they are stained in their nuclei. There are also labelled cells in the proximal parts of the limbs (figure 6.10a-b) but there is also a strong background staining which makes it challenging to identify the labelled cells. Optimising the stainings to reduce background signal, as well as staining later embryos would give a more clear picture of Ss antisera staining pattern in the limbs. All the antisera raised in mice give similar stainings, except one (mouse #3), which labels axons with a pattern similar to that of the acetylated tubulin stainings. The stainings with antibodies raised in chicken did not work on embryos.

In regenerating legs, SoxF antisera label large cells in the interior sections of the legs, similar to those seen in stainings with Emc-1 and By antisera (Figure 6.10c). The number and distribution of the cells in these stainings varies from one limb to the other. As with the other antibodies, the rabbit antisera did not give a signal. I have not yet tried to stain embryos with SoxF antisera.

Conclusions

In this study, I first developed an antibody staining protocol for regenerating *Parhyale* legs. The protocol is robust and works reliably in my hands, with almost 100% efficiency. The morphology of the limb is conserved very well during the staining procedure, so that it is possible to recognise individual nuclei after antibody staining, and to match them with nuclei in the live-imaged regenerating limb. Thus, antibody stainings can be used to identify different cell types in a regenerating limb and to track their progenitors. Of note, the H2B::mRFPRuby fluorescence in the limbs is undetectable after the antibody staining protocol. This does not pose any problem for my experiments since nuclei can be easily and efficiently labeled with DAPI.

I used TrpE fusions to express and purify proteins in bacteria, because this technique is supposed to be quick and straightforward. However, the protocol did not work reliably in my hands: protein production was more difficult and took longer than expected, and the purity of the antigen varied. Due to these problems, I did not manage to produce three out of eight target proteins. In some cases (Pros and Mys-2 protein fragments) this was due to inefficient protein expression in bacteria (possibly because these proteins are toxic in bacteria), in other cases (Repo protein fragment) this was due to inefficient purification in inclusion bodies. For future experiments, it would be useful to test other approaches for protein production and purification, as well as other parts of the target proteins.

I aimed to generate antibodies to stain the accessory cells of *Parhyale* sensilla. Antibodies recognising By, Ss, Emc-1, and SoxF proteins are expected to stain the outer accessory cells, therefore, I expected to see labelling of the cells embedded or closely associated with the epidermal layer. However, the IHF on regenerating legs with antisera raised against these proteins labels the cytoplasm of some large cells located in the deeper sections of the leg, away from the epidermis. Occasionally, in late stages of regeneration, labelled cells are found in the space between the cuticle and the regenerating leg tissue. The number and the positions of these cells are not fixed, but change from one leg to the other. Based on these observations, I suggest that the stained cells do not belong to the sensilla but are likely to represent circulating cells like macrophages.

The staining patterns on embryos are hard to predict since the target genes might have multiple roles during embryogenesis. Both By and Emc-1 antisera give cytoplasmic staining in cells that are dispersed throughout the body of late embryos. It is not clear whether some of those cells are related to developing sensilla or not. It is

possible that both antisera are staining the same cells in embryos and regenerating legs.

Ss antisera staining is mainly detected in the nuclei of cells associated with the central nervous system; at least some of the stained nuclei are associated with developing neurons. But I could not detect any nuclear labeling when regenerated legs are stained. Overall, these staining patterns observed by IHF using the antibodies I raised against By, Emc-1, Ss, and SoxF are unexpected and it is hard to draw any solid conclusions before further testing and optimisation to eliminate non-specific staining (by adjusting the dilutions of these sera, affinity purification, or preadsorbtion).

The sera are expected to contain many antibodies besides the ones raised against the *Parhyale* target proteins, including antibodies already present prior to immunization, antibodies developed against the TrpE protein, and also against bacterial proteins that were co-purified and injected along with the desired antigen. The unspecific staining could be coming from either of these antibodies, if they cross-react with unknown antigens in the tissues of *Parhyale*. There are two ways to reduce this unspecific staining: purification of the desired antibodies from the serum by affinity purification, or depleting the unwanted antibodies in the serum via preadsorbtion with the target tissues. To extract specific antibodies by affinity purification, the target antigen is used to bind and purify the antibodies. To preadsorb unwanted antibodies, the serum can be incubated either with *Parhyale* tissues, with protein extracts from *Parhyale* tissues (usually extracted using acetone), or with protein preparations containing likely contaminants (e.g. trpE or bacterial proteins).

None of the antibody preparations obtained from rabbits or chicken gave a specific cell staining in my initial IHF trials. Affinity purification may help to purify and concentrate any specific antibodies present in those sera.

In summary, I have a few antibodies to work with in *Parhyale*. The repo antisera I produced, work well on regenerating legs and seem to be labelling glial cells. The PxPros antibody is labelling some cells associated with axons and the epidermis, which suggest some of the labelled cells could be the sheath cells. Also, antibodies for acetylated tubulin can be used to visualise axons (although some axons are associated with several nuclei, which makes it difficult to identify the nuclei of the neurons). Moreover, in a recent study, commercially available ChAT and VaChT antibodies raised against human antigens were shown to label sensory neurons in locust *Schistocerca gregaria* (Ehrhardt and Boyan, 2020). There is a good chance that these antibodies would be useful to label the sensory neurons in *Parhyale*. The next step will be to co-stain live imaged regenerating legs with these antibodies, to trace the labelled

cells during regeneration and identify the clonal and developmental relationships between them.

Overall, combining live recordings with antibody stainings is a promising new approach for tracing the progenitors of specific cell types during *Parhyale* leg regeneration. The main difference of this method, compared with the transgenic approaches described in chapters 4 and 5, is that we will miss information of the initial cell type of the progenitor cells. Therefore, it cannot answer directly whether trans- or de-differentiation is involved during the regeneration process.

7. Conclusion and perspectives

A major quest in the study of regeneration is to identify the progenitors of the regenerated tissues, and to describe their properties and developmental potential. In this thesis I focused on the limb regeneration of *Parhyale* hawaiensis, aiming to develop a methodology for tracing back the progenitors of terminally differentiated cell types. As a pilot to this approach, I concentrated in the cells that constitute the sensory organs. In this chapter I would like to discuss my results and observations in a broader context, considering the findings from studying regeneration in other animals, and to discuss the future of *Parhyale* regeneration studies.

High fidelity regeneration of *Parhyale* peripheral sensory organs

I demonstrated that the regenerated limbs are near-perfect replicas of uninjured limbs in terms of the setae they bear on their surface. In the same line with my results, single-nucleus RNA sequencing studies revealed all cell types of *Parhyale* limbs are recovered during regeneration (Alba Almazán, unpublished). The morphological and physiological recovery of the sensory organs during limb regeneration raises questions about the nature of their progenitors, which underpin the rest of my project. These results also have implications beyond the tracing of progenitor cells, related to the fidelity of regeneration.

Previous reports comparing regenerated organs with their uninjured counterparts tend to focus on their differences, *e.g.* regenerated lizard tails lacking bone tissue (Goss, 1969), or regenerated axolotl legs bearing anomalies in their muscles (Diogo et al., 2014). Single-cell RNA sequencing experiments have revealed that the cellular diversity is recovered during Zebrafish caudal fin regeneration (Hou et al., 2020), yet differences in bony ray bifurcation and pigmentation are reported in the regenerated fins (Azevedo et al., 2012), most likely resulting from patterning errors/differences. Unlike these examples, *Parhyale* regeneration is capable of restoring both the diversity and the pattern of cells/tissues generated during embryogenesis. To my knowledge, *Parhyale* limbs are the first example where a high fidelity regeneration is documented among the species that do not have the capacity for whole-body regeneration. Future analysis of other *Parhyale* cell types and tissues will clarify this point further.

From a different viewpoint, the recovery of sensory organs has some exciting implications for sensory neuron development. The functional recovery of at least one type of sensilla indicates that the axons of newly born sensory neurons have reached the ventral nerve cord and made the relevant synaptic connections. Interestingly, this

can happen (in the fastest cases) within 3-4 days from limb amputation. It would be interesting to investigate how this rapid axonal growth and synaptogenesis take place. In the crayfish *Procambarus clarkii*, severed axons of sensory neurons live on when the limbs are amputated, and could potentially serve as guidance cues for the new sensory axons (Cooper, 1998). This could also be happening in *Parhyale*. Alternatively, the axons of newly born neurons may fuse with the severed axons of old sensory neurons (axonal fusion is reported in some invertebrates (Neumann et al., 2019)). With the existing and developing tools and techniques, *Parhyale* could serve as a good model to study underlying mechanisms of axon pathfinding during regeneration.

Live recording the entire course of *Parhyale* limb regeneration

Several methods are available for lineage tracking in different organisms; direct observation via imaging is one of the most accurate methods (reviewed in (Garcia-Marques et al., 2020)). Live imaging and cell tracking during regeneration is particularly challenging, due to the long duration of this process, difficulties to immobilise animals during long periods, and the opaqueness of tissues. Relatively rapid progression of regeneration in *Parhyale*, together with the small size and transparency of its tissues, which are encased in a rigid exoskeleton, have made it possible to develop a live imaging technique for *Parhyale* limbs (Alwes et al., 2016). I improved this technique to record the entire course of regeneration, from amputation to completion, with a resolution that allows us to track individual cells. Live tissue monitoring approaches have been developed for axolotl and zebrafish appendages as well (Currie et al., 2016; Tornini et al., 2016); these utilise sparsely labeled cells as landmarks for registering images acquired of the regenerating tissue over long time intervals. The low complexity of cell labels and the limited temporal resolution of this approach constrain the ability to precisely track individual cells through successive cell divisions. Hence, our ability to perform direct cell tracking in *Parhyale* is unparalleled.. The current method allows us to visualize and track the cells in the posterior half of the limb. Currently, one animal can be imaged at a time. In the future, we intend to adapt this approach to allow imaging multiple animals simultaneously and/or imaging deeper tissues. My live recording protocol can serve as a good reference for future improvements in live imaging techniques, especially if the ultimate goal is cell tracking.

Transgenesis in *Parhyale*

My attempts at generating cell-type-specific transgenic markers, via CRISPR-mediated knock-ins and *cis*-regulatory element (CRE) reporters, were unfruitful. Establishing these transgenic approaches would be very helpful for future work on *Parhyale*

regeneration because, beyond visualising different cell types, being able to target different cell/tissue types would make it possible to manipulate them. There are additional *Parhyale* genomic resources available today, compared to the time when I started these projects: we have an improved genome assembly, more extensive transcriptomic datasets from embryonic and adult limbs, single nucleus RNAseq and ATACseq from adult limbs. These resources combined with the experience from my work, would help future transgenesis approaches.

The single nucleus RNAseq datasets are invaluable for identifying marker genes for different cell types (see chapter 6 for details). The improved genome assembly and the bulk limb transcriptomes would be useful to identify the most highly expressed genes and the least polymorphic regions for CRISPR targeting. Along with these improvements, I believe that a more systematic study for CRISPR-mediated knock-ins for *Parhyale* is required. The two previous reports of CRISPR knock-ins in *Parhyale* were not systematic efforts to test and quantify CRISPR approaches in this species, but isolated cases where knock-ins were reported. Rather than taking these reports as the basis for future CRISPR knock-ins, it could be beneficial to take a more systematic approach, *e.g.* to test different types of repair constructs like short/long single-stranded DNA molecules (Quadros et al., 2017) or double-stranded linear DNA molecules with homology arms (Yao et al., 2018), and use drugs that block the NHEJ repair pathway to enhance knock-in rates (Maruyama et al., 2015). These experiments can be time consuming, but I believe in the long term this would be a valuable investment.

For CRE-reporters, one obvious way to continue is to clone and test the activity of the genomic regions highlighted in the ATACseq datasets. In addition to the ATACseq data from embryonic limbs, we now have single-nucleus ATACseq datasets from adult *Parhyale* limbs, which can be used to select regions that are potentially active in a cell-type-specific manner. Yet, even with these prediction tools, the identification of useful reporter constructs is likely to be laborious. Therefore, more efficient methods are required for both cloning and testing these reporter constructs. DNA synthesis technologies are becoming more cost-efficient and may provide an avenue for generating the constructs. To test the constructs can be injected as pools rather than individually. If large numbers of constructs are tested, advanced sample pooling methods like the hypercube placement used for Covid19 testing (Mutesa et al., 2020), can be used.

Combining antibody staining with live imaging to identify progenitors in regenerating limbs

Combining antibody staining with live imaging is a promising approach for identifying the cellular progenitors of specific cell types during limb regeneration. Both live imaging and the antibody staining protocols are robust and, therefore, the method can be applied to many cell types once antibodies are raised to label them. With the techniques and resources at hand, in my experience it will be more rapid and easier to generate antibodies than to generate transgenic animals with cell-type-specific reporter expression. We will have a clearer idea on this point once the generated antibodies are fully tested, so that we know the actual success rate of antibody production.

There are two fundamental differences between tracing the progenitors using transgenic markers or with the antibody staining method. First, the transgenic approach would allow us to observe cases of trans- and de-differentiation, since it allows the identification of the initial cell type of the progenitors; this is not possible with the antibody stainings. Second, the antibody staining will be more useful than the transgenic approaches to identify multipotent progenitors, if there are any, because it allows easy multiplexing of different cell-type markers. For example, in order to reveal if the neuron, glia, and sheath cells can arise from the same progenitors, transgenic animals carrying four distinct transgenes would be required (three cell type-specific markers, and a ubiquitous nuclear reporter for cell tracking); combining four transgenes in a single animal would be very laborious in *Parhyale* (taking several years by simple crosses). On the other hand, it is much easier to combine antibody stainings to identify different cell fates in a regenerating limb.

Once cell-type specific transgenic animals are generated, combining these two methods could be very powerful. For example, transgenic animals with glia-specific reporters could be used to identify whether glial cells are contributing to the regenerating limbs as progenitors, and if so, the antibody staining could be used to unveil the fates of their progeny.

Studying cell and tissue dynamics during *Parhyale* limb regeneration

The live recordings also provide useful datasets to study cellular dynamics, such as cell division, migration and tissue rearrangements during limb regeneration. Since the datasets cover the entire course of regeneration and the field over which regeneration is occuring, with good spatiotemporal resolution, these datasets give a good overview of the cellular dynamics underlying regeneration. My colleague Severine Urdy is comparing the cellular behaviors of limb development during regeneration and embryogenesis, taking advantage of the datasets I generated and the embryonic recording generated by (Wolff et al., 2018). A similar comparison of the progenitors of

limb cells can be made by applying the cell fate identification and tracking techniques to the embryos.

