TÉCNICO LISBOA

Investigating Arx role in mushroom body development in *Platynereis dumerilii*

Joaquim António da Silva Melo Contradanças

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors: Dr. Detlev Arendt
Prof. Jorge Humberto Gomes Leitão

Examination Committee

Chairperson: Prof. Gabriel António Amaro Monteiro

Supervisor: Prof. Jorge Humberto Gomes Leitão

Member of the Committee: Dr. Nuno Filipe Santos Bernardes

November 2017

Acknowledgements

It was a pleasure and an honor to be part of EMBL for seven months. I am especially grateful to Dr. Detlev Arendt for the many opportunities, and to Dr. Kaia Achim and Dr. Thomas Chartier for everything they taught me and all the support in developing such a challenging and ambitious project. Thanks to all the other lab members for giving important feedback.

I want to acknowledge my thesis supervisor at Técnico, Professor Jorge Leitão, for the help in writing the thesis.

Last but not least, I want to express my gratitude towards my parents and Rita for always being present, even at a distance.

Resumo

O sistema nervoso central é partilhado pela maioria dos animais com simetria bilateral e a sua origem é uma das grandes questões em Biologia Evolutiva e do Desenvolvimento. O anelídeo *Platynereis dumerilii* é um organismo modelo simples e adequado ao estudo do ancestral comum dos animais com simetria bilateral (Urbilateria). Possui corpos de cogumelo (CC), considerados centros de integração sensorial que poderão apresentar características ancestrais.

Neste trabalho, investigou-se a função de Arx, um fator de transcrição que é especificamente expresso nos CC em desenvolvimento. Usando o sistema CRISPR-Cas9, e após testar diversos gRNAs em conjunto com proteína ou mRNA de Cas9 em diversas concentrações, duas das condições experimentais induziram mutações com tamanhos que variaram entre a eliminação de 9 pares de bases (pb) e a adição de 21 pb a jusante do homeobox de *Arx*. A taxa de mutação foi ~30%. Iniciou-se o estudo do fenótipo morfológico associado, mas os resultados não foram conclusivos. Este é um passo importante na compreensão do papel de Arx no desenvolvimento dos CC, abrindo caminho ao estabelecimento de uma linha de animais com este gene mutado.

Adicionalmente, estudou-se as respostas neuronais na cabeça de uma larva com 6 dias face à exposição a 1 mM de dopamina, fenilacetaldeído e putrescina. Os resultados obtidos sugerem que os CC são ativados por fenilacetaldeído e putrescina. Contudo, não foi possível retirar mais conclusões. Esta experiência prova que se pode testar múltiplos estímulos químicos num mesmo chip, sem que haja contaminação das outras correntes de entrada.

Palavras-chave: evolução do sistema nervosa central, *Platynereis dumerilii*, corpos de cogumelo, Arx, CRISPR-Cas9, imagiologia de cálcio

Abstract

A major research field in evolutionary developmental biology is the origin of the central nervous system, shared by most animals with a bilateral symmetry (bilaterians). For its phylogenetic position and slow-evolving characteristics, the annelid *Platynereis dumerilii* has become a simple system to unveil the last common ancestor of bilaterians (Urbilateria). It has mushroom bodies (MBs), a presumed sensory-associative brain center homologous to the vertebrate pallium that might display ancestral features.

The role of Arx, a homeodomain transcription factor specifically expressed in developing MBs, was studied. A CRISPR-Cas9-mediated knockout of *Arx* was established after trying different gRNAs with Cas9 mRNA or protein at various concentrations. Two experimental conditions induced mutations that varied between deletions of up to 9 bp and insertions of up to 21 bp around the downstream end of Arx homeobox. The estimated mutation rate was ~30%. Immunohistochemical analyses of the associated morphological phenotype were started, yet with inconclusive results. This is an important step towards understanding Arx role in MB development, opening the door to the establishment of an Arx mutant line.

Additionally, head neuronal responses to dopamine, phenylacetaldehyde and putrescine at 1 mM were studied on a single 6-day-old larva ubiquitously expressing the fluorescent calcium indicator GCaMP6s, after immobilization on a microfluidics chip, testing the hypothesis that MBs are chemosensory. MBs might be activated by phenylacetaldehyde and putrescine, but no stronger conclusions could be drawn. This experiment showed it is possible to do multiple chemical tests on a single chip without contamination of the other inflow currents.

Keywords: central nervous system evolution, *Platynereis dumerilii*, mushroom bodies, Arx, CRISPR-Cas9, calcium imaging

Contents

Acknowledgements i

Resumo iii

Abstract v

List of figures ix

List of tables xi

List of abbreviations xiii

1 Introduction 1

1.1 Evolutionary developmental biology 1

1.2 The evolution of the nervous system in Bilateria 1

1.3 *Platynereis dumerilii* as a model to study the evolution of the nervous system 2

1.4 *Platynereis dumerilii* lifecycle 4

1.5 Mushroom bodies 6

1.6 Mushroom bodies in *Platynereis dumerilii* 7

1.7 The Arx transcription factor 9

1.8 Motivation 11

2 Materials and methods 14

2.1 *Platynereis dumerilii* culture 14

2.2 PCR amplifications 14

2.3 Design of the guide RNAs 15

2.4 Production of the gRNAs, Cas9 and GCaMP6s mRNA 15

2.5 Cas9 *in vitro* cleavage assay 16

2.6 Delivery of the gRNAs, Cas9 and GCaMP6s mRNA 17

2.7 Validation of the *Arx* knockout 18

2.8 Immunostaining of larvae at 6 dpf 19

2.9 Imaging of stained larvae 20

2.10 Calcium imaging 20

3 Results and discussion 23

3.1 The knockout of the *Arx* gene 23

3.1.1 Design of the gRNAs 23

3.1.2 Cas9 *in vitro* cleavage assay 28

3.1.3 Validation of the *Arx* knockout 29
3.1.4 Genotyping with a restriction digestion 32
3.1.5 Morphological analysis of *Arx* mutants 33
3.2 Imaging of chemically-evoked neuronal responses 36
4 Conclusion and future perspectives 40
Bibliography 44
Appendix A 49
Appendix B 51