Using the semi-automated cell tracking tool generated by my colleague Ko Sugawara, we can currently track a large number of lineages within 2-4 weeks, which is considerably shorter than manual tracking.. This means, in a near future, we can have several regenerating specimens where the majority of cell lineages have been tracked. This will open new opportunities for statistical analysis and classification of cells based on their behaviors during regeneration, such as their division patterns and orientations, and movements.

APPENDICES

A - Material and Methods

Parhyale husbandry

Parhyale stocks listed in table A.1 are kept for many generations in the laboratory and raised as described previously (Browne et al., 2005).

Table A.1: List of *Parhyale* strains used

Strain name	Details	Reference
Chicago-F	Isofemale wild type *Paryale* line	{Parchem:2010ik}
PhHS-H2B::mRFPRuby	Transgenic line expressing a GFP-histone fusion under a heat-shock promoter	{Wolff:2018et}
DC5-H2B::eGFP	Transgenic line expressing a GFP-histone fusion in the CNS (including motor neurons)	{Konstantinides:2016wl}
PhHS-lyn::tdTomato-2A-H2B::EGFP	Transgenic line expressing tdTomato with the lyn tag (associated with cortical actin) and a GFP-histone fusion under a heat-shock promoter	{Alwes:2016ff}

Animal preparation for live imaging

Prior to imaging, animals were kept 1 to 4 weeks in the dark to prevent the build-up of algae on the cuticle. Animals were initially heat-shocked for 45 minutes at 37°C 16-24 hours prior to mounting and imaging. The next day, animals were anesthetized using 0.02% clove oil and glued on 35mm glass-bottom Petri dishes (Ibidi, 81158) using surgical glue (2-octyl cyanoacrylate, Dermabond) as described by (Alwes et al., 2016). Once the glued animals woke up from anesthesia, glued T4 and T5 limbs were amputated at the distal end of the carpus podomere using a microsurgical knife (Premier Edge PE3045, 45° stab knife).

Antibody staining

Antibody stainings on embryos and regenerating limbs are performed as described in detailed protocols. The antibodies used are listed in table A.2.

Table A.2: List of antibodies used

	Reference no	Host / isotype	Clone	RRID	Dilution
α-acetylated tubulin	Sigma-Aldrich, T6793	Mouse IgG monoclonal	6-11B-1	AB_477585	1:1000
α-acetylated tubulin	Thermo Fisher Scientific, MA5-33079	rabbit IgG monoclonal	RM318	AB_2810171	1:1000
α-tyrosinated tubulin	Abcam, ab6160	rat IgG monoclonal	YL1/2	AB_305328	1:1000
α-Lamin Dm0	DSHB, adl67.10	mouse IgG monoclonal	-	AB_528336	1:100
α-Pax3/7	{Davis:2005eg}	mouse IgG monoclonal	DP311	-	1:50
α-Glutamine synthetase	BD Biosciences, 610517	mouse IgG monoclonal	-	AB_397879	1:200
α-PxPros	{Perry:2016bv}	rat IgG polyclonal	-	-	1:100
α-mouse IgG-Alexa 488	Life Technologies, A11001	goat polyclonal	-	AB_2534069	1:1000
α-mouse IgG-Alexa 647	Invitrogen, A21235	goat polyclonal	-	AB_2535804	1:1000
α-rat IgG-Alexa 594	abcam, ab150168	goat polyclonal	-	-	1:1000
α-rabbit IgG-Alexa 488	Life Technologies, A11008	goat polyclonal	-	AB_143165	1:1000
α-rabbit IgG-Alexa 555	Life Technologies, A21428	goat polyclonal	-	AB_2535849	1:1000

Microinjections for CRISPR

The protocol for injections can be found in the detailed protocols section. The sgRNAs used are listed in table A.3.

Table A.3: List of sgRNAs used

name	sequence	type
Antp sgRNA	GGGCCTATAGTCAGGGTATG	GG18$_N$-NGG
Pax3/7-2 sgRNA1	GGGGCAAGTATCATAATATT	Gx18$_N$-NGG
Pax3/7-2 sgRNA3	GGCTGGTAACAGTCAGGCTG	Gx18$_N$-NGG
futsch sgRNA1	GGGCCTTTGGGTGGCTAGTA	xG18$_N$-NGG
futsch sgRNA2	GGGACTATGACACTTTCTGC	xG18$_N$-NGG
futsch sgRNA4	GGAAGTCTCGTGGTCCTGT	GG17$_N$-NGG
futsch sgRNA5	GGGTAGATTGTCGTGTAAA	GG17$_N$-NGG
futsch sgRNA6	GGTTGTCGTGTAAAAGGTTG	Gx18$_N$-NGG
futsch sgRNA7	GGGCAGGTCGGAGAACACGG	GG18$_N$-NGG
futsch sgRNA9	GGGCTTTCAGCTGTATGCAC	GG18$_N$-NGG
sens sgRNA3	GGAAGGAGGCATCGCCCCTG	xG18$_N$-NGG
sens sgRNA4	GGGGACGATGGGGAGACGTG	Gx18$_N$-NGG
sens sgRNA5	GGGCCTGACGGCAGCCGCAA	Gx18$_N$-NGG
sens sgRNA6	GGCCTCGATGTGGATCGAA	GG17$_N$-NGG
sens sgRNA7	GGTCGGTGTTGGTGTCCATG	xG18$_N$-NGG
sens sgRNA8	GGATCTGGACAAGTCCACGT	xG18$_N$-NGG
cut sgRNA1	GGGGGTGTGATGAGTGTGG	GG17$_N$-NGG
cut sgRNA2	GGCGTAGGGGCTGGTGCAGG	Gx18$_N$-NGG
cut sgRNA7	GGCAGATCTTGTCGAAAAAA	GG18$_N$-NGG
elav sgRNA3	GGGGGAGGTAGTTCACGATG	GG18$_N$-NGG
elav sgRNA4	GGGCGAGCGTATCTCCTCCT	Gx18$_N$-NGG

Image acquisition and analysis

Antibody stained tissues are imaged with a Zeiss LSM 800 scanning laser confocal microscope. Cuticle preps and embryos stained with RNA *in situ* hybridization are imaged with a Zeiss AxioZoom microscope.

Live imaging of limb regeneration:

The petri dish with a glued animal was placed on an inverted Zeiss LSM 800 laser scanning confocal microscope, equipped with an incubation chamber and a heat-adjustable stage pre-adjusted to 26°C. During the live imaging, the animals were heat-shocked by setting the temperature of the stage to 40°C for two hours, without changing the temperature of the chamber. The live recordings were performed using a dry 20x objective (Zeiss 420650-9901-000, N.A. 0.8), with 0,15 - 0,50% laser power and 650-750 mV detector gain, and with 20-minute time intervals. Once the imaging was completed, if the recording was divided into several files, the datasets were concatenated using ImageJ. The image drift caused by the heat-shock was corrected by registering the individual time points in the datasets using the ImageJ plug-in "Template-Adjustment".

Calculation of CR and SNR:

For the CR and SNR calculations in chapter 2, the same frame from the same limb was imaged successively with different speed and averaging settings. To analyze the images, a binary mask in which the nuclei are labeled with black and the internuclear space is labeled with white, was manually prepared, marking the nuclei and internuclear spaces. The CR and SNR values of the images were calculated by using the ImageJ plugin generated by (Ulman et al., 2017).

ATACseq

Three samples were prepared for ATACseq; one with a whole stage S24 embryo, another with all legs of four-stage S24 embryos, and the last one with a T4 leg of an adult male. For each sample, the protocol (see in detailed protocols) is followed and at the PCR amplification step the embryonic samples were amplified for 9 cycles and the adult sample was amplified for 12 cycles. Then, the libraries were sequenced with Illumina NextSeq500 at IGFL Sequencing Platform. Processing of the raw sequencing reads was performed by Mathilde Paris, as follows. First, the raw reads were mapped to the genome using Bowtie2 software package (Langmead et al., 2018), and the enrichment analysis was performed with MACS2 (Gaspar, 2018) to reveal the peaks. The peaks were visualized with IGV software (Robinson et al., 2011).

Detailed Protocols

Fixation of adult *Parhyale hawaiensis* for SEM

! Prefer animals kept in the dark for ~1 month to avoid algae !

1. Anesthetize the animals in 1:5000 clove oil:FASW (filtered artificial seawater)
2. Fix animals with 1% Glutaraldehyde in FASW for 2h
3. Wash 3X5min with FASW
4. Fix animals with 1% Osmium Tetroxide in FASW for 2h
5. Wash 3X15min with FASW
6. Dehydrate the animals with serial washes of 25%, 50%, 75%, and 90% EtOH in FASW; 30min each
7. Wash 2X30min with absolute EtOH
8. Wash with 90% EtOH in FASW for 30min
9. Store the samples in fresh 90% EtOH in FASW

RNA *in situ* hybridization in *Parhyale hawaiensis* embryos*

*This is a slight modification of a protocol that has been established by Matt Giorgianni, later modified by Danielle Liubicich both from Nipam Patel Lab, and adapted by Anastasios Pavlopoulos from Averof Lab.

Probe Synthesis

1. Amplify the fragment of interest (500-1000bp) from cDNA or gDNA using *Taq* DNA polymerase and clone it into pGEM-T plasmid.
2. Confirm the sequence and orientation by Sanger sequencing.
3. Amplify the template for RNA probe synthesis using M13 forward and reverse primers, and purify with a PCR purification kit.
4. Set the reaction to synthesize RNA probe and incubate @37°C for 2h (up to 4h)

4X reaction buffer	5µl
10X DIG or Fluorescein labeled nucleotides	2µl
100mM DTT	2µl
DNA Template	0.1 - 0.3µg
SP6 or T7 RNA Polymerase	1µl
Water	up to 20µl

5. Add 30µl water and 30µl 8M LiCl and store @-20°C o/n.
6. Centrifuge @max speed for 30min @4°C
7. Remove the supernatant and wash the pellet with 70% EtOH
8. Air-dry the pellet and dissolve in water.
9. Check the concentration with nanodrop, aliquot and store @-80°C

Embryo preparation

10. Collect embryos at desired stages (before cuticle deposition: S28 / 216h / 8.5d) and dissect them in a fresh fixation buffer under the chemical hood.
11. Incubate at room temperature with mild nutation for 1h to 4h.
12. Wash embryos with PBT 8x15min
13. Dehydrate embryos in serial washes of (5min each): 10, 25, 50, 75% MetOH in PBT, and 100% MetOH. Replace MetOH with a fresh one and store the embryos @-20°C (up to several years).

Day 1: Hybridization

14. Take the embryos out of -20°C and incubate @RT for 10min.
15. Rehydrate with serial washes of (5min each) 90, 75, 50, 25%MetOH in PBT, and 100% PBT.
16. Wash embryos with PBT 2x5min
17. Fix embryos for 30min to 2h with 3,6% PFA in PBT.
18. Wash embryos with PBT 6x5min

19. Incubate embryos in detergent solution for 30min
20. Wash embryos with PBT 6x5min
21. Incubate embryos in 1:1 Hyb Buffer:PBT for 10min
22. Incubate embryos in Hyb Buffer for 10min
23. Replace the Hyb Buffer with the fresh one, incubate @65°C for 3h (up to o/n).
24. Prepare 0.1-1 ng/µl probe dilution in Hyb buffer. Boil it @90°C for 1min and incubate @65°C for 5min. Transfer embryos into it and incubate @65°C o/n.

Day 2: Antibody Staining

25. Recover the probe solution and store @-20°C (can be used several times)
26. Wash the embryos in Hyb Buffer 9x20min @65°C
27. Wash the embryos in Hyb Buffer 2x5min @RT
28. Serial washes in (5min each) 90, 70, 50, and 25% Hyb Buffer in TBST
29. Wash the embryos in TBST 4x15min
30. Incubate the embryos in $TBST^{+}$ solution for 1h
31. Dilute anti-DIG antibody (1:3000) or anti-Fluorescein (1:4000) in $TBST^{+}$, transfer the embryos in it and incubate @ 4°C o/n with mild nutation.

Day 3: AP Reaction

32. Wash the embryos in TBST 8x15min + 1x1h
33. Wash the embryos in AP buffer 3x5min
34. Prepare the reaction solution (4,5µl NBT and 3,5µl BCIP in 1ml AP buffer) and incubate embryos in the dark.
35. When the color developed to the desired level stop the reaction by washing several times with TBST.
36. Mount in 70% glycerol

Solutions:

1M Hepes: Dissolve 238.3g Hepes ($C_8H_{18}N_2O_4S$) in water for one liter of solution and adjust pH to 6.9 with NaOH.

20mM Magnesium Sulfate: Dissolve 4.93g $MgSO_4 7H_2O$ in water for one liter of solution

0.5M EDTA: Dissolve 186.12g in water for one liter of solution and adjust pH 8.0 with NaOH (pH 8.0 is required for salt to be dissolved).

0.5M EGTA: Dissolve 190.2g in water for one liter of solution and adjust pH 8.0 with NaOH (pH 8.0 is required for salt to be dissolved).

20% Tween-20: Dissolve 10ml of Tween-20 in water for 50ml solution

10X PBS: 2.56g $NaH_2PO_4 H_2O$ (18.6mM), 11.94g Na_2HPO_4 (84.1 mM) and 102.2g NaCl (1750 mM) per litre of water. Adjust pH to 7.4 (check pH again after diluting into 1XPBS).

20% SDS: Dissolve 200g SDS in water for one liter solution.

20X SSC: 175.3g (3M) NaCl and 88.2g (0.3M) Sodium Citrate, dihydrate ($CH_6H_8Na_3O_7.H_2O$) per liter of water. pH 7.0 and sterilize by autoclaving. Adjust the pH to 4.5 just before using (pH adjustment is very important!).

1M Tris pH 7.5: Dissolve 121.1g Tris Base in water for one liter solution and adjust pH with HCl at room temperature.

1M Tris pH 9.0: Dissolve 121.1g Tris Base in water for one liter solution and adjust pH with HCl at room temperature.

1M Tris pH 9.5: Dissolve 121.1g Tris Base in water for one liter solution and adjust pH with HCl at room temperature.

TBST: 12.5 mL 1M Tris pH 7.5 (250 mM), 13.7 mL 5M NaCl (1.37 M), 1.25 mL 1M KCl (25mM) and 250µl 20% Tween-20 in water for 50ml solution.

$TBST^+$: Mix 500 µl 10% BSA with 50 mL TBST.

5M NaCl: Dissolve 292.2g NaCl in water for one liter solution.

1M $MgCl_2$: Dissolve 203.3g $MgCl_2$, hexahydrate ($MgCl_2$6H2O) in water for 1 liter of solution.

20mg/ml Heparin: Dissolve 200mg Heparin for 10ml solution, store at -20 in aliquots.