List of figures

1.1. Simplified phylogenetic tree of the main bilaterian clades 3
1.2. *Platynereis dumerilii* lifecycle 4
1.3. *Platynereis dumerilii* at 6 dpf 5
1.4. Mushroom bodies in *P. dumerilii* at 32-segment stage 7
1.5. Mushroom bodies in *P. dumerilii* at 6 dpf 8
1.6. Neuronal projection of the MBs dorsal peduncle to the head surface in *P. dumerilii* at 6 dpf....9
1.7. Expression of the Arx in *P. dumerilii* at 6 dpf 10
1.8. CRISPR-Cas9 genetic editing via non-homologous end joining and homology-directed repair....12
2.1. Schematic representation of the microinjection apparatus used to inject *P. dumerilii* zygotes....18
2.2. Chip design and principle of stimulus delivery for calcium imaging experiments 21
2.3. Stimulus protocol for calcium imaging experiments....21
2.4. Picture of the setup used for calcium imaging 22
3.1. The longest ORF in transcriptome contig #18590, predicted to be the *Arx* coding sequence....24
3.2. Possible off-targets of the crRNAs chosen....26
3.3. Target genomic loci 27
3.4. Results of the *in vitro* cleavage assay....28
3.5. Mutations in the *Arx* gene induced by the CRISPR-Cas9 system 31
3.6. Results of digesting with HpyAV the amplicons from gDNA of animals injected with the sgRNA E....32
3.7. *P. dumerilii* larvae at 6 dpf stained with DAPI and anti-acetylated tubulin antibody 35
3.8. Calcium signal in the head of a 6-day-old *P. dumerilii* larva injected with GCaMP6s mRNA....37

List of tables

3.1. crRNA sequences chosen and crRNA sequences with the highest predicted efficiencies.....24
3.2. Oligos used to produce the sgRNAs designed 27
3.3. Concentrations of the Cas9 and the gRNAs in each injection mixture tested.....29

List of abbreviations

Arx	*aristaless*-related homeobox gene
bp	base pair(s)
BSA	bovine serum albumin
C. elegans	*Caenorhabditis elegans*
Cas	CRISPR associated
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
dpf	day(s) post fertilization
DSB	double-strand break
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
Evo-devo	evolutionary developmental biology
FNSW	filtered natural sea water
fps	frame(s) per second
gDNA	genomic DNA
gRNA	guide RNA
GFP	green fluorescent protein
GS	genome scaffold
HDR	homology-directed repair
hpf	hour(s) post fertilization
HyD	hybrid detector
MB	mushroom body
mpf	minute(s) post fertilization
mRNA	messenger RNA
NHEJ	non-homologous end joining
NSW	natural sea water
nt	nucleotide
ORF	open reading frame
PCR	polymerase chain reaction
P. dumerilii	*Platynereis dumerilii*
PAM	protospacer-adjacent motif
PDMS	polydimethylsiloxane
PFA	paraformaldehyde
PMT	photomultiplier tube
RNA	ribonucleic acid
RFP	red fluorescent protein
S. pyogenes	*Streptococcus pyogenes*
sgRNA	single guide RNA
SNP	single nucleotide polymorphism
SpCas9	*Streptococcus pyogenes* Cas9
TAE	tris-acetate-EDTA
TALEN	transcription activator-like effector nucleases
TC	transcriptome contig
tracrRNA	transactivating crRNA
UTR	untranslated region

1 Introduction

1.1 Evolutionary developmental biology

Evolutionary developmental biology (evo-devo) is a multidisciplinary field that emerged in the beginning of the 1980s when methodological advances provided evidence that very distinct animals share developmental regulatory genes [1] and spatiotemporal gene expression patterns. For instance, the highly conserved Hox transcription factors are present in almost all animals (with exception of Porifera, Ctenophora, and Placozoa), where they control fundamental processes of embryonic development [2,3].

Evolutionary change does not happen by directly transforming adult ancestors into adult descendants. Developmental processes need to first be modified and then produce new phenotypes [1]. Consequently, evo-devo relates the evolution of ontogeny (origination and development of an organism [4]) with the phenotypic change during evolution, considering the environmental conditions [5,6]. It aims at linking the "genetically programmed and cyclical" development with the "non-programmed and contingent" evolution, both involving morphological changes over time [1]. For this, knowledge from embryology, morphology, genetics, developmental biology, evolutionary theory and paleontology needs to be integrated and compared [7,8].

Initially, classical morphological comparisons and embryology helped to establish phylogenetic relationships between closely related animals. Nevertheless, they easily led to ambiguous situations as homologous or homoplasic features (similar traits inherited or not inherited from a common ancestor) were difficult to discriminate and phylogenetic relationships among distantly related animals were not possible to establish [9–11].

More recently, developments such as gene cloning, visualization of gene expression in embryos, DNA sequencing and the increase in computational power have enabled comparisons of genomes, molecular developmental processes [1,5,12], gene regulatory networks [13] and cell types (defined by a unique combination of genes that specify distinct morphological and physiological features [9]), redefining the tree of life [14,15].

1.2 The evolution of the nervous system in Bilateria

Currently, one major research focus in evo-devo is the origin of the nervous system. How did the casually described as the most complex system in the universe [16] came into existence?

The nervous system is mainly composed of specialized cells (neurons) and supporting cells, known as the glia. It receives sensory input that is integrated and processed, before transmitting signals to control and coordinate effector organs. This capacity to interact with the environment is essential to maintain homeostasis [17].

During 0.6 to 1.2 billion years of metazoan evolution [16], animals have diversified into numerous forms, and so has the nervous system. Sea anemones, sponges, jellyfishes and other animals belonging to

non-bilaterian phyla (Porifera, Placozoa, Ctenophora and Cnidaria) have simple non-centralized nervous systems, organized as diffuse nerve nets, that receive sensory input and process locomotor or neurosecretory output only locally, without central integration [18,19].

In contrast, the majority of bilaterian animals have complex centralized nervous systems that integrate and process sensory information coming from the periphery, prior to initiating locomotor or neurosecretory responses [18,19]. Most of the neurons are concentrated on one side of the body and form different structures, including nuclei, ganglia, neuropils and nerve tracts [18,20].

Bilateria is a vast clade that comprises animals with a bilateral symmetry. They have anterior (head), posterior (tail), dorsal (back), ventral (belly), left and right sides, whereas non-symmetrical or radially symmetrical animals have no identifiable front or back [21]. Bilateria includes: Deuterostomia, a group containing humans, mice and fishes, among other organisms that have a dorsal nerve cord (a part of the central nervous system) [21,22]; Ecdysozoa, comprising flies, nematodes, honeybees and others; and Lophotrochozoa, to which annelids and mollusks belong to [23,24]. Together, Ecdysozoa and Lophotrochozoa form Protostomia, a group of animals that have a ventral nerve cord [21,25].