10mg/ml Torula Yeast RNA: Dissolve 1gr of Torula Yeast RNA in 0.1M NaOAc buffer [13.6g NaOAc and 6ml glacial acetic acid for one liter solution adjust pH 5.0 with NaOH] to make 100ml of solution. aliquot and store at -20°C.

in situ fixation buffer: 500µl 1M Hepes, pH 6.9 [100mM]; 500µL 20mM Magnesium Sulfate [2mM]; 10µl 0.5M EGTA [1mM], 500µl 36% PFA, and 500µl 10X PBS for 5ml solution. Make fresh!

PBT: 5ml 10X PBS; 250µl 20% Tween-20 for 50ml solution (pH 7.4).

Detergent Solution: 2.5ml 20%SDS [1%]; 1.25ml 20% Tween-20 [0.5%]; 2.5ml 1M Tris-HCl (pH 7.5) [50mM]; 100µl 0.5M EDTA (pH 8.0) [1mM] and 1.5ml 5M NaCl [150mM] for 50ml solution.

SDS Hyb Buffer: 25ml Formamide [50%]; 12.5ml 20XSSC (ph 4.5) [5XSSC]; 125µl 20mg/ml heparin [50µg/ml]; 625µl 20% Tween-20 [0.25%]; 2.5ml 20% SDS [1%]; 2.5ml 10mg/ml Torula Yeast RNA [500µg/ml] for 50ml solution. The pH should be between 5.0 and 6.0! If not check the starting solution and start over (you can also adjust pH but you have to filter after). You can prepare fresh and keep at 65°C during the whole process.

AP Reaction Buffer: 250µl MgCl2 [5mM]; 1ml 5M NaCl [100mM]; 5ml 1M Tris-HCL pH 9.5 [100mM] and 250µl 20% Tween-20 [0.1%] for 50ml solution. Make just before use!

in vitro* sgRNA synthesis

1. Design F primer by replacing N with target gRNA

   F: gaaattaatacgactcactatagg'nnnnnnnnnnnnnnnnnnn'gttttagagctagaaatagc
   R: aaaagcaccgactcggtgccacttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaac

2. Set up a pcr reaction as: 20µl Phusion HF buffer, 67µl ddH2O, 2µl 10mM dNTPs, 5µl 10 µM CRISPR F primer, 5µl 10 µM CRISPR sgR primer and 1µl Phusion DNA polymerase
3. Cycle samples under the following conditions: 98°C 30", 35 cycles of [98°C 10",60°C 30" and 72°C 15"], 72°C 10', 4°C hold.
4. Run 5µl of the reaction to see the around 100bp band.
5. Purify the rest with pcr purification kit (elute in 30µl water) and it supposed to be around 150ng/µl
6. Synthesise the gRNA with Ambion MEGAshortscript kit:

T7 reaction buffer (10X)	2µl
dNTP mix (75mM)	8µl
DNA template (around 2µl)	300 ng
T7 enzyme mix	2µl
nuclease free water	up to 20µl

7. Incubate at 37°C for 4 hours.
8. Add 1µl of turbo DNAse and incubate at 37°C for 15 min.
9. Add 15µl of Ammonium Acetate stop solution (provided with the kit) and 115µl DEPC-water and mix.
10. Add 300µl of EtOH and put to -20°C overnight.
11. Centrifuge with max speed for 5 min at 4°C and resuspend the pellet in 40µl DEPC-water
12. Quantify with nanodrop, dilute into 1 mg/µl and store at -80°C.

Microinjection to *Parhyale embryos*

Needle preparation

Quartz capillaries with filaments (Sutter QF100-50-10) are pulled into injection needles with P-2000 micropipette puller (Sutter Instrument) with following settings:

Heat: 700
Filament: 4
Velocity: 50
Delay: 145
Pull: 175

Injection mix for HDR-mediated knock-ins*

*As performed by Arnaud Martin (Serano et al., 2016).

Prepare 6µl of injection mix is as:

sgRNA	200 ng/µl
Repair plasmid	200 ng/µl
Cas9	333 ng/µl
Phenol Red	%0,05

Injection mix for NHEJ-mediated knock-ins*

*As performed by Anastasios Pavlopoulos (Kao et al., 2016).

Prepare 6µl of injection mix is as:

sgRNA	40 ng/µl
Repair plasmid	10 ng/µl
Cas9	400 ng/µl
Phenol Red	%0

! Centrifuge the mix with maximum speed @4°C for 30min !

! Keep sgRNAs in ethanol @ -80 without precipitation !

! After precipitation keep the sgRNA @ -80 up to a month !

! After injecting, check the embryos 2-3 hours later and discard the ones that are not single cell stage !

! Trust nanodrop accuracy for sgRNA concentration but it'd be nice to check them on a gel for the first time after synthesis !

! sgRNA is working nicely between 20-200 ng/µl range !

Genomic DNA extraction from *Parhyale* embryos*

*Protocol received from Nikos Konstantinides

1. Place 15 to 50 embryos (depending on age) into a 1,5 ml microcentrifuge tube
2. Wash them 2-3 times with PBS (to get rid of salt)
3. Add 150µl of grinding buffer and homogenize
4. Rinse the debris on the pestle into a tube with 150µl more of the grinding buffer.
5. Vortex the tube for 10 sec and place immediately into 65°C for 30 min.
6. Add 600µl 5M KAc, mix by inverting, and place on ice for 30 min.
7. Transfer the sample into a 2mL microcentrifuge tube containing 900µl phenol-chloroform (1:1) shake well and spin with max speed for 15 min at 4°C
8. Transfer the supernatant into a new microcentrifuge tube containing 900µl Chloroform; shake well and spin with max speed for 15 min at 4°C
9. Transfer the supernatant into a new microcentrifuge tube containing 800µl isopropanol, mix well and keep on bench for 20 min.
10. Centrifuge at 13000rpm for 15 min at 4°C
11. Remove the liquid and wash the pellet with 600µl of 70% EtOH.
12. Centrifuge at 13000rpm for 15 min at 4°C, remove the liquid and air dry the pellet
13. Add 20µl of water to dissolve the pellet.

Grinding Buffer: 0.1 M Tris pH 9.2, 0.2M Sucrose, 50mM EDTA, 0.5% SDS

T7 endonuclease assay*

*This is a slight modification of a protocol received from Johannes Schinko

1. Amplify the desired genomic piece in a 50µl PCR reaction (30 to 35 cycles)
   ! Use a high fidelity DNA polymerase !
2. Add PCR reaction in a microcentrifuge tube with 300µl Phenol-Chloroform (1:1)
3. Add 250µl water, shake well, and centrifuge with the max speed at 4°C for 15 min.
4. Transfer the upper layer into a new microcentrifuge tube containing 300µl Chloroform, shake well, and centrifuge with the max speed at 4°C for 15 min.
5. Transfer the upper layer into a new tube with 300µl isopropanol and 30µl of 3M NaOAc, incubate on the bench for 20 min.
6. Centrifuge with the max speed at 4°C for 15 min.

7. Remove the liquid and wash the pellet with 500µl 70% EtOH and centrifuge with the max speed at 4°C for 15 min.
8. Remove the liquid and air dry the pellet
9. Dissolve pellet in 17,5µl of water and transfer into a PCR tube and add 2µl of buffer 2 (NEB restriction enzyme buffer)
10. Melt the DNA with the following protocol in a PCR machine:
11. Add 0.5µl of T7 Endonuclease and incubate at 37°C for 15 min.

    ! Add the enzyme directly into the solution, not to the tube wall !
12. Load all reaction on an agarose gel

Antibody staining on *Parhyale* embryos

1. Dissect embryos in cold PBS (max for 20min).
2. Transfer them to 3.6% PFA and incubate at room temperature for 15min.
3. Wash 3X10 min with PBX.
4. !!! Steps 4-6 are optional !!!
5. Dehydrate embryos with serial washes of 25, 50, 75 and 100% MetOH in PBX for 5min each @RT (you can store in 100% MetOH @ -20).
6. Rehydrate embryos with serial washes of 75, 50, 25% MetOH in PBX for 5min each @RT
7. Wash 3X5 min with PBX @RT.
8. Wash 60min with PAXD-1 (up to 2 hours) @RT.
9. Put the primary antibodies in PAXD-1 and incubate @4°C o/n.
10. Wash 3X15 min with PBX @RT.
11. Put secondary antibodies in PAXD-1 and incubate @4°C o/n (or @RT 2hours).
12. Wash 3X15 min with PBX @RT.
13. Transfer into 50% Glycerol (you can add DAPI here) for 30 min (or o/n @4°C).
14. Transfer into Vectashield and incubate @RT for 30 min (or o/n @4°C), and mount.

PBX: 0.1% Triton X-100 in PBS

PAXD-1: 0.1% Triton-X-100, 0.1% Sodium Deoxycholate, 5% BSA, 0.5% NGS in PBS

ATACseq on *Parhyale* embryos*

*Adapted from the protocol of Amanda Ackerman from Klaus Kaestner Lab, who derived it from (Buenrostro et al., 2013; 2015).

Cell preparation

1. Cut the lid of a eppendorf tube and place the embryo along with 50µl cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630). Dissect and remove the membrane, dissociate cells quickly by gentle pipetting and spin down immediately at 800 xg for 10 min, 4°C.
2. Discard supernatant and immediately continue to transposition reaction.

   !!! Pellet is not visible !!!
3. Transposition reaction and purification
4. Make sure the cell pellet is set on ice.
5. To make the transposition reaction mix, combine the following:

2x TD Buffer (Illumina Cat #FC-121-1030)	25 µl
Tn5 Transposes (Illumina Cat #FC-121-1030)	1.5 µl
Nuclease Free H_2O	22.5 µl

6. Resuspend nuclei in the transposition reaction mix.
7. The transposition reaction is carried out at 37°C for 30 min.
8. Following transposition reaction, the sample is purified using a Qiagen MinElute kit.
9. Elute transposed DNA in 11.5 µl Elution Buffer (10 mM Tris buffer pH 8.0).
10. 1 µl can be used for Qubit analysis, although we have seen concentrations between 0.3 ng/µl – 10 ng/µl all work.

    ! Purified DNA can be stored at -20°C !

PCR Amplification

11. To amplify transposed DNA fragments, combine the following in a PCR tube:

Tagmented DNA	10 µl
Nuclease Free H_2O	10 µl
Nextera fw PCR primer (general fw primer)	2.5 µl
Nextera rev PCR primer (specific rev primer, contains barcode)	2.5 µl
NEBNext High-Fidelity 2x PCR Master Mix (Bioke Cat #M0541)	25 µl

12. Cycle as follows:

    72°C, 5 min
    98°C, 30 sec
    98°C, 10 sec
    63°C, 30 sec
    72°C, 1 min
    Repeat steps 3-5, 9-12x
    Hold at 4°C

Library Purification

13. Warm AMPure XP beads to room temperature, and vortex to resuspend.

    ! For single left-sided bead purification (to remove primer dimers) !

14. Transfer each PCR sample to a microcentrifuge tube, add 1.8X volume (81 µl) AMPure XP beads, pipet up and down 10x to mix thoroughly.
15. Incubate at room temperature for 10 minutes.
16. Place microcentrifuge tubes in a magnetic rack for 5 minutes.
17. Discard supernatant.
18. Wash beads with 200 µl 80% EtOH (freshly made), pipet EtOH over beads 10x, then discard EtOH.
19. Leave the tube on a magnetic rack with the cap open for 10 minutes.

    ! Ensure all EtOH is removed !

20. Resuspend beads in 20 µl nuclease-free H2O, pipet up and down 10 times to mix thoroughly.
21. Place the microcentrifuge tube in a magnetic rack for 1-5 minutes.
22. Transfer supernatant to a new microcentrifuge tube.
23. Check the sample with TapeStation, quantify with Qubit and proceed with sequencing.

    !!! Purified libraries should be stored @ -20°C !!!

Total RNA extraction from *Parhyale* tissues

1. Put desired tissues into a 1,5mL microcentrifuge tube with 1mL Trizol and homogenise quickly with pestle (work under the hood!)
2. Put the mix into Qiashredder columns (700µl max) and spin for 2min @max speed.
3. Keep @RT for 3min
4. Add 200µl of chloroform and shake vigorously.
5. Incubate @RT for 3min
6. Centrifuge with max speed @4°C for 10min
7. Remove supernatant into a new eppendorf tube with 500µl of isopropanol.
8. Incubate @RT for 5min and centrifuge with max speed for 30min @4°C
9. Remove liquid and wash the pellet with 70% EtOH (RNAse free!).
10. Air dry the pellet and dissolve in 50µl of water.

B - Tracking cell lineages in 3D by incremental deep learning

Ko Sugawara[1,2,*], Cagri Cevrim[1,2] and Michalis Averof[1,2,*]

[1]Institut de Génomique Fonctionnelle de Lyon (IGFL), École Normale Supérieure de Lyon, 32 avenue Tony Garnier, 69007 Lyon, France

[2]Centre National de la Recherche Scientifique (CNRS), France

[*]Authors for correspondence (ko.sugawara@ens-lyon.fr, michalis.averof@ens-lyon.fr)

Abstract

Deep learning is emerging as a powerful approach for bioimage analysis, but its wider use is limited by the scarcity of annotated data for training. We present ELEPHANT, an interactive platform for cell tracking in 4D that seamlessly integrates annotation, deep learning and proofreading. ELEPHANT's user interface supports cycles of incremental learning starting from sparse annotations, yielding accurate, user-validated cell lineages with a modest investment in time and effort.

Main text

Recent progress in deep learning has led to significant advances in bioimage analysis[1-4]. As deep learning is data-driven, it is adaptable to a variety of datasets once an appropriate model architecture is selected and trained with adequate data[1]. In spite of its powerful performance, deep learning remains challenging for non-experts to utilize, for three reasons. First, pre-trained models can be inadequate for new tasks and the preparation of new training data is laborious. Because the quality and quantity of the training data are crucial for the performance of deep learning, users must invest significant time and effort in annotation at the start of the project [1]. Second, an interactive user interface for deep learning, especially in the context of cell tracking, is lacking. Third, deep learning applications are often limited by accessibility to computing power (high-end GPU). We have addressed these challenges by establishing ELEPHANT (Efficient learning using sparse human annotations for nuclear tracking), an interactive web-friendly platform for cell tracking, which seamlessly integrates manual annotation with deep learning and proofreading of the results. ELEPHANT implements two algorithms optimized for incremental deep learning using sparse annotations, one for detecting nuclei in 3D and a second for linking these nuclei across timepoints in 4D image datasets. Incremental learning allows models to be trained in a stepwise fashion on a given dataset, starting from sparse annotations that are incrementally enriched by human proofreading, leading to a rapid increase in performance (Figure 1a). ELEPHANT is implemented as an extension of Mastodon (https://github.com/mastodon-sc/mastodon), an open source framework for large-scale tracking based on Fiji[5]. It works on a client-server model, in which the server provides a deep learning environment equipped with sufficient GPU (Supplementary Figure XXX).