Analyses of spatiotemporal gene expression patterns support the hypothesis that deuterostomes and protostomes share a common ancestor, dubbed Urbilateria [18,19], thought to have lived in the Ediacaran Period, 600 million years ago [26–29]. Naturally, its reconstitution is essential to know how and when the central nervous system came into existence, what it looked like and how it functioned.

Fossils could contribute to direct evidence of Urbilateria, but no representative has been found or has been recognized as so. Knowing that there are no living candidates for the type of animal that Urbilateria was, phylogenetic inferences about gene content, cell types, development and morphological features have to be drawn from the comparison of extant animals representative of the different bilaterian lineages [1,30]. Ancestral features, innovations or loss of ancestral features can then be inferred.

Progress has been made with modern approaches [30–38]. However, much is still to be discovered, especially concerning the homology or convergence of the centralized nervous systems [18,39–42].

1.3 *Platynereis dumerilii* as a model to study the evolution of the nervous system

Classical model organisms are not sufficient to unveil urbilaterian characteristics since they belong either to Ecdysozoa (*Drosophila melanogaster* and *Caenorhabditis elegans*) or Deuterostomia (*Danio rerio* and *Mus musculus*), neglecting the whole Lophotrochozoa clade [43] (Figure 1.1). Moreover, the best-understood model organisms are fast-evolving and more evolutionary-derived [44,45]. Most likely, ancestral traits have been lost or significantly modified in these lineages.

To further elucidate the appearance and the evolutionary steps of the central nervous system, lophotrochozoans need to be explored. Good candidates can be found in Annelida, one phylum of this vast clade allegedly possessing ancestral or less-derived characteristics [44]. Indeed, annelid fossils from the Cambrian period have been found [46,47], suggesting that annelids may conserve some urbilaterian characteristics. In addition, these animals are particularly interesting to neurobiology because they

commonly have well-developed central nervous systems connected to a broad repertoire of sensory appendages [21].

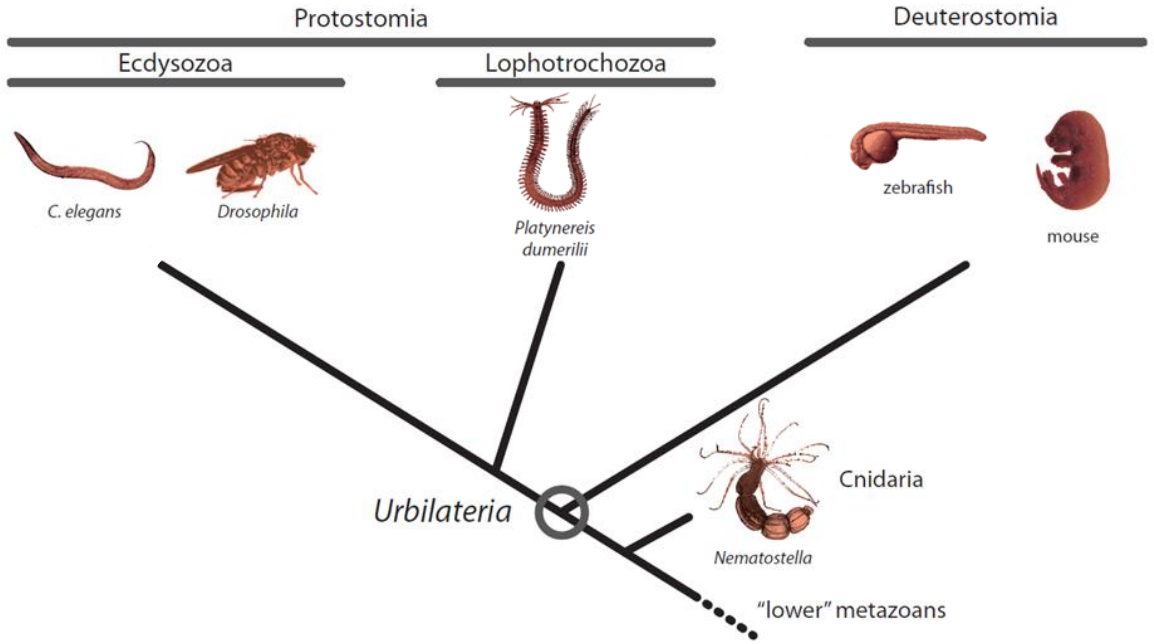


Figure 1.1. Simplified phylogenetic tree of the main bilaterian clades. Animals shown are model organisms representative of each group. Cnidaria is an outgroup of Bilateria. Adapted from [48].

Within Annelida, *Platynereis dumerilii* (*P. dumerilii*) is a recent model organism for neurobiological evo-devo studies. It is a bristle worm (polychaete) that naturally lives in sea grass beds and on macroalgae along Atlantic and Mediterranean shores [49] and that has been successfully kept and bred in lab conditions since 1953 [43]. It has a rope ladder nervous system and is especially adequate to dig into the origins of the nervous system. It has genes more conserved with their vertebrate homologues than with those of *Drosophila* and *C. elegans* (some have actually been lost from these genomes) [35], and its gene structure and sequence [34] and cell type complement [50,51] are less derived, suggesting that it keeps more ancestral features [46,52] than other model organisms. For these reasons, *P. dumerilii* is a good connecting link between the major organism models and it is adequate for a multi-level comparison with other slow-evolving animals, meaning it can contribute to shed light on key steps of the nervous system evolution [9].

Additionally, *P. dumerilii* has other characteristics relevant for neurobiological developmental studies [43]. A single female can produce up to thousands of eggs, its development is stereotypic and synchronized, and eggs, embryos and larvae are transparent until about 6 days post fertilization (dpf). Thus, *P. dumerilii* is amenable to a broad range of techniques and several hundreds of larvae at the same developmental stage can be simultaneously studied. As its central nervous system develops early on in its life, complex behaviors can be observed at stages where there are only a few hundred neurons, allowing for a cellular understanding of the system [38].