ELEPHANT employs the tracking-by-detection paradigm[6], which involves initially the *detection* of nuclei in 3D and subsequently their *linking* over successive timepoints to generate tracks. In both steps, the nuclei are represented as ellipsoids, using the data model of Mastodon (Figure 1b, c). In the detection phase, voxels are labeled as *background*, *nucleus center* or *nucleus periphery*, or left unlabeled (Figure 1b, top right). The *nucleus center* and *nucleus periphery* labels are generated by the annotation of nuclei, and the *background* can be annotated either manually or by intensity thresholding. Sparse annotations (e.g. of a few nuclei in a single frame) are sufficient to start training. A U-Net convolutional neural network (U-Net CNN[7,8]) is then trained on these labels (ignoring the unlabeled voxels) to generate voxel-wise probability maps for background, nucleus center or nucleus periphery across the entire image dataset (figure 1b, bottom right). Post-processing on these probability maps

yields predictions of nuclei, which are available for visual inspection and proofreading (validation or rejection of each predicted nucleus) by the user (figure 1b, bottom left). Human-computer interaction is facilitated by color-coding of the annotated nuclei (as true positive, false positive, true negative, false negative, or unevaluated, see Supplementary Materials XXX) based on the proofreading. The cycles of training and prediction are rapid because only a small amount of training data are added each time (in the order of seconds, see Supplementary Materials XXX). As a result, users can enrich the annotations by proofreading the output almost simultaneously, enabling incremental training of the model in an efficient manner.

In the linking phase, we found that nearest neighbour approaches for tracking nuclei over time[9] perform poorly in challenging datasets, when the cells are dividing (see below), hence we turned to optical flow modeling to improve linking[10-12]. A second U-Net CNN, optimized for optical flow estimation, is trained on manually generated/validated links between nuclei in successive timepoints (Figure 1c, top left). Unlabeled regions are ignored, hence training can be performed on sparse linking annotations (figure 1c, top right). The flow model is used to generate voxel-wise 3D flow maps, representing predicted x, y and z displacements over time (figure 1c, bottom right), which are then combined with nearest neighbour linking to generate links between the detected nuclei (see Methods). Users proofread the linking results to finalize the tracks and to update the labels for the next iteration of training (figure 1c, bottom left).

We evaluated the performance of ELEPHANT using two types of 4D confocal microscopy recordings in which nuclei were visualized by fluorescent markers: one type capturing the embryonic development of *Caenorhabditis elegans* (CE datasets), which has been used before for benchmarking of tracking methods[13,14], and another capturing limb regeneration in *Parhyale hawaiensis* (PH dataset, imaging adapted from[15]), which presents greater challenges for image analysis (see below). On both types of dataset, we find that less than ten annotated nuclei are sufficient to initiate a virtuous cycle of training, prediction and proofreading, which efficiently yields cell tracks and validated cell lineages in highly dynamic tissues.

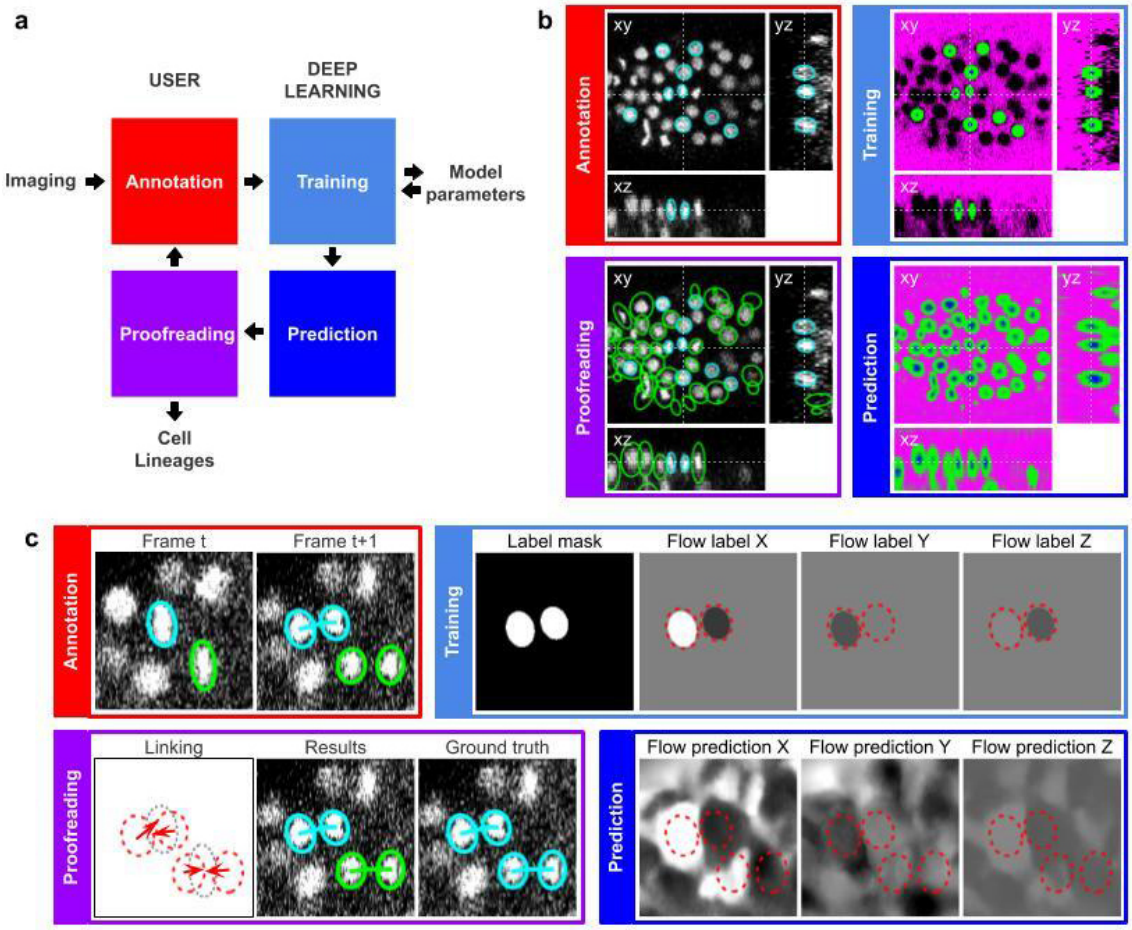


Figure 1: Overview of ELEPHANT.
a, Schematic illustration of incremental learning with ELEPHANT. Imaging data are fed into a cycle of annotation, training, prediction and proofreading to generate cell lineages. At each iteration, model parameters are updated and saved. This workflow applies to both detection and linking phases. b, Detection workflow, illustrated with orthogonal views on the CE1 dataset. Top left: The user annotates nuclei with ellipsoids in 3D; newly generated annotations are colored in cyan. Top right: The detection model is trained with the labels generated from the sparse annotations of nuclei and from the annotation of background (in this case by intensity thresholding); background, nucleus center, nucleus periphery and unlabelled regions are indicated in magenta, blue, green and black, respectively. Bottom right: The trained model generates voxel-wise probability maps for background (red), nucleus center (blue) or nucleus periphery (green). Bottom left: The user validates or rejects the predictions; predicted nuclei are shown in green, predicted and validated nuclei in cyan. c, Linking workflow, illustrated on the CE1 dataset. Top left: The user annotates links by connecting detected nuclei in successive timepoints; annotated/validated nuclei and links are shown in cyan, non-validated ones in green. Top right: The flow model is trained with optical flow labels coming from annotated nuclei with links (voxels indicated in the label mask), which consist of displacements in X, Y and Z; greyscale values indicate displacements along a given axis, annotated nuclei with link labels are outlined in red. Bottom right: The trained model generates voxel-wise flow maps for each axis; greyscale values indicate displacements, annotated nuclei are outlined in red. Bottom left: The user validates or rejects the predictions; predicted links are shown in green, predicted and validated links in cyan.

Interactive cycles of manual annotation, deep learning and proofreading on ELEPHANT reduce the time required to complete tracks (figure 2a). On the CE1 dataset, a complete cell lineage was built over 195 timepoints, from scratch, using ELEPHANT's semi-automated workflow (figure 2c). The detection model was trained

incrementally starting from sparse annotations (four nuclei) on the first timepoint. On this dataset, linking could be performed using the nearest neighbor algorithm (without flow modeling) and manual proofreading. In this way, we were able to annotate in less than 8 hours a total of 23,829 nuclei (across all timepoints), of which ~2% were manually annotated (483 nuclei) and the remaining nuclei were collected by validating predictions of the deep-learning model.

Although ELEPHANT works efficiently without prior training, cell tracking can be accelerated by starting off from models trained on image data with similar characteristics. To illustrate this, we used nuclear annotations in a separate dataset, CE2, to train a model for detection, which was then applied to CE1. This pre-trained model allowed us to track CE1 much more rapidly and effortlessly than with an untrained model (figure 2a, blue versus orange lines). For benchmarking, the detection and linkage models trained with the annotations from the CE1 and CE2 lineage trees were then tested on unseen datasets with similar characteristics, as part of the Cell Tracking Challenge[6,14]. In this test, our models with assistance of flow-based interpolation (Supplementary Materials XXX) outperformed state-of-the-art tracking algorithms[16,17] in detection (DET) and linking (TRA) metrics (figure 2b).

The PH dataset presents greater challenges for image analysis, such as larger variations in the shape, intensity and distribution of nuclei, lower temporal resolution, and more noise (Supplementary Figure XXX). ELEPHANT has allowed us to grapple with these issues by supporting the continued training of the models through visual feedback from the user (annotation of missed nuclei, validation and rejection of predictions). Using ELEPHANT, we annotated and validated over 260,000 nuclei in this dataset, across 504 timepoints that span 168 hours of imaging.

We observed that the conventional nearest neighbor approach was inadequate for linking in the PH dataset, leaving many errors in the lineage trees (Figure 2d). This is likely due to the lower temporal resolution in this dataset (20 minutes in PH, versus 1-2 minutes in CE) and the fact that daughter nuclei often show large displacements at the end of mitosis. We trained optical flow using 78,316 validated links (including 168 links for 84 cell divisions). These sparse annotations were sufficient to generate 3D optical flow predictions for the entire dataset (Supplementary Figure XXX), which significantly improved the linking performance (Figure 2d, Supplementary Figure XXX): the number of false positive and false negative links decreased by ~60% (from 1,704 to 744) and ~40% (from 1,542 to 957), respectively, among a total of 259,071 links. We could also obtain a similar linking performance after training with a much smaller

number of links (1,162 validated links, see Supplementary Materials XXX), indicating that even a modest degree of training can be effective.

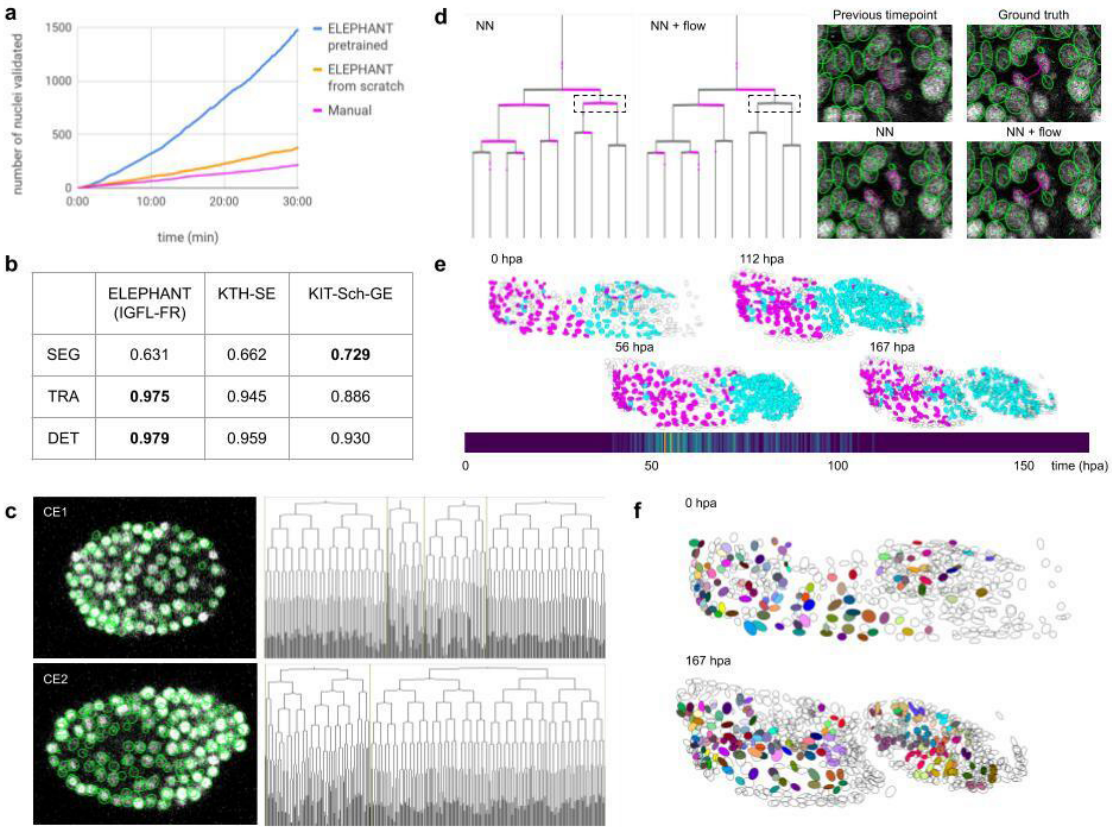


Figure 2: Evaluation of performance and applications.
a, Comparison of the speed of detection and validation of nuclei on the CE1 dataset, by manual annotation (magenta), semi-automated detection without a pre-trained model (orange) and semi-automated detection using a pre-trained model (blue) using ELEPHANT. **b**, Performance of ELEPHANT compared with two state-of-the-art algorithms [16,17], using the metrics of the Cell Tracking Challenge on unseen CE datasets. ELEPHANT outperforms the other methods in detection and linking accuracy (DET and TRA metrics, respectively); it performs less well in segmentation accuracy (SEG). **c**, Tracking results obtained with ELEPHANT. Left panels: Tracked nuclei in the CE1 and CE2 datasets at timepoints 194 and 189, respectively. Representative optical sections are shown with tracked nuclei shown in green; out of focus nuclei are shown as green spots. Right panels: Corresponding lineage trees. **d**, Comparison of tracking results obtained on the PH dataset, using the nearest neighbor algorithm (NN) with and without optical flow prediction (left panels); linking errors are highlighted in magenta on the correct lineage tree. The panels on the right focus on the nuclear division that is marked by a dashed line rectangle. Without optical flow prediction, the dividing nuclei (in magenta) are linked incorrectly. **e**, Spatial and temporal distribution of dividing cells in the regenerating leg of *Parhyale* over a 1-week time course (PH dataset), showing that cell proliferation is concentrated at the distal part of the regenerating leg stump and peaks after a period of proliferative quiescence, as described in [15]. Top: Nuclei in lineages that contain at least one division are colored in cyan, nuclei in non-dividing lineages are in magenta, and nuclei in partly tracked lineages that left division status undetermined are blank. Bottom: Heat map of the temporal distribution of nuclear divisions; hpa, hours post amputation. The number of divisions per 20-minute time interval ranges from 0 (purple) to 9 (yellow). **f**, Fate map of regenerating leg of *Parhyale*, encompassing 109 fully tracked lineage trees (202 cells at 167 hpa). Each clone is assigned a unique color and contains 1-9 cells at 167 hpa. Partly tracked nuclei are blank. In panels e and f the amputation plane (distal end of the limb) is located on the right.

By applying ELEPHANT's human-in-the-loop semi-automated workflow, we succeeded in reconstructing 109 complete and fully-validated cell lineage trees encompassing the duration of leg regeneration in *Parhyale*, each lineage spanning a period of ~1 week (504 timepoints, Supplementary Materials XXX). Using analysis and visualization modules implemented on Mastodon and ELEPHANT we could illustrate the distribution of cell divisions in time and space (figure 2e) and produce a fate map of the regenerating leg of *Parhyale* (figure 2f). This lineaging task, which would have required several months of manual annotation, was achieved in ~1 month of interactive cell tracking in ELEPHANT, without prior training. Re-using the trained models and transferring the best performing ones to new data could improve tracking efficiency even further.

Acknowledgements

We are grateful to Anna Kreshuk and Constantin Pape for training in machine learning, to Jean-Yves Tinevez and Tobias Pietzsch for support in developing ELEPHANT as a Mastodon plugin, and to the NEUBIAS community for feedback on the software. We also thank Sebastien Tosi, Carsten Wolff and XXX for comments on the manuscript. This research was supported by the European Research Council, under the European Union Horizon 2020 programme, grant ERC-2015-AdG #694918.

Competing interests

KS is employed part-time by LPixel Inc.

Contributions

KS and MA conceived the project; KS designed and produced the software, and evaluated its performance; CC acquired the image dataset on regenerating limbs; KS and CC generated the annotations and tested the software; KS and MA wrote the manuscript.

References

1. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat Meth* **16,** 1233–1246 (2019).
2. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36,** 460–468 (2018).
3. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Meth* **15,** 1090–1097 (2018).
4. Caicedo, J. C. *et al.* Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16,** 1247–+ (2019).
5. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Meth* **9,** 676–682 (2012).
6. Maška, M. *et al.* A benchmark for comparison of cell tracking algorithms. *Bioinformatics* **30,** 1609–1617 (2014).
7. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* **9351,** 234–241 (Springer, Cham, 2015).
8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* **9901,** 424–432 (Springer, Cham, 2016).
9. Crocker, J. C. & Grier, D. G. Methods of Digital Video Microscopy for Colloidal Studies. *Journal of Colloid and Interface Science* **179,** 298–310 (1996).
10. Horn, B. & Schunck, B. G. Determining Optical Flow Artificial Intelligence Vol. 17. *Artificial Intelligence* **17,** 185–203 (1981).
11. Lucas, B. D. & Kanade, T. An iterative image registration technique with an application to stereo vision. *Proceedings of the th international joint conference on Artificial intelligence* **2,** 674–679 (1981).
12. Amat, F., Myers, E. W. & Keller, P. J. Fast and robust optical flow for time-lapse microscopy using super-voxels. *Bioinformatics* **29,** 373–380 (2012).
13. Murray, J. I. *et al.* Automated analysis of embryonic gene expression with cellular resolution in C. elegans. *Nat Meth* **5,** 703–709 (2008).
14. Ulman, V. *et al.* An objective comparison of cell-tracking algorithms. *Nat Meth* **14,** 1141–1152 (2017).
15. Alwes, F., Enjolras, C. & Averof, M. Live imaging reveals the progenitors and cell dynamics of limb regeneration. *Elife* **5,** 73 (2016).
16. Scherr, T., Löffler, K., Böhland, M. & Mikut, R. Cell Segmentation and Tracking using CNN-Based Distance Predictions and a Graph-Based Matching Strategy. *arXiv.org* **cs.CV,** (2020).
17. Magnusson, K. E. G., Jaldén, J., Gilbert, P. M. & Blau, H. M. Global linking of cell tracks using the Viterbi algorithm. *IEEE Trans Med Imaging* **34,** 911–929 (2015).
18. Wolff, C. *et al.* Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb. *Elife* **7,** 375 (2018).
19. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. BigDataViewer: visualization and processing for large image data sets. *Nat Meth* **12,** 481–483 (2015).
20. Harris, C. R. *et al.* Array programming with NumPy. *Nature Publishing Group* **585,** 357–362 (2020).
21. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Meth* **17,** 261–272 (2020).
22. van der Walt, S. *et al.* scikit-image: Image processing in Python. *Peer J* **cs.MS,** e453 (2014).
23. Tseng, Q. *et al.* A new micropatterning method of soft substrates reveals that different tumorigenic signals can promote or reduce cell contraction levels. *Lab Chip* **11,** 2231–2240 (2011).
24. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9,** 671–675 (2012).
25. He, K., Zhang, X., Ren, S. & Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *IEEE International Conference on Computer Vision* 1026–1034 (2015). doi:doi:10.1109/ICCV.2015.123
26. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015).
27. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (2013).
28. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* **13,** 600–612 (2004).
29. Matula, P. *et al.* Cell Tracking Accuracy Measurement Based on Comparison of Acyclic Oriented Graphs. *PLoS ONE* **10,** e0144959 (2015).

C – References

Adli M. 2018. The CRISPR tool kit for genome editing and beyond. *Nat Comms* **9**:1–13. doi:10.1038/s41467-018-04252-2

Aldaz S, Escudero LM. 2010. Imaginal discs. *Curr Biol* **20**:R429–31. doi:10.1016/j.cub.2010.03.010

Alwes F, Enjolras C, Averof M. 2016. Live imaging reveals the progenitors and cell dynamics of limb regeneration. *eLife Sciences*. doi:10.7554/eLife.19766.001

Ando K, Shibata E, Hans S, Brand M, Kawakami A. 2017. Osteoblast Production by Reserved Progenitor Cells in Zebrafish Bone Regeneration and Maintenance. *Developmental Cell* **43**:643–650.e3. doi:10.1016/j.devcel.2017.10.015

Anzalone AV, Koblan LW, Liu DR. 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**:824–844. doi:10.1038/s41587-020-0561-9

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149–157. doi:10.1038/s41586-019-1711-4

Arnone MI, Davidson EH. 1997. The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124**:1851–1864.

Ayyar S, Negre B, Simpson P, Stollewerk A. 2010. An arthropod cis-regulatory element functioning in sensory organ precursor development dates back to the Cambrian. *BMC Biol* **8**. doi:10.1186/1741-7007-8-127

Azevedo A, Sousa S, Jacinto A, Saúde L. 2012. An amputation resets positional information to a proximal identity in the regenerating zebrafish caudal fin. *BMC Dev Biol* **12**:24–10. doi:10.1186/1471-213X-12-24

Baghdadi MB, Tajbakhsh S. 2018. Regulation and phylogeny of skeletal muscle regeneration. *Developmental Biology* **433**:200–209. doi:10.1016/j.ydbio.2017.07.026

Baguna J, Salo E, Auladell C. 1989. Regeneration and pattern formation in planarians. *Development* **107**:77–86.

Bai H, Liu L, An K, Lu X, Harrison M, Zhao Y, Yan R, Lu Z, Li S, Lin S, Liang F, Qin W. 2020. CRISPR/Cas9-mediated precise genome modification by a long ssDNA template in zebrafish. *BMC Genomics* **21**:1–12. doi:10.1186/s12864-020-6493-4

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *CELL* **27**:299–308. doi:10.1016/0092-8674(81)90413-X

Barker DL, Herbert E, Hildebrand JG, Kravitz EA. 1972. Acetylcholine and lobster sensory neurones. *The Journal of Physiology* **226**:205–229. doi:10.1113/jphysiol.1972.sp009981

Baudouin-Gonzalez L, Schoenauer A, Harper A, Blakeley G, Seiter M, Arif S, Sumner-Rooney L, Russell S, Sharma PP, McGregor AP. 2020. The evolution of Sox gene repertoires and regulation of segmentation in arachnids. *bioRxiv* **9**:M68–57. doi:10.1101/2020.06.04.133389

Beltrami AP, Barlucchi L, Torella D, Baker M, Limana F, Chimenti S, Kasahara H, Rota M, Musso E, Urbanek K, Leri A, Kajstura J, Nadal-Ginard B, Anversa P. 2003. Adult cardiac stem cells are multipotent and support myocardial regeneration. *CELL* **114**:763–776. doi:10.1016/s0092-8674(03)00687-1

Bely AE, Nyberg KG. 2010. Evolution of animal regeneration: re-emergence of a field. *Trends in Ecology & Evolution* **25**:161–170. doi:10.1016/j.tree.2009.08.005

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim Y-G, Chandrasegaran S. 2001. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Molecular and Cellular Biology* **21**:289–297. doi:10.1128/MCB.21.1.289-297.2001

Blanco J, Girard F, Kamachi Y, Kondoh H, Gehring WJ. 2005. Functional analysis of the chicken δ1-crystallin enhancer activity in Drosophila reveals remarkable evolutionary conservation between chicken and fly. *Development* **132**:1895–1905. doi:10.1242/dev.01738

Blochlinger K, Bodmer R, Jan LY, Jan YN. 1990. Patterns of expression of cut, a protein required for external sensory organ development in wild-type and cut mutant Drosophila embryos. *Genes & Development* **4**:1322–1331.

Bosch T, David CN. 1987. Stem-Cells of Hydra-Magnipapillata Can Differentiate Into Somatic-Cells and Germ Line Cells. *Developmental Biology* **121**:182–191. doi:10.1016/0012-1606(87)90151-5

Bosch TCG. 2007. Why polyps regenerate and we don't: Towards a cellular and molecular framework for Hydra regeneration. *Developmental Biology* **303**:421–433. doi:10.1016/j.ydbio.2006.12.012

Boukhatmi H, Bray S. 2018. A population of adult satellite-like cells in Drosophila is maintained through a switch in RNA-isoforms. *eLife Sciences* **7**. doi:10.7554/elife.35954

Brandt A. 1988. Morphology and ultrastructure of the sensory spine, a presumed mechanoreceptor of Sphaeroma hookeri (Crustacea, Isopoda), and remarks on similar spines in other peracarids. *J Morphol* **198**:219–229. doi:10.1002/jmor.1051980208

Browne WE, Price AL, Gerberding M, Patel NH. 2005. Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. *Genesis* **42**:124–149. doi:10.1002/gene.20145

Buchwalow I, Samoilova V, Boecker W, Tiemann M. 2011. Non-specific binding of antibodies in immunohistochemistry: fallacies and facts. *Sci Rep* **1**. doi:10.1038/srep00028

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Meth* **10**:1213–1218. doi:10.1038/nmeth.2688

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**:21.29.1–21.29.9. doi:10.1002/0471142727.mb2129s109

Campos AR, Grossman D, White K. 1985. Mutant alleles at the locus elav in Drosophila melanogaster lead to nervous system defects. A developmental-genetic analysis. *J Neurogenet* **2**:197–218. doi:10.3109/01677068509100150

Carlson BM. 2007. Principles of Regenerative Biology. Academic Press.

Carvell GE, Simons DJ. 1990. Biometric analyses of vibrissal tactile discrimination in the rat **10**:2638–2648. doi:10.1523/JNEUROSCI.10-08-02638.1990

Cenik C, Derti A, Mellor JC, Berriz GF, Roth FP. 2010. Genome-wide functional analysis of human 5' untranslated region introns. *Genome Biology* **11**:1–17. doi:10.1186/gb-2010-11-3-r29

Charmantier-Daures M, Vernet G. 2004. Moulting, autotomy, and regeneration In: Forest J, Vaupel Klein von C, Schram F, editors. Treatise on Zoology - Anatomy, Taxonomy, Biology - the Crustacea. Leiden. pp. 161–254.

Chaturvedi D, Reichert H, Gunage RD, VijayRaghavan K. 2017. Identification and functional characterization of muscle sateilte cells in Drosophila. *eLife Sciences* **6**. doi:10.7554/eLife.30107

Chen Y, Lin G, Slack JMW. 2006. Control of muscle regeneration in the Xenopus tadpole tail by Pax7. *Development* **133**:2303–2313. doi:10.1242/dev.02397

Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F, Herrera PL. 2014. Diabetes recovery by age-dependent conversion of pancreatic δ-cells into insulin producers. *Nature* **514**:503–507. doi:10.1038/nature13633

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**:757–761. doi:10.1534/genetics.110.120717

Clapier CR, Cairns BR. 2009. The Biology of Chromatin Remodeling Complexes **78**:273–304. doi:10.1146/annurev.biochem.77.062706.153223

Clarac F. 1990. Proprioception from Chordotonal Organs in Crustacean LimbsFrontiers in Crustacean Neurobiology. Basel: Birkhäuser, Basel. pp. 262–270. doi:10.1007/978-3-0348-5689-8_30

Clare AS, Lumb G, Clare PA, Costlow JD. 1990. A morphological study of wound response and telson regeneration in postlarval Limulus polyphemus (L.). *Invertebrate Reproduction & Development* **17**:77–87. doi:10.1080/07924259.1990.9672091

Clark-Hachtel CM, Tomoyasu Y. 2020. Two sets of candidate crustacean wing homologues and their implication for the origin of insect wings. *Nature Ecology & Evolution* **13**:1–16. doi:10.1038/s41559-020-1257-8

Colombelli J, Lorenzo C. 2014. Light sheet fluorescence microscopy applications for multicellular systemsFluorescence Microscopy. Elsevier. pp. 109–120. doi:10.1016/B978-0-12-409513-7.00008-7

Colombrita C, Silani V, Ratti A. 2013. ELAV proteins along evolution: Back to the nucleus? *Molecular and Cellular Neuroscience* **56**:447–455. doi:10.1016/j.mcn.2013.02.003

Cooper RL. 1998. Development of sensory processes during limb regeneration in adult crayfish **201**:1745–1752.