In the last years, research done with *P. dumerilii* has progressed and different techniques have been successfully employed. The genome is sequenced and assembled (unpublished data from Dr. Oleg Simakov and Dr. Thomas Larsson — EMBL, Heidelberg) and there are several transcriptomes available [53]. Microinjections in animals at the one-cell stage of development are an established delivery

method [48,54,55], essential to the development of knockouts and knockdowns with TALEN nucleases [56], morpholino-oligomers [57,58] and the CRISPR-Cas9 system [48]. First transgenic lines have been established [59]. Gene expression throughout development is often studied by doing whole mount *in situ* hybridizations [37,60,61]. RNA-sequencing protocols have been efficaciously employed and, combined with *in situ* hybridizations, have allowed the creation of whole body gene expression atlases with cellular resolution for different developmental stages [37,38]. Immunohistochemistry is routinely used to assess morphology [55,62,63]. A part of the connectome has been reconstructed with transmission electron microscopy [64,65]. Pharmacological treatments have been performed by just dissolving the drug in water, as the cuticle of the animals is permeable [66,67]. Specific behaviors such as larvae swimming and settling have also been studied [58,68]. And lastly, neuronal activity has been monitored using calcium indicators [55,64] and electrophysiology [67]. Using all these techniques, some advances have been made in the nervous system centralization [66], circadian rhythms [67], segmentation [69] and photoreception research [70], and evolutionary relationships have been established with the vertebrates [37,50,71].

1.4 *Platynereis dumerilii* lifecycle

P. dumerilii ontogeny has been descriptively detailed [43,72]. Its lifecycle comprises a pelagic and a benthic phase, illustrated in Figure 1.2. As development is temperature-dependent, all time references correspond to a development at 18 °C.

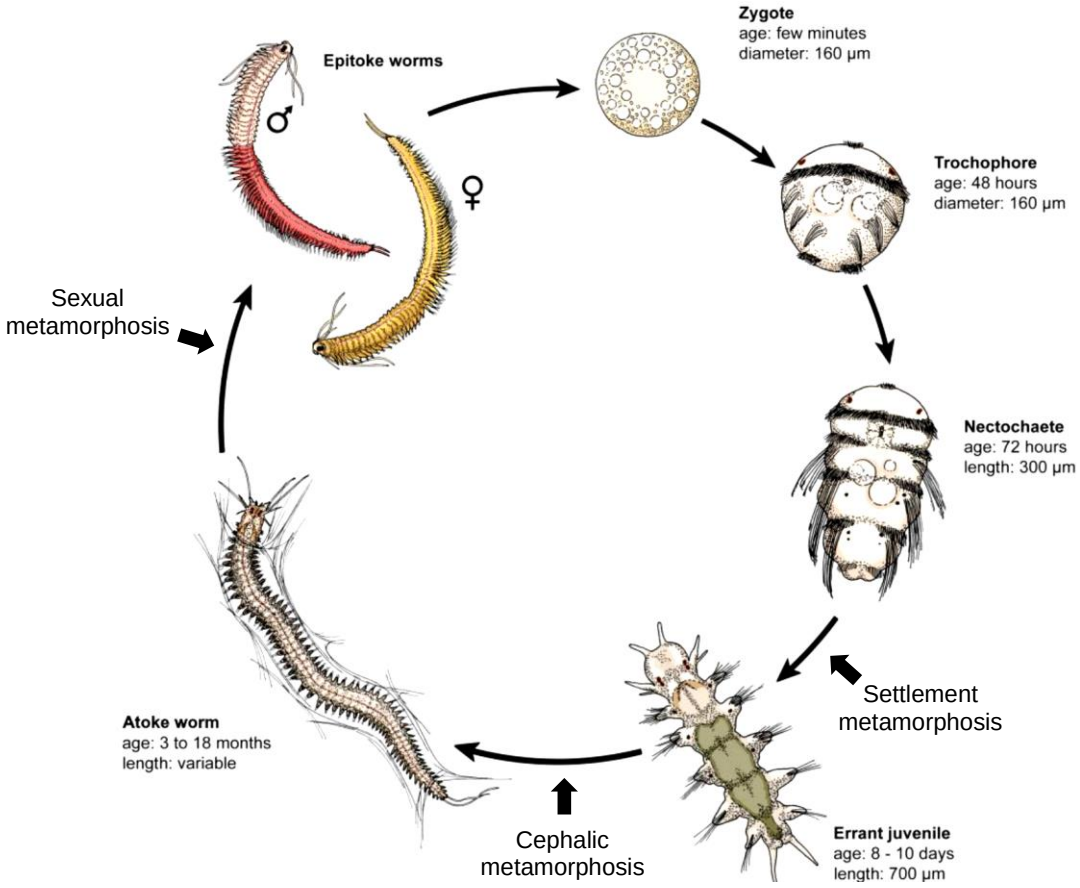


Figure 1.2. *Platynereis dumerilii* lifecycle. Time indications correspondent to a development at 18 °C. Adapted from [73].

The pelagic phase comprises three developmental larval stages: trochophore, metatrochopore (not represented in Figure 1.2) and nectochaete.

Upon fertilization, a protective jelly coat forms around the eggs, avoiding other sperm cells. At 2 hours post fertilization (hpf) the zygotes start dividing by spiralian cleavage and at 24 hpf they reach the trochophore stage. The larvae have then a spherical shape with 2000-4000 cells [48]. They have an apical organ and are part of the plankton, freely swimming in a helical fashion towards light through ciliary bands.

At around 48 hpf the metatrochophore stage starts and the larvae assume a conical shape with three segments. The lifestyle is the same as in the previous phase. Several bilateral and segmental structures rapidly develop.

After 66 hpf the larvae start to resemble adult worms, entering the nectochaete stage. There is a rapid growth of the brain, an elongation of the trunk, development of muscles, nerves and other structures (including antennae, tentacular cirri, palps and anal cirri). The lifestyle is mixed pelago-benthic, meaning the larvae alternate between swimming and crawling on the benthos with the parapodia.

After 5 dpf, the digestive tract becomes functional and the larvae begin to feed on detritus and algae. Several anatomical modifications continue to take place. Jaws grow rapidly, antennae elongate, palps forms on both sides of the mouth opening, ciliary bands start to get abolished and larva-specific structures are lost. At the same time, the central nervous system is already well-developed, including a brain and a ventral nerve cord, which is split into two parallel columns. The larvae already possess a vast number of mechano- and chemosensory structures. They resemble tiny adult worms with around 8000 cells [48] and are still transparent. Consequently, 6 dpf has emerged as a reference stage in multiple recent studies exploring whole-organism gene expression patterns [48] and chemosensation [55]. Figure 1.3 presents a larva at 6 dpf. The settlement metamorphosis initiated before is completed and the lifestyle becomes fully benthic. From then on, the development is no longer synchronous and stereotypical between animals. The animals' age is then measured by the number of body segments.

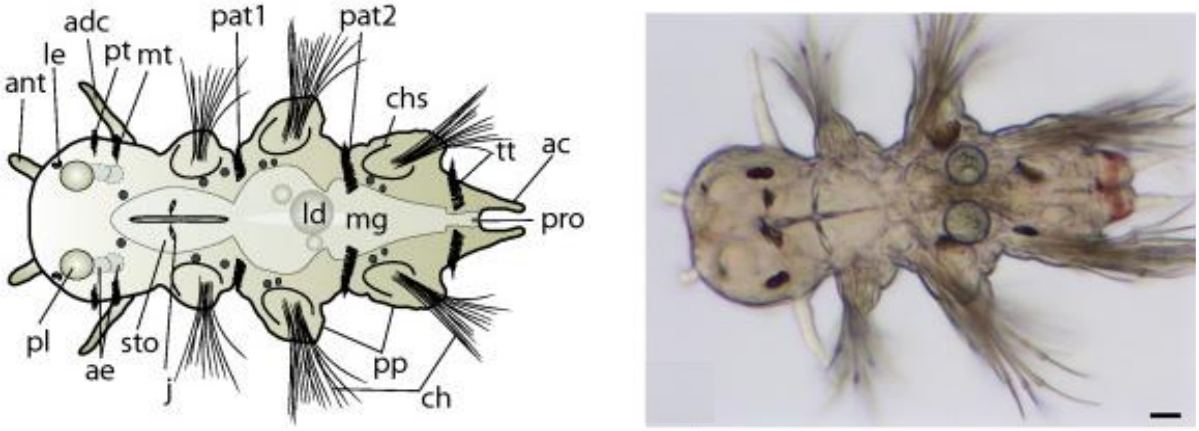


Figure 1.3. *Platynereis dumerilii* larva at 6 dpf. On the left, a scheme of a larva with several structures indicated and, on the right, a conventional light microscopy image from the dorsal side. Anterior side is on the left and posterior on the right. Scale bar indicates 20 µm. ac: anal cirri; adc: anterior dorsal cirrus; ae: adult eyes; ant: antenna; ch: chaetae; chs: chaetal sac; j: jaw; ld: lipid droplets; le: larval eye; mg: midgut; mt: metatroch; pat1 and pat2: first and second paratrochs; pl: palp; pp: parapodia; pro: proctodeum; pt: prototroch; sto: stomodeum; tt: telotroch. Adapted from [43].

During the benthic phase, larval eyes and ciliary bands disappear and the animals further develop as errant juveniles and then as asexual atokes, growing by a continuous addition of body segments at the posterior end, up to around 70. A cephalic metamorphosis takes place after the fifth segment is formed. The first pair of parapodia is transformed into part of the head and into the posterior pair of tentacular cirri. At some point, each animal builds a proteinaceous tube where it lives [74].

After 3-18 months, the worms reach up to 2-3 cm in length and undergo the last metamorphosis, becoming sexually mature epitokes. As a sexually dimorphic organism, sexes become distinguishable. Females are now yellow and males are red and white. Both maturation and spawning are synchronized by the moon cycle. The epitoke mature worms then swim to look for the opposite sex, while secreting and sensing sexual pheromones. After a nuptial dance, gametes are released into the sea water, external fertilization of the eggs occurs and the worms die [75].

1.5 Mushroom bodies

One remarkable structure in the head of protostomes is the mushroom bodies (MBs), the most-studied higher brain center in these animals [76]. Also called *corpora pedunculata*, they were first discovered in 1850 [77] and have become a major topic of research [78].

Mushroom bodies are lobed neuropils (synaptically dense regions containing a relatively low number of cell bodies) that comprise long and roughly parallel axon bundles originating from clusters of minute globuli cells located dorsally in the anterior part of the central nervous system of several protostome species [78]. Since very early, they have been thought to be involved in intelligent behavior and sensory-motor integration [77]. More recently, they have been implicated in: olfactory memory in ants [79]; short term memory in honeybees [80,81]; place memory formation in cockroaches [82]; associative learning and olfactory memory formation [83–85], regulation of sleep [86,87], control of locomotor activity [88] and decisions based on visual cues [89] in *Drosophila*. Additionally, mushroom body-like centers have been shown to share proteins required for memory formation across phyla [90].

MBs are present in many invertebrate lineages, both in Ecdysozoa and Lophotrochozoa [90], raising the question whether they share common ancestry or they are the result of convergent evolution. The latter has been supported by the lack of MBs in some protostome lineages [91] and the high diversity in the morphology and sensory inputs [92–94] of this structure. As an example, MBs of insects, spiders, onychophorans and annelids receive chemosensory input [94–96], but in spiders, honeybees and cockroaches they additionally receive visual input [97].

Furthermore, some argue that crustacean hemiellipsoid bodies (which resemble MBs but lack proper peduncles [78,90]) are not homologous to MBs [98]. Yet, some consider that hemiellipsoid bodies are highly derived MBs, as there is a common mushroom body ground pattern, including globuli-type cell bodies [90,99]. All these dissimilarities can be adaptations to control distinct behaviors across different species. Some variations have been suggested to be correlated with behavioral complexities [78,93,94].

Studies at the molecular level have tried to complement anatomical and histological investigations, paying special attention to the cellular molecular fingerprint (unique combination of genes active in a given cell type [100]) and gene regulatory networks of the MBs. In the last years, research on *P. dumerilii*

supported the hypothesis of a common ancestry. By comparing the structure and molecular fingerprint of MBs in *P. dumerilii* (lophotrochozoan) with those of insects (ecdysozoans) and the pallium (the site of learning and memory and the most highly developed part of the forebrain in mammals [101]) of vertebrates, a mushroom body/pallium sensory-associative precursor structure has been suggested to be already present in Urbilateria [37]. It would have been a simple chemosensory-integrative brain center involved in distinguishing food, integrating previous experiences into some sort of ‘‘learning’’ and coordinating locomotion. Further ahead on evolution, MBs could have diversified, becoming more histological and functional complex through independent evolution in different lineages, according to their ecological niches [37].

1.6 Mushroom bodies in *Platynereis dumerilii*

Mushroom bodies are found in many marine annelids [78] and in *P. dumerilii* they were first described in 1973 [102]. As shown in Figure 1.4, MBs are a huge paired neuronal structure, consisting of densely packed globuli-like cell bodies that form a dorsal and a ventral lobe, located anterior to the eyes, lateral to the antennal nerves and dorsal to the palps. These lobes are separated from other structures by a thin membrane and send posteriorily axon bundles known as peduncles [55,62,63].