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, Zhou D, Luo S, Vasicek TJ, Daly MJ, Wolfsberg TG, Collins FS. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Research* **16**:123–131. doi:10.1101/gr.4074106

Crawford K, Quiroz JFD, Koenig KM, Ahuja N, Albertin CB, Rosenthal JJC. 2020. Highly Efficient Knockout of a Squid Pigmentation Gene. *Current Biology* **30**:3484–3490.e4. doi:10.1016/j.cub.2020.06.099

Cui Y, Xu J, Cheng M, Liao X, Peng S. 2018. Review of CRISPR/Cas9 sgRNA Design Tools. *Interdiscip Sci* **10**:455–465. doi:10.1007/s12539-018-0298-z

Currie JD, Kawaguchi A, Traspas RM, Schuez M, Chara O, Tanaka EM. 2016. Live Imaging of Axolotl Digit Regeneration Reveals Spatiotemporal Choreography of Diverse Connective Tissue Progenitor Pools. *Developmental Cell* **39**:411–423. doi:10.1016/j.devcel.2016.10.013

Dan H. n.d. Sanes, Thomas A. Reh, William A. Harris: Development of the nervous system, -Elsevier.-ISBN 978-0-12-618621-5.

Das S. 2015. Morphological, Molecular, and Hormonal Basis of Limb Regeneration across Pancrustacea. *Integrative and Comparative Biology* **55**:869–877. doi:10.1093/icb/icv101

Dasyani M, Tan WH, Sundaram S, Imangali N, Centanin L, Wittbrodt J, Winkler C. 2019. Lineage tracing of col10a1 cells identifies distinct progenitor populations for osteoblasts and joint cells in the regenerating fin of medaka (Oryzias latipes). *Developmental Biology* **455**:85–99. doi:10.1016/j.ydbio.2019.07.012

David CN. 2012. Interstitial stem cells in Hydra: multipotency and decision-making. *Int J Dev Biol* **56**:489–497. doi:10.1387/ijdb.113476cd

Denker A, de Laat W. 2016. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes & Development* **30**:1357–1382. doi:10.1101/gad.281964.116

Diaspro A, Federici F, Robello M. 2002. Influence of refractive-index mismatch in high-resolution three-dimensional confocal microscopy. *Appl Opt, AO* **41**:685–690. doi:10.1364/AO.41.000685

Diep CQ, Ma D, Deo RC, Holm TM, Naylor RW, Arora N, Wingert RA, Bollig F, Djordjevic G, Lichman B, Zhu H, Ikenaga T, Ono F, Englert C, Cowan CA, Hukriede NA, Handin RI, Davidson AJ. 2011. Identification of adult nephron progenitors capable of kidney regeneration in zebrafish. *Nature* **470**:1–7. doi:10.1038/nature09669

Dillaman RM, Roer R, Shafer T, Modla S. 2013. The Crustacean Integument: Structure and Function, Functional Morphology and Diversity. Oxford University Press. doi:10.1093/acprof:osobl/9780195398038.003.0005

Diogo R, Nacu E, Tanaka EM. 2014. Is Salamander Limb Regeneration Really Perfect? Anatomical and Morphogenetic Analysis of Forelimb Muscle Regeneration in GFP-Transgenic Axolotls as a Basis for Regenerative, Developmental, and Evolutionary Studies. *Anat Rec* **297**:1076–1089. doi:10.1002/ar.22906

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**:1262–U130. doi:10.1038/nbt.3026

Domínguez-Bendala J, Qadir MMF, Pastori RL. 2019. Pancreatic Progenitors: There and Back Again. *Trends in Endocrinology & Metabolism* **30**:4–11. doi:10.1016/j.tem.2018.10.002

Doyle HJ, Kraut R, Levine M. 1989. Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. *Genes & Development* **3**:1518–1533. doi:10.1101/gad.3.10.1518

Echeverri K, Tanaka EM. 2002. Ectoderm to mesoderm lineage switching during axolotl tail regeneration. *Science* **298**:1993–1996. doi:10.1126/science.1077804

Egner A, Hell SW. 2006. Aberrations in Confocal and Multi-Photon Fluorescence Microscopy Induced by Refractive Index MismatchHandbook of Biological Confocal Microscopy. Boston, MA: Springer, Boston, MA. pp. 404–413. doi:10.1007/978-0-387-45524-2_20

Ehrhardt E, Boyan G. 2020. Evidence for the cholinergic markers ChAT and vAChT in sensory cells of the developing antennal nervous system of the desert locust Schistocerca gregaria. *Invertebrate Neuroscience* **20**:1–8. doi:10.1007/s10158-020-00252-4

Ehrhardt E, Kleele T, Boyan G. 2015. A method for immunolabeling neurons in intact cuticularized insect appendages. *Dev Genes Evol* **225**:187–194. doi:10.1007/s00427-015-0499-y

Eltoum I, Fredenburgh J, Myers RB, Grizzle WE. 2001. Introduction to the theory and practice of fixation of tissues. *Journal of Histotechnology* **24**:173–190. doi:10.1179/his.2001.24.3.173

Fei J-F, Schuez M, Knapp D, Taniguchi Y, Drechsel DN, Tanaka EM. 2017. Efficient gene knockin in axolotl and its use to test the role of satellite cells in limb regeneration. *Proc Natl Acad Sci USA* **114**:12501–12506. doi:10.1073/pnas.1706855114

Felsenfeld G, Miles HT. 1967. The physical and chemical properties of nucleic acids. **36**:407–448. doi:10.1146/annurev.bi.36.070167.002203

Field LH, Matheson T. 1998. Chordotonal organs of insects. *Advances in insect physiology*.

Fields DM, Shaeffer DS, Weissburg MJ. 2002. Mechanical and neural responses from the mechanosensory hairs on the antennule of Gaussia princeps. *Marine Ecology Progress Series* **227**:173–186.

Florencio-Silva R, Sasso GRDS, Sasso-Cerri E, Simões MJ, Cerri PS. 2015. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *BioMed Research International* **2015**:421746.

Fox DT, Cohen E, Smith-Bolton R. 2020. Model systems for regeneration: Drosophila. *Development* **147**:dev173781–12. doi:10.1242/dev.173781

Frank U, Nicotra ML, Schnitzler CE. 2020. The colonial cnidarian Hydractinia. *EvoDevo* **11**:1–6. doi:10.1186/s13227-020-00151-0

Gaete M, Muñoz R, Sánchez N, Tampe R, Moreno M, Contreras EG, Lee-Liu D, Larraín J. 2012. Spinal cord regeneration in Xenopus tadpoles proceeds through activation of Sox2-positive cells. *Neural Dev* **7**:13–17. doi:10.1186/1749-8104-7-13

Gahan JM, Bradshaw B, Flici H, Frank U. 2016. The interstitial stem cells in Hydractinia and their role in regeneration. *Current Opinion in Genetics & Development* **40**:65–73. doi:10.1016/j.gde.2016.06.006

Gaj T, Sirk SJ, Shui S-L, Liu J. 2016. Genome-Editing Technologies: Principles and Applications. *Cold Spring Harbor Perspectives in Biology* a023754–21. doi:10.1101/cshperspect.a023754

Galland R, Grenci G, Aravind A, Viasnoff V, Studer V, Sibarita J-B. 2015. 3D high- and super-resolution imaging using single-objective SPIM. *Nat Meth* **12**:641–644. doi:10.1038/nmeth.3402

Gao R, Ustinov J, Pulkkinen MA, Lundin K, Korsgren O, Otonkoski T. 2003. Characterization of endocrine progenitor cells and critical factors for their differentiation in human adult pancreatic cell culture. *Diabetes* **52**:2007–2015. doi:10.2337/diabetes.52.8.2007

Garcia-Marques J, Espinosa-Medina I, Lee T. 2020. The art of lineage tracing: From worm to human. *Progress in Neurobiology* 101966. doi:10.1016/j.pneurobio.2020.101966

Gargioli C, Slack JMW. 2004. Cell lineage tracing during Xenopus tail regeneration. *Development* **131**:2669–2679. doi:10.1242/dev.01155

Garm A. 2004. Revising the definition of the crustacean seta and setal classification systems based on examinations of the mouthpart setae of seven species of decapods. *Zoological Journal of the Linnean Society* **142**:233–252.

Garm A, Watling L. 2013. The Crustecean Integument: Setae, Setules, and ther Ornamentation In: Watling L, Thiel M, editors. Functional Morphology and Diversity. pp. 167–198.

Gaspar JM. 2018. Improved peak-calling with MACS2. *bioRxiv* 496521. doi:10.1101/496521

Gebhardt JCM, Suter DM, Roy R, Zhao ZW, Chapman AR, Basu S, Maniatis T, Xie XS. 2013. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat Meth* **10**:421–426. doi:10.1038/nmeth.2411

Gehrke AR, Srivastava M. 2016. Neoblasts and the evolution of whole-body regeneration. *Current Opinion in Genetics & Development* **40**:131–137. doi:10.1016/j.gde.2016.07.009

Gilles AF, Averof M. 2014. Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *EvoDevo* **5**:43. doi:10.1186/2041-9139-5-43

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research* **17**:877–885. doi:10.1101/gr.5533506

Gladfelter AS. 2015. How nontraditional model systems can save us. *Mol Biol Cell* **26**:3687–3689. doi:10.1091/mbc.E15-06-0429

Rubin GM, Spradling AC. 1982. Genetic transformation of Drosophila with transposable element vectors. *Science* **218**:348–353. doi:10.1126/science.6289436

Gold K, Cotton JA, Stollewerk A. 2009. The role of Notch signalling and numb function in mechanosensory organ formation in the spider Cupiennius salei. *Developmental Biology* **327**:121–131. doi:10.1016/j.ydbio.2008.12.004

González-Rosa JM, Guzmán-Martínez G, Marques IJ, Sánchez-Iranzo H, Jiménez-Borreguero LJ, Mercader N. 2014. Use of Echocardiography Reveals Reestablishment of Ventricular Pumping Efficiency and Partial Ventricular Wall Motion Recovery upon Ventricular Cryoinjury in the Zebrafish. *PLoS ONE* **9**:e115604–18. doi:10.1371/journal.pone.0115604

Goss RJ. 1992. The Evolution of Regeneration - Adaptive or Inherent **159**:241–260. doi:10.1016/s0022-5193(05)80704-0

Goss RJ. 1969. Principles of Regeneration. New York: Academic Press.

Goulding SE, Lage zur P, Jarman AP. 2000. amos, a Proneural Gene for Drosophila Olfactory Sense Organs that Is Regulated by lozenge. *Neuron* **25**:69–78. doi:10.1016/S0896-6273(00)80872-7

Greiss F, Deligiannaki M, Jung C, Gaul U, Braun D. 2016. Single-Molecule Imaging in Living Drosophila Embryos with Reflected Light-Sheet Microscopy. *Biophysical Journal* **110**:939–946. doi:10.1016/j.bpj.2015.12.035

Grillo M, Konstantinides N, Averof M. 2016. Old questions, new models: unraveling complex organ regeneration with new experimental approaches. *Current Opinion in Genetics & Development* **40**:23–31. doi:10.1016/j.gde.2016.05.006

Guse GW. 1983. Ultrastructure, development, and moulting of the aesthetascs of Neomysis integer and Idotea baltica (Crustacea, Malacostraca). *Zoomorphology*.

Hallberg E, Hansson BS. 1999. Arthropod sensilla: morphology and phylogenetic considerations. *Microsc Res Tech* **47**:428–439. doi:10.1002/(SICI)1097-0029(19991215)47:6<428::AID-JEMT6>3.0.CO;2-P

Halstead MM, Kern C, Saelao P, Chanthavixay G, Wang Y, Delany ME, Zhou H, Ross PJ. 2020. Systematic alteration of ATAC-seq for profiling open chromatin in cryopreserved nuclei preparations from livestock tissues. *Sci Rep* **10**:1–12. doi:10.1038/s41598-020-61678-9

Hanaoka R, Ohmori Y, Uyemura K, Hosoya T, Hotta Y, Shirao T, Okamoto H. 2004. Zebrafish gcmb is required for pharyngeal cartilage formation **121**:1235–1247. doi:10.1016/j.mod.2004.05.011

Hartenstein V. 2005. Development of Insect Sensilla. Elsevier. pp. 379–419. doi:10.1016/B0-44-451924-6/00012-0

Hartenstein V, Posakony JW. 1990. A dual function of the Notch gene in Drosophila sensillum development. *Developmental Biology* **142**:13–30. doi:10.1016/0012-1606(90)90147-B

Hartenstein V, Younossi-Hartenstein A, Lekven A. 1994. Delamination and Division in the Drosophila Neurectoderm: Spatiotemporal Pattern, Cytoskeletal Dynamics, and Common Control by Neurogenic and Segment Polarity Genes. *Developmental Biology* **165**:480–499. doi:10.1006/dbio.1994.1269

Havemann J, Müller U, Berger J, Schwarz H, Gerberding M, Moussian B. 2008. Cuticle differentiation in the embryo of the amphipod crustacean Parhyale hawaiensis. *Cell Tissue Res* **332**:359–370. doi:10.1007/s00441-007-0571-7

Hoey T. 2001. Expression and purification of lacZ and trpE fusion proteins. *Curr Protoc Mol Biol* **Chapter 16**:Unit16.5–16.5.6. doi:10.1002/0471142727.mb1605s28

Holstein TW, Hobmayer E, Technau U. 2003. Cnidarians: An evolutionarily conserved model system for regeneration? *Dev Dyn* **226**:257–267. doi:10.1002/dvdy.10227

Hou Y, Lee HJ, Chen Y, Ge J, Osman FOI, McAdow AR, Mokalled MH, Johnson SL, Zhao G, Wang T. 2020. Cellular diversity of the regenerating caudal fin. *Sci Adv* **6**:eaba2084. doi:10.1126/sciadv.aba2084

Howat WJ, Wilson BA. 2014. Tissue fixation and the effect of molecular fixatives on downstream staining procedures. *Methods* **70**:12–19. doi:10.1016/j.ymeth.2014.01.022

Hummel T, Krukkert K, Roos J, Davis G, Klämbt C. 2000. Drosophila Futsch/22C10 Is a MAP1B-like Protein Required for Dendritic and Axonal Development. *Neuron* **26**:357–370. doi:10.1016/S0896-6273(00)81169-1

Jafar-Nejad H, Acar M, Nolo R, Lacin H, Pan H, Parkhurst SM, Bellen HJ. 2003. Senseless acts as a binary switch during sensory organ precursor selection. *Genes & Development* **17**:2966–2978. doi:10.1101/gad.1122403

Jamur MC, Oliver C. 2009. Permeabilization of Cell MembranesImmunocytochemical Methods and Protocols, Methods in Molecular Biology. Totowa, NJ: Humana Press. pp. 63–66. doi:10.1007/978-1-59745-324-0_9

Jannuzi AL, Bunch TA, Brabant MC, Miller SW, Mukai L, Zavortink M, Brower DL. 2002. Disruption of C-terminal cytoplasmic domain of betaPS integrin subunit has dominant negative properties in developing Drosophila. *Mol Biol Cell* **13**:1352–1365. doi:10.1091/mbc.01-08-0429

Janssen JM, Chen X, Liu J, Gonçalves MAFV. 2019. The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. *Molecular Therapy - Nucleic Acids* **16**:141–154. doi:10.1016/j.omtn.2019.02.009

Jarman AP, Brand M, Jan LY, Jan YN. 1993. The regulation and function of the helix-loop-helix gene, asense, in Drosophila neural precursors. *Development* **119**:19–29.

Jarman AP, Sun Y, Jan LY, Jan YN. 1995. Role of the proneural gene, atonal, in formation of Drosophila chordotonal organs and photoreceptors. *Development* **121**:2019–2030.

Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* **45**:W24–W29. doi:10.1093/nar/gkx346

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**:816–821. doi:10.1126/science.1225829

Jonkman J, Brown CM. 2015. Any Way You Slice It—A Comparison of Confocal Microscopy Techniques. *J Biomol Tech* **26**:54–65. doi:10.7171/jbt.15-2602-003

Jopling C, Boue S, Belmonte JCI. 2011. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nature Reviews Molecular Cell Biology* **12**:79–89. doi:10.1038/nrm3043

Junkunlo K, Söderhäll K, Söderhäll I. 2020. A transcription factor glial cell missing (Gcm) in the freshwater crayfish Pacifastacus leniusculus. *Dev Comp Immunol* **113**:103782. doi:10.1016/j.dci.2020.103782

Kaars C, Greenblatt S, Fourtner CR. 1984. Patterned regeneration of internal femoral structures in the cockroach, Periplaneta americana L. *J Exp Zool* **230**:141–144. doi:10.1002/jez.1402300119

Kaina B. 2003. DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling. *Biochemical Pharmacology* **66**:1547–1554. doi:10.1016/S0006-2952(03)00510-0

Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & Development* **15**:1272–1286. doi:10.1101/gad.887101

Kao D, Lai AG, Stamataki E, Rosic S, Konstantinides N, Jarvis E, Di Donfrancesco A, Pouchkina-Stancheva N, Sémon M, Grillo M, Bruce H, Kumar S, Siwanowicz I, Le A, Lemire A, Eisen MB, Extavour C, Browne WE, Wolff C, Averof M, Patel NH, Sarkies P, Pavlopoulos A, Aboobaker A. 2016. The genome of the crustacean Parhyale hawaiensis,a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife Sciences* **5**:1–45. doi:10.7554/eLife.20062

Kassmer SH, Langenbacher AD, De Tomaso AW. 2020. Integrin-alpha-6+ Candidate stem cells are responsible for whole body regeneration in the invertebrate chordate Botrylloides diegensis. *Nat Comms* **11**:1–11. doi:10.1038/s41467-020-18288-w

Kassmer SH, Nourizadeh S, De Tomaso AW. 2019. Cellular and molecular mechanisms of regeneration in colonial and solitary Ascidians. *Developmental Biology* **448**:271–278. doi:10.1016/j.ydbio.2018.11.021

Kelley ML, Strezoska Ž, He K, Vermeulen A, Smith AVB. 2016. Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing. *Journal of Biotechnology* **233**:74–83. doi:10.1016/j.jbiotec.2016.06.011

Klann M, Stollewerk A. 2017. Evolutionary variation in neural gene expression in the developing sense organs of the crustacean Daphnia magna. *Developmental Biology* **424**:50–61. doi:10.1016/j.ydbio.2017.02.011

Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nature Publishing Group* **20**:1–14. doi:10.1038/s41576-018-0089-8

Knopf F, Hammond C, Chekuru A, Kurth T, Hans S, Weber CW, Mahatma G, Fisher S, Brand M, Schulte-Merker S, Weidinger G. 2011. Bone Regenerates via Dedifferentiation of Osteoblasts in the Zebrafish Fin. *Developmental Cell* **20**:713–724. doi:10.1016/j.devcel.2011.04.014

Koerner TJ, Hill JE, Myers AM, Tzagoloff A. 1991. High-expression vectors with multiple cloning sites for construction of trpE fusion genes: pATH vectors. *Methods Enzymol* **194**:477–490. doi:10.1016/0076-6879(91)94036-c

Konstantinides N, Averof M. 2014. A common cellular basis for muscle regeneration in arthropods and vertebrates. *Science* **343**:788–791. doi:10.1126/science.1243529

Kouyama N, Shimozawa T. 1982. The Structure of a Hair Mechanoreceptor in the Antennule of Crayfish (Crustacea). *Cell Tissue Res* **226**:565–578.

Kragl M, Knapp D, Nacu E, Khattak S, Maden M, Epperlein HH, Tanaka EM. 2009. Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**:60–65. doi:10.1038/nature08152

Kvon EZ. 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* **106**:185–192. doi:10.1016/j.ygeno.2015.06.007

Kvon EZ, Stampfel G, Yáñez-Cuna JO, Dickson BJ, Stark A. 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes & Development* **26**:908–913. doi:10.1101/gad.188052.112

Lai AG, Aboobaker AA. 2018. EvoRegen in animals: Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes. *Developmental Biology* **433**:118–131. doi:10.1016/j.ydbio.2017.10.010

Lai EC, Orgogozo V. 2004. A hidden program in Drosophila peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Developmental Biology* **269**:1–17. doi:10.1016/j.ydbio.2004.01.032

Langmead B, Wilks C, Antonescu V, Charles R. 2018. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**:421–432. doi:10.1093/bioinformatics/bty648

Levine M. 2010. Transcriptional Enhancers in Animal Review Development and Evolution. *Current Biology* **20**:R754–R763. doi:10.1016/j.cub.2010.06.070

Liang XJ, Liu AQ, Lim CS, Ayi TC, Yap PH. 2007. Determining refractive index of single living cell using an integrated microchip. *Sensors and Actuators A: Physical* **133**:349–354. doi:10.1016/j.sna.2006.06.045

Littleton JT, Serano TL, Rubin GM, Ganetzky B, Chapman ER. 1999. Synaptic function modulated by changes in the ratio of synaptotagmin I and IV. *Nature* **400**:757–760. doi:10.1038/23462

Liu M, Rehman S, Tang X, Gu K, Fan Q, Chen D, Ma W. 2019. Methodologies for Improving HDR Efficiency. *Front Genet* **9**:569. doi:10.3389/fgene.2018.00691

Londono R, Wenzhong W, Wang B, Tuan RS, Lozito TP. 2017. Cartilage and Muscle Cell Fate and Origins during Lizard Tail Regeneration. *Front Bioeng Biotechnol* **5**:1–9. doi:10.3389/fbioe.2017.00070

Lyman DF, Yedvobnick B. 1995. Drosophila Notch receptor activity suppresses Hairless function during adult external sensory organ development. *Genetics* **141**:1491–1505.

Lysy PA, Weir GC, Bonner-Weir S. 2013. Making β Cells from Adult Cells Within the Pancreas. *Curr Diab Rep* **13**:695–703. doi:10.1007/s11892-013-0400-1

Manning L, Doe CQ. 2017. Immunofluorescent antibody staining of intact Drosophila larvae. *Nature Publishing Group* **12**:1–14. doi:10.1038/nprot.2016.162

Manning L, Doe CQ. 1999. Prospero distinguishes sibling cell fate without asymmetric localization in the Drosophila adult external sense organ lineage. *Development* **126**:2063.

Marques IJ, Lupi E, Mercader N. 2019. Model systems for regeneration: zebrafish. *Development* **146**:dev167692–13. doi:10.1242/dev.167692

Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, Barker CA, O'Connell LC, Patel NH. 2016. CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. **26**:14–26. doi:10.1016/j.cub.2015.11.021

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**:538–542. doi:10.1038/nbt.3190

Maruzzo D, Bortolin F. 2013. Arthropod Regeneration In: Minelli A, Boxshall G, Fusco G, editors. Arthropod Biology and Evolution. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 149–169. doi:10.1007/978-3-642-36160-9_7

Mashal RD, Koontz J, Sklar J. 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat Genet* **9**:177–183. doi:10.1038/ng0295-177

Mathews EA, Mullen GP, Manjarrez JR, Rand JB. 2015. Unusual Regulation of Splicing of the Cholinergic Locus in Caenorhabditis elegans. *Genetics* **199**:729–737. doi:10.1534/genetics.114.173765

McHedlishvili L, Epperlein HH, Telzerow A, Tanaka EM. 2007. A clonal analysis of neural progenitors during axolotl spinal cord regeneration reveals evidence for both spatially restricted and multipotent progenitors. *Development* **134**:2083–2093. doi:10.1242/dev.02852

McHedlishvili L, Mazurov V, Grassme KS, Goehler K, Robl B, Tazaki A, Roensch K, Duemmler A, Tanaka EM. 2012. Reconstitution of the central and peripheral nervous system during salamander tail regeneration. *Proc Natl Acad Sci USA* **109**:E2258–66. doi:10.1073/pnas.1116738109

Mees J, Fockedey N, Dewicke A, Janssen CR, Sorbe J-C. 1995. Aberrant individuals of Neomysis integer and other Mysidacea: Intersexuality and variable telson morphology. *Netherlands Journal of Aquatic Ecology* **29**:161–166. doi:10.1007/BF02336046

Miller SW, Avidor-Reiss T, Polyanovsky A, Posakony JW. 2009. Complex interplay of three transcription factors in controlling the tormogen differentiation program of Drosophila mechanoreceptors. *Developmental Biology* **329**:386–399. doi:10.1016/j.ydbio.2009.02.009

Mojica FJM, Díez-Villaseñor CS, García-Martínez J, Soria E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**:174–182. doi:10.1007/s00239-004-0046-3

Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ. 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Meth* **12**:982–988. doi:10.1038/nmeth.3543

Morgan TH. 1901. Regeneration. New York: The MacMillan Company.

Moulins M. 1976. Ultrastructure of chordotonal organs In: Mill PJ, editor. Structure and Function of Proprioceptors in the Invertebrates. London. pp. 387–426.

Mutesa L, Ndishimye P, Butera Y, Souopgui J, Uwineza A, Rutayisire R, Ndoricimpaye EL, Musoni E, Rujeni N, Nyatanyi T, Ntagwabira E, Semakula M, Musanabaganwa C, Nyamwasa D, Ndashimye M, Ujeneza E, Mwikarago IE, Muvunyi CM, Mazarati JB, Nsanzimana S, Turok N, Ndifon W. 2020. A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* **14**:1–5. doi:10.1038/s41586-020-2885-5

Naciff JM, Behbehani MM, Misawa H, Dedman JR. 1999. Identification and Transgenic Analysis of a Murine Promoter that Targets Cholinergic Neuron Expression. *Journal of Neurochemistry* **72**:17–28. doi:10.1046/j.1471-4159.1999.0720017.x

Neumann B, Linton C, Giordano-Santini R, Hilliard MA. 2019. Axonal fusion: An alternative and efficient mechanism of nerve repair. *Progress in Neurobiology* **173**:88–101. doi:10.1016/j.pneurobio.2018.11.004

Novák B, Bentrup FW. 1972. An electrophysiological study of regeneration in Acetabularia mediterranea. *Planta* **108**:227–244. doi:10.1007/BF00384111

Parchem RJ, Poulin F, Stuart AB, Amemiya CT, Patel NH. 2010. BAC library for the amphipod crustacean, Parhyale hawaiensis. *Genomics* **95**:261–267. doi:10.1016/j.ygeno.2010.03.005

Park J, Bae S, Kim J-S. 2015. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* **31**:4014–4016. doi:10.1093/bioinformatics/btv537

Patel NH. 1994. Imaging neuronal subsets and other cell types in whole-mount Drosophila embryos and larvae using antibody probes. *Methods Cell Biol* **44**:445–487. doi:10.1016/s0091-679x(08)60927-9

Pavlopoulos A, Averof M. 2005. Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis. *Proc Natl Acad Sci USA* **102**:7888–7893. doi:10.1073/pnas.0501101102

Pavlopoulos A, Kontarakis Z, Liubicich DM, Serano JM, Akam M, Patel NH, Averof M. 2009. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proc Natl Acad Sci USA* **106**:13897–13902. doi:10.1073/pnas.0902804106

Perry M, Kinoshita M, Saldi G, Huo L, Arikawa K, Desplan C. 2016. Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. *Nature* **535**:280–284. doi:10.1038/nature18616

Pinto-Teixeira F, Viader-Llargués O, Torres-Mejía E, Turan M, González-Gualda E, Pola-Morell L, López-Schier H. 2015. Inexhaustible hair-cell regeneration in young and aged zebrafish. *Biol Open* **4**:903–909. doi:10.1242/bio.012112

Pioro HL, Stollewerk A. 2006. The expression pattern of genes involved in early neurogenesis suggests distinct and conserved functions in the diplopod Glomeris marginata. *Dev Genes Evol* **216**:417–430. doi:10.1007/s00427-006-0078-3

Powell C, Cornblath E, Elsaeidi F, Wan J, Goldman D. 2016. Zebrafish Müller glia-derived progenitors are multipotent, exhibit proliferative biases and regenerate excess neurons. *Sci Rep* **6**:24851–10. doi:10.1038/srep24851

Power RM, Huisken J. 2017. A guide to light-sheet fluorescence microscopy for multiscale imaging. *Nat Meth* **14**:360–373. doi:10.1038/nmeth.4224

Prykhozhij SV, Rajan V, Gaston D, Berman JN. 2015. CRISPR MultiTargeter: A Web Tool to Find Common and Unique CRISPR Single Guide RNA Targets in a Set of Similar Sequences. *PLoS ONE* **10**:e0119372–18. doi:10.1371/journal.pone.0119372

Quadros RM, Miura H, Harms DW, Akatsuka H, Sato T, Aida T, Redder R, Richardson GP, Inagaki Y, Sakai D, Buckley SM, Seshacharyulu P, Batra SK, Behlke MA, Zeiner SA, Jacobi AM, Izu Y, Thoreson WB, Urness LD, Mansour SL, Ohtsuka M, Gurumurthy CB. 2017. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biology* **18**:1–15. doi:10.1186/s13059-017-1220-4

Chari R, Mali P, Moosburner M, Church GM. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Publishing Group* **12**:823–826. doi:10.1038/nmeth.3473

Ramos AP, Gustafsson O, Labert N, Salecker I, Nilsson D-E, Averof M. 2019. Analysis of the genetically tractable crustacean Parhyale hawaiensis reveals the organisation of a sensory system for low-resolution vision. *BMC Biol* **17**:1–19. doi:10.1186/s12915-019-0676-y

Reddien PW. 2018. The Cellular and Molecular Basis for Planarian Regeneration. *CELL* **175**:327–345. doi:10.1016/j.cell.2018.09.021

Reddien PW. 2013. Specialized progenitors and regeneration. *Development* **140**:951–957. doi:10.1242/dev.080499

Reddy GV, Rodrigues V. 1999. Sibling cell fate in the Drosophila adult external sense organ lineage is specified by prospero function, which is regulated by Numb and Notch. *Development* **126**:2083.