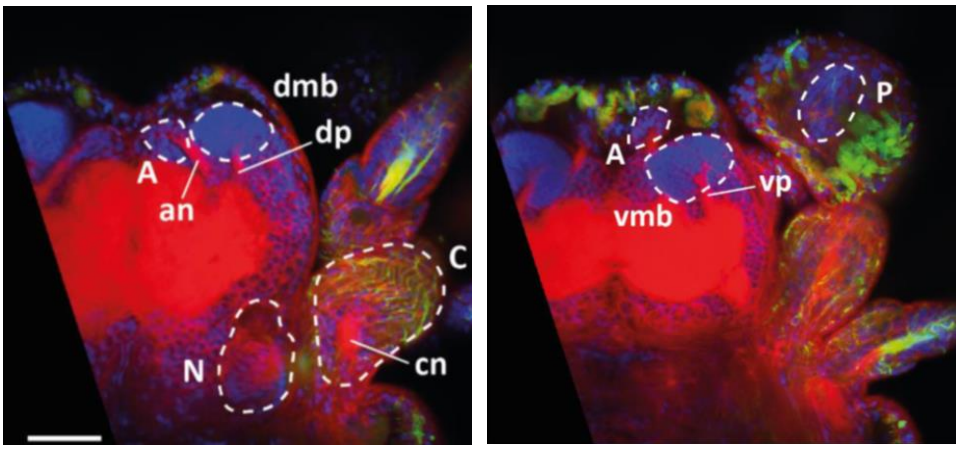


Figure 1.4. Mushroom bodies in *P. dumerilii* at 32-segment stage. The images are from two different planes of the head in dorsal view. FM 4-64 FX staining (membrane marker) is in red, anti-acetylated tubulin staining (marker of nerve fibers and cilia) is in green and DAPI (nuclear marker) is in blue. Scale bar corresponds to 50 µm. Dashed lines indicate the presumed boundaries of the areas containing the neurons of antenna (A), tentacular cirrus (C), nuchal organ (N), palp (P), and dorsal (dmb) and ventral (vmb) lobe of the MBs. Other anatomical landmarks are the antennal nerve (an), cirral nerve (cn), dorsal (dp) and ventral (vp) peduncle of the MBs. Adapted from [55].

However, in early developmental stages (animals with 3-20 segments), the MB lobes are not visible and are only recognizable by the peduncles. Moreover, 6-day-old larvae possess really thin axon bundles (less than 2 µm long and 0.15 µm wide) that are hard to spot since there are many similar brain fibers around them [62]. Thus, at these developmental stages, their identification relies on reference points more easily recognizable. As described recently, the dorsal peduncle is located laterally to the antennal nerve and is mediated towards the lateral surface, whereas the ventral peduncle is dorsal to the palpal nerve and is directed more inwardly to the medial midline of the head [55] (Figure 1.5).

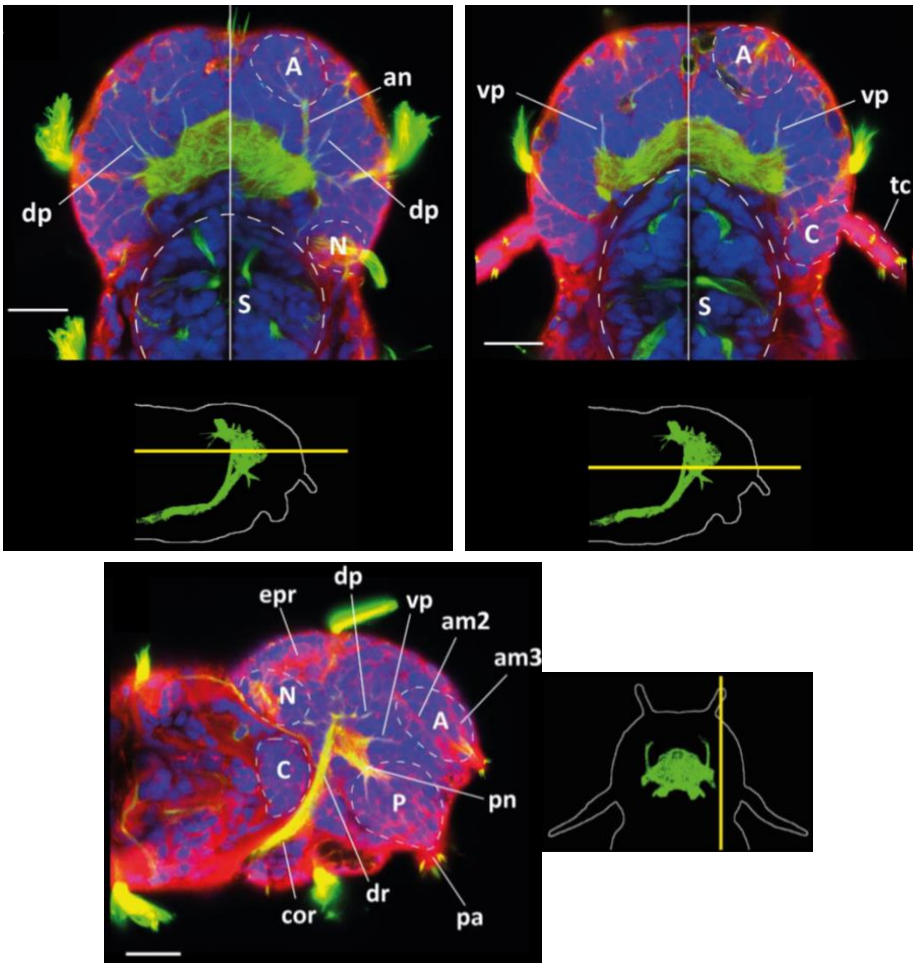


Figure 1.5. Mushroom bodies in *P. dumerilii* at 6 dpf. The images are from different planes of the head in dorsal view with anterior side up (top) and side view with anterior side on the right (bottom), as indicated by the schematic drawings where the body outline is in grey, a simplified nervous system in green and the corresponding plane in yellow. mCLING-ATTO 647N staining (membrane marker) is in red, anti-acetylated tubulin staining (marker of nerve fibers and cilia) is in green and DAPI (nuclear marker) is in blue. Scale bar corresponds to 20 µm. Dashed lines indicate the presumed boundaries of the areas containing the neurons of antenna (A), tentacular cirrus (C), nuchal organ (N), palp (P), stomodeum (S) and dorsal (dmb) and ventral (vmb) lobe of the MBs. Other anatomical landmarks are the antennal muscles (am2 and am3), antennal nerve (an), circum-oesophageal ring (cor), dorsal root of the circum-oesophageal ring (dr), eye photoreceptor cell (epr), dorsal (dp) and ventral (vp) peduncle of the MBs, palp (pa), palpal nerve (pn) and tentacular cirrus (tc). Adapted from [55].

Even though MBs functions have been investigated in many different organisms, not much is known about their role in *P. dumerilii* [55]. It is tempting to speculate that they are important to learning and memory [103,104], considering all the studies on insects mentioned before. Therefore, associative learning on *P. dumerilii* has started to be explored at 6 dpf, making use of a microfluidics chips [55].

MBs are also thought to be involved in chemosensation, the ability to detect chemical compounds. They have connections with the ventral nerve cord commissures and may receive neuronal inputs from sensory organs, suggesting that they might send signals to control the movement of the animal [63]. Once again, microfluidic devices have been used to dig into the chemosensation properties of different head structures at 6 dpf [55].

One should question whether MBs at such an early stage are already functional or have the same features as in the adult, recalling that the huge clusters of globuli-like cells are not apparent yet. Actually,

MBs have been proposed to play different roles throughout *P. dumerilii* developmental stages. 12 out of 30 6-day-old larvae analyzed in a recent study displayed neuronal projections from the dorsal peduncle to the head surface, occasionally resulting in cilia at the surface [62] (Figure 1.6). This could indicate that MBs are able to directly receive sensory inputs, but only in early developmental stages, as in the adults these projections no longer exist [62]. Such a morphological change and hypothetical loss of function would also accompany the development of globuli-like cells that constitute the MB lobes.

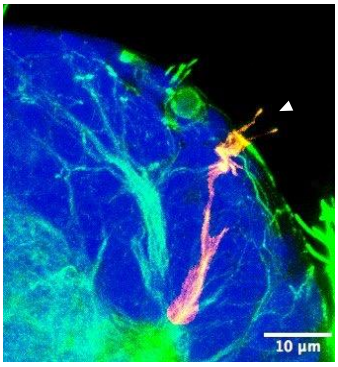


Figure 1.6. Neuronal projection of the mushroom body dorsal peduncle to the surface in *P. dumerilii* at 6 dpf. The image is from the right side of the head. The larva is in dorsal view with the anterior side up. Anti-acetylated tubulin staining (marker of nerve fibers and cilia) is in green and DAPI (nuclear marker) is in blue. The neuronal projection of the dorsal peduncle is in yellow and the cilia at the head surface are indicated by the white arrow. Adapted from [62].

Knowing that these extensions have never been reported in other species exhibiting MBs [62], much can be speculated. In the planktonic phase, these structures could be used to detect food or chemical cues indicative of danger. Indeed, at 6 dpf MBs appear to respond to some chemical stimuli [55] and putative chemoreceptor proteins are expressed in this region of the brain (unpublished results from the Arendt lab). During the benthic phase, a considerable amount of time is spent by the larvae inside a proteinaceous tube, suggesting that this characteristic would not be so relevant.

Assuming that MBs derive from a precursor structure already present in Urbilateria [105], a direct chemosensory role of the MBs could mean that not only urbilaterian MB-like structure was a sensory-associative brain center [37], but also a sensory organ. As a slow-evolving species, *P. dumerilii* would still conserve some traces of this feature. Throughout evolution, MBs would have become more specialized in integrating information and direct chemosensory capacities would have been lost in some lineages, according to each ecological niche. This hypothesis would also explain MBs diversity.

1.7 The Arx transcription factor

For the normal development of any body structure, the expression of numerous genes in a combinatorial and spatiotemporal-specific manner is required. As an example, in 10-day-old *P. dumerilii*, Emx, Pax6, Dach, Svp, Slp, Tll, Ngn, Asc and Arx are among the transcription factors expressed in the MBs, being part of the MBs molecular fingerprint [62]. Interestingly, Arx is specifically expressed where the MBs are developing at 4 dpf (unpublished data from the Arendt lab), 6 dpf [38] (Figure 1.7), 10 dpf [37,62] and 12 dpf [37]. There is still no data about later stages.

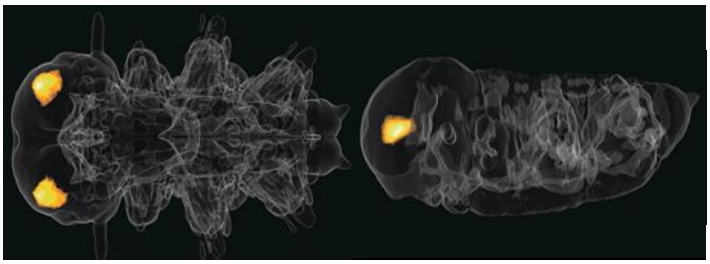


Figure 1.7. Expression of the Arx in *P. dumerilii* at 6 dpf. The images are three-dimensional representations of a larva in ventral (on the left) and lateral (on the right) position. Arx expression region is in yellow. Adapted from [38].

Arx (*aristaless*-related homeobox) is a homeodomain transcription factor [106], as it has a particular DNA-binding domain called homeodomain, encoded by a homeobox. The prefix "homeo-" comes from "homeosis", a mutational phenotype characterized by the replacement of one body part with another body part, which frequently happens when homeobox genes are mutated in animals. This was first described in *Drosophila*, where certain mutations originate flies with legs in the place of the antennae, the so-called antennapedia mutation [107].

Homeodomains are highly specific and conserved protein domains, composed of around 60 amino acids. Transcription factors containing homeodomains are responsible for regulating the expression levels of genes involved in the anatomical development (morphogenesis) of animals, plants and fungi in a temporal, spatial and tissue-specific manner [108]. In animals, homeodomain transcription factor genes are expressed since embryonic stages, where they model axial patterning and cell identity and proliferation [109,110], for instance. Naturally, their disruption leads to developmental disorders [3,111–114].

Arx has an *aristaless* domain (C-peptide) and a paired-like homeobox domain. It belongs to group II of the *aristaless*-related protein family, whose members are involved in the nervous system development [115]. In mouse, human, and even the African clawed frog (*Xenopus laevis*), Arx is involved in several processes of brain development: neuronal proliferation and migration, patterning, cell maturation and differentiation, as well as axonal outgrowth and connectivity [111,116]. Mutations in the human Arx are associated with highly pleiotropic developmental and behavioral anomalies, including severe neuronal migration defects and mental retardation without apparent brain abnormalities [106,115,117]. These can lead to X-linked lissencephaly with abnormal genitalia, infantile spasms and epilepsy [107,118,119]. Additionally, Arx expression may be important outside the brain, as in mouse it is detected in the heart, skeletal muscle, liver, pancreas and testes [106,120,121].

In *Drosophila*, Arx is involved in the development of terminal appendages (wings, legs and aristae, terminal antennal appendages required for auditory functions and hygrosensation) [120,122,123]. In *Caenorhabditis elegans* it has been suggested to function in different pathways to regulate chemosensory and motoneuron development [124–126]. Furthermore, Arx might be necessary for navigation and communication skills in honeybees [113].

All this data on Arx homologs and its specific expression in the developing MBs of *P. dumerilii* supports the hypothesis that Arx is connected to the sensory-integrative and putative chemosensory features of the MBs. However, no reports on studies about the relevance of this transcription factor in the MBs development are available at the online databases.

1.8 Motivation

One of the hot topics in evo-devo theory is the origin of the nervous system. For its characteristics, *Platynereis dumerilii* is an adequate model to study the nervous system, as well as to investigate the origin of this system. *P. dumerilii* possesses mushroom bodies, a sensory-associative brain center homologous to the vertebrate pallium [38], that might still display urbilaterian ancestral features. At early stages of development, it has neuronal projections to the head surface that might confer the ability to directly detect chemosensory stimuli, a trait further lost and never detected in other species [37].

The main objective of this thesis was to establish a CRISPR-Cas9-mediated knockout of the Arx, a transcription factor particularly interesting. It is specifically expressed in *P. dumerilii* MBs at early stages.

The CRISPR-Cas system (CRISPR stands for clustered regularly interspaced short palindromic repeats; Cas means CRISPR associated endonuclease) originates from the prokaryotic adaptive immune system [113]. It is one of the most used genome editing technologies, enabling easy, inexpensive, precise and efficient targeting [127–130].

As in many recent studies [131,132], *Streptococcus pyogenes* Cas9 (SpCas9) was used in the present work. The Cas9 endonuclease associates with a 20-nt RNA molecule (CRISPR RNA or crRNA) and a transactivating crRNA (tracrRNA), forming a complex that targets a specific DNA sequence complementary to the crRNA and adjacent to a 3-nucleotide locus downstream of the crRNA. This 3-nucleotide locus is called protospacer-adjacent motif, or PAM, and for SpCas9 it has the sequence NGG [133,134]. Importantly, crRNA and tracrRNA can be replaced by a chimeric single guide RNA (sgRNA), containing a specific 20-nt sequence to target the desired locus [135–137].

The complex formed by Cas9, crRNA and tracrRNA introduces double-strand breaks (DSBs) that can be repaired in two different ways (Figure 1.8). Non-homologous end joining (NHEJ) is more frequent and error-prone, resulting in indels (insertion or deletion mutations) at the desired genomic locus. This may lead to disruption of the target gene. Most likely there is a shift in the reading frame and nonfunctional proteins are produced. However, in-frame mutations can also take place. In case there is an exogenous donor DNA template with homology to the target site, homology-directed repair (HDR) can happen instead of NHEJ. This allows knocking in desired sequences at specific loci, for instance [138].

Although there are different ways of delivering CRISPR-Cas9 system components [134,136,137,139], in this work the delivery was done by microinjecting zygotes at the one-cell stage, soon after fertilization. Cas9 was injected as protein or as mRNA. The guide RNA (gRNA) was injected as sgRNA, or as crRNA mixed with tracrRNA. Since there was no donor DNA template, the knockout was established by NHEJ.

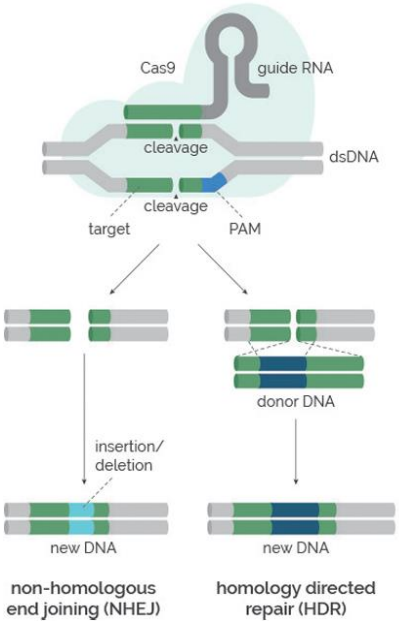


Figure 1.8. CRISPR-Cas9 genetic editing via non-homologous end joining and homology-directed repair. Adapted from ERS Genomics webpage (http://www.ersgenomics.com/crispr-cas-9-technology.php).

After generating mutants using the CRISPR-Cas9 system, a preliminary study on the overall morphology and neuronal circuitry of *Arx* mutants at 6 dpf was carried out by staining with DAPI and anti-acetylated tubulin antibody.

Immunohistochemical methods are powerful tools that take advantage of antibodies that specifically bind to antigens (for example proteins) in biological samples [135], enabling the selective imaging of antigens in cells of tissues.

In invertebrates, nerve cell axons are rich in acetylated α-tubulin. Therefore, as in many other studies [140], an anti-acetylated α-tubulin primary antibody was used to mark the axonal pathways in the nervous system, while the somata of central neurons remains unstained [62,63,141–144]. Then, a secondary antibody was added, specifically binding to the primary antibody. Since the secondary antibody was conjugated to a fluorophore, axons were visualized by fluorescent microscopy, revealing the neuronal circuits.

Furthermore, DAPI (4',6-diamidino-2-phenylindole) was applied to mark the cell bodies. DAPI is a DNA-specific fluorescent stain that binds to AT-rich regions in double-stranded DNA [145].

The whole process of staining was long and comprised multiple steps. Generically, Mg^{2+} ions from $MgCl_2$ were used as a muscle relaxant to avoid distortions of the animals [146]. Proteinase K was added to break down the cuticle of the larvae and permeabilize cells, allowing the entry of chemicals. Glycine was used to stop the digestion with proteinase K because it is an inhibitor of this enzyme. Paraformaldehyde was applied to crosslink proteins and DNA, stabilizing the morphology and preserving the specimens [147].

After investigating the morphology of *Arx* mutants at 6 dpf, head neuronal activity triggered by three chemical stimuli was briefly analyzed, exploring the putative chemosensory role of the MBs at early stages of development.

The rapid and large increase of intracellular free calcium during neuronal spiking and synaptic activity [148] can be followed with several fluorescent calcium indicators [149–151]. One of these calcium indicators is GCaMP6 [149,150,152], an ultrasensitive version of the genetically encoded calcium sensor GCaMP [153]. It is derived from the green fluorescent protein, fused with calmodulin and the M13 domain of the myosin light chain kinase. In presence of free calcium, calmodulin binds four calcium ions and M13 can bind calmodulin in turn, due to a conformational change. This