Rehm EJ, Hannibal RL, Chaw RC, Vargas-Vila MA, Patel NH. 2009. Antibody Staining of Parhyale hawaiensis Embryos. *Cold Spring Harbor Protocols* **2009**:pdb.prot5129–pdb.prot5129. doi:10.1101/pdb.prot5129

Rinkevich Y, Lindau P, Ueno H, Longaker MT, Weissman IL. 2011. Germ-layer and lineage-restricted stem/ progenitors regenerate the mouse digit tip. *Nature* **476**:1–6. doi:10.1038/nature10346

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**:24–26. doi:10.1038/nbt.1754

Rodrigo Albors A, Tazaki A, Rost F, Nowoshilow S, Chara O, Tanaka EM. 2015. Planar cell polarity-mediated induction of neural stem cell expansion during axolotl spinal cord regeneration. *eLife Sciences* **4**:664. doi:10.7554/eLife.10230

Rothstein RJ. 1989. One-Step Gene Disruption in YeastRecombinant DNA Methodology. Academic Press. pp. 331–340. doi:10.1016/B978-0-12-765560-4.50024-1

Rouet P, Smih F, Jasin M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology* **14**:8096–8106. doi:10.1128/mcb.14.12.8096

Ruiz-Gómez M, Ghysen A. 1993. The expression and role of a proneural gene, achaete, in the development of the larval nervous system of Drosophila. *The EMBO Journal* **12**:1121–1130.

Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects **24**:2033–2040. doi:10.1101/gr.177881.114

Saina M, Benton R. 2013. Visualizing olfactory receptor expression and localization in Drosophila. *Methods Mol Biol* **1003**:211–228. doi:10.1007/978-1-62703-377-0_16

Salvaterra PM, Kitamoto T. 2001. Drosophila cholinergic neurons and processes visualized with Gal4/UAS-GFP. *Brain Res Gene Expr Patterns* **1**:73–82.

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF, Peterson RT, Yeh J-RJ, Joung JK. 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Publishing Group* **8**:67–69. doi:10.1038/nmeth.1542

Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, Furlong EEM. 2007. A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes & Development* **21**:436–449. doi:10.1101/gad.1509007

Sandoval-Guzmán T, Wang H, Khattak S, Schuez M, Roensch K, Nacu E, Tazaki A, Joven A, Tanaka EM, Simon A. 2014. Fundamental Differences in Dedifferentiation and Stem Cell Recruitment during Skeletal Muscle Regeneration in Two Salamander Species. *Stem Cell* **14**:174–187. doi:10.1016/j.stem.2013.11.007

Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic Regulation of Nucleosome Positioning in the Human Genome. *CELL* **132**:887–898. doi:10.1016/j.cell.2008.02.022

Schubiger G. 1971. Regeneration, duplication and transdetermination in fragments of the leg disc of Drosophila melanogaster. *Developmental Biology* **26**:277–295. doi:10.1016/0012-1606(71)90127-8

Sehring IM, Weidinger G. 2019. Recent advancements in understanding fin regeneration in zebrafish. *WIREs Dev Biol* **9**:1321–16. doi:10.1002/wdev.367

Serano JM, Martin A, Liubicich DM, Jarvis E, Bruce HS, La K, Browne WE, Grimwood J, Patel NH. 2016. Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. *Developmental Biology* **409**:297–309. doi:10.1016/j.ydbio.2015.10.029

Singh SP, Holdway JE, Poss KD. 2012. Regeneration of Amputated Zebrafish Fin Rays from De Novo Osteoblasts. *Developmental Cell* **22**:879–886. doi:10.1016/j.devcel.2012.03.006

Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. 1985. Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature* **317**:230–234. doi:10.1038/317230a0

Somorjai IML, Somorjai RL, Garcia-Fernàndez J, Escrivà H. 2012. Vertebrate-like regeneration in the invertebrate chordate amphioxus. *Proc Natl Acad Sci USA* **109**:517–522. doi:10.1073/pnas.1100045109

Sousa S, Afonso N, Bensimon-Brito A, Fonseca M, Simoes M, Leon J, Roehl H, Cancela ML, Jacinto A. 2011. Differentiated skeletal cells contribute to blastema formation during zebrafish fin regeneration. *Development* **138**:3897–3905. doi:10.1242/dev.064717

Stewart S, Stankunas K. 2012. Limited dedifferentiation provides replacement tissue during zebrafish fin regeneration. *Developmental Biology* **365**:339–349. doi:10.1016/j.ydbio.2012.02.031

Stinson BM, Moreno AT, Walter JC, Loparo JJ. 2020. A Mechanism to Minimize Errors during Non-homologous End Joining. *Molecular Cell* **77**:1080–1091.e8. doi:10.1016/j.molcel.2019.11.018

Stollewerk A, Seyfarth E-A. 2008. Evolutionary changes in sensory precursor formation in arthropods: Embryonic development of leg sensilla in the spider Cupiennius salei. *Developmental Biology* **313**:659–673. doi:10.1016/j.ydbio.2007.11.003

Subramanian A, Wayburn B, Bunch T, Volk T. 2007. Thrombospondin-mediated adhesion is essential for the formation of the myotendinous junction in Drosophila. *Development* **134**:1269–1278. doi:10.1242/dev.000406

Sun Y, Miao N, Sun T. 2019. Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas* **156**:1–9. doi:10.1186/s41065-019-0105-9

Suryamohan K, Halfon MS. 2014. Identifying transcriptional cis-regulatory modules in animal genomes. *WIREs Dev Biol* **4**:59–84. doi:10.1002/wdev.168

Suzuki Y, Chou J, Garvey SL, Wang VR, Yanes KO. 2019. Evolution and Regulation of Limb Regeneration in ArthropodsEvo-Devo: Non-Model Species in Cell and Developmental Biology, Results and Problems in Cell Differentiation. Cham: Springer, Cham. pp. 419–454. doi:10.1007/978-3-030-23459-1_17

Südhof TC. 2002. Synaptotagmins: why so many? *Journal of Biological Chemistry* **277**:7629–7632. doi:10.1074/jbc.R100052200

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW.1983.The double-strand-break repair model for recombination.*Cell* **33**:25–35. doi:10.1016/0092-8674(83)90331-8

Takeuchi R, Lambert AR, Mak AN-S, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR, Stoddard BL. 2011. Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci USA* **108**:13077–13082. doi:10.1073/pnas.1107719108

Talchai C, Xuan S, Lin HV, Sussel L, Accili D. 2012. Pancreatic β Cell Dedifferentiation as a Mechanism of Diabetic β Cell Failure. *CELL* **150**:1223–1234. doi:10.1016/j.cell.2012.07.029

Tanaka HV, Ng NCY, Yu ZY, Casco-Robles MM, Maruo F, Tsonis PA, Chiba C. 2019. A developmentally regulated switch from stem cells to dedifferentiation for limb muscle regeneration in newts. *Nat Comms* **7**:1–8. doi:10.1038/ncomms11069

Thomas KR, Folger KR, Capecchi MR. 1986. High frequency targeting of genes to specific sites in the mammalian genome. *CELL* **44**:419–428. doi:10.1016/0092-8674(86)90463-0

Thorel F, Népote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL. 2010. Conversion of adult pancreatic α-cells to β-cells after extreme β-cell loss. *Nature* **464**:1–6. doi:10.1038/nature08894

Thorpe HM, Smith MCM. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA* **95**:5505–5510. doi:10.1073/pnas.95.10.5505

Tiozzo S, Copley RR. 2015. Reconsidering regeneration in metazoans: an evo-devo approach. *Front Ecol Evol* **3**:1–12. doi:10.3389/fevo.2015.00067

Tornini VA, Puliafito A, Slota LA, Thompson JD, Nachtrab G, Kaushik A-L, Kapsimali M, Primo L, Di Talia S, Poss KD. 2016. Live Monitoring of Blastemal Cell Contributions during Appendage Regeneration. **26**:2981–2991. doi:10.1016/j.cub.2016.08.072

Trebels B, Dippel S, Schaaf M, Balakrishnan K, Wimmer EA, Schachtner J. 2020. Adult neurogenesis in the mushroom bodies of red flour beetles (Tribolium castaneum , Herbst) is influenced by the olfactory environment. *Sci Rep* **10**:1–11. doi:10.1038/s41598-020-57639-x

Tu S, Johnson SL. 2011. Fate Restriction in the Growing and Regenerating Zebrafish Fin. *Developmental Cell* **20**:725–732. doi:10.1016/j.devcel.2011.04.013

Ulman V, Maška M, Magnusson KEG, Ronneberger O, Haubold C, Harder N, Matula P, Matula P, Svoboda D, Radojevic M, Smal I, Rohr K, Jaldén J, Blau HM, Dzyubachyk O, Lelieveldt B, Xiao P, Li Y, Cho S-Y, Dufour AC, Olivo-Marin J-C, Reyes-Aldasoro CC, Solis-Lemus JA, Bensch R, Brox T, Stegmaier J, Mikut R, Wolf S, Hamprecht FA, Esteves T, Quelhas P, Demirel Ö, Malmström L, Jug F, Tomancak P, Meijering E, Muñoz-Barrutia A, Kozubek M, Ortiz-de-Solorzano C. 2017. An objective comparison of cell-tracking algorithms. *Nat Meth* **14**:1141–1152. doi:10.1038/nmeth.4473

Ungerer P, Eriksson BJ, Stollewerk A. 2012. Unravelling the evolution of neural stem cells in arthropods: Notch signalling in neural stem cell development in the crustacean Daphnia magna. *Developmental Biology* **371**:302–311. doi:10.1016/j.ydbio.2012.08.025

Van De Bor V, Walther R, Giangrande A. 2000. Some fly sensory organs are gliogenic and require glide/gcm in a precursor that divides symmetrically and produces glial cells. *Development* **127**:3735–3743.

Vedel JP. 1985. Cuticular Mechanoreception in the Antennal Flagellum of the Rock Lobster Palinurus-Vulgaris. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* **80**:151–158. doi:10.1016/0300-9629(85)90532-8

Ventura T, Stewart MJ, Chandler JC, Rotgans B, Elizur A, Hewitt AW. 2018. Molecular aspects of eye development and regeneration in the Australian redclaw crayfish, Cherax quadricarinatus. *Aquaculture and Fisheries* 1–10. doi:10.1016/j.aaf.2018.04.001

Vicinanza C, Aquila I, Scalise M, Cristiano F, Marino F, Cianflone E, Mancuso T, Marotta P, Sacco W, Lewis FC, Couch L, Shone V, Gritti G, Torella A, Smith AJ, Terracciano CM, Britti D, Veltri P, Indolfi C, Nadal-Ginard B, Ellison-Hughes GM, Torella D. 2017. Adult cardiac stem cells are multipotent and robustly myogenic: c-kit expression is necessary but not sufficient for their identification. *Cell Death Differ* **24**:2101–2116. doi:10.1038/cdd.2017.130

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**:1–6. doi:10.1038/nature07730

Wagh DA, Rasse TM, Asan E, Hofbauer A, Schwenkert I, Dürrbeck H, Buchner S, Dabauvalle M-C, Schmidt M, Qin G, Wichmann C, Kittel R, Sigrist SJ, Buchner E. 2006. Bruchpilot, a Protein with Homology to ELKS/CAST, Is Required for Structural Integrity and Function of Synaptic Active Zones in Drosophila. *Neuron* **49**:833–844. doi:10.1016/j.neuron.2006.02.008

Wagner DE, Wang IE, Reddien PW. 2011. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* **332**:811–816. doi:10.1126/science.1203983

Watling Les. 1989. A classification system for crustacean setae based on the homology conceptFunctional Morphology of Feeding and Grooming in Crustacea. CRC Press. pp. 15–26. doi:10.1201/9781003079354-2

Whitear M. 1960. Chordotonal Organs in Crustacea. *Nature* **187**:522–523. doi:10.1038/187522a0

Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA. 2014. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes & Development* **28**:2778–2791. doi:10.1101/gad.251694.114

Wolff C, Tinevez J-Y, Pietzsch T, Stamataki E, Harich B, Guignard L, Preibisch S, Shorte S, Keller PJ, Tomancak P, Pavlopoulos A. 2018. Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb. *eLife Sciences* **7**:375. doi:10.7554/eLife.34410

Wong N, Liu W, Wang X. 2015. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biology* **16**:1–8. doi:10.1186/s13059-015-0784-0

Xu H, Xiao T, Chen C-H, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS, Brown M, Liu XS. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Research* **25**:1147–1157. doi:10.1101/gr.191452.115

Yao X, Zhang M, Wang X, Ying W, Hu X, Dai P, Meng F, Shi L, Sun Y, Yao N, Zhong W, Li Y, Wu K, Li W, Chen Z-J, Yang H. 2018. Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells. *Developmental Cell* **45**:526–536.e5. doi:10.1016/j.devcel.2018.04.021

Yeh E, Zhou L, Rudzik N, Boulianne GL. 2000. Neuralized functions cell autonomously to regulate Drosophila sense organ development **19**:4827–4837. doi:10.1093/emboj/19.17.4827

Young AP, Jackson DJ, Wyeth RC. 2020. A technical review and guide to RNA fluorescence in situ hybridization. *PeerJ* **8**:e8806. doi:10.7717/peerj.8806

Yun M. 2015. Changes in Regenerative Capacity through Lifespan. *IJMS* **16**:25392–25432. doi:10.3390/ijms161025392

Zhao Q, Ren H, Zhu D, Han Z. 2009. Stem/progenitor cells in liver injury repair and regeneration. *Biol Cell* **101**:557–571. doi:10.1042/BC20080105

Zimmer AR, Araujo PB, Bond-Buckup GB. 2009. Diversity and arrangement of the cuticular structures of Hyalella (Crustacea: Amphipoda: Dogielinotidae) and their use in taxonomy. *Zoologia*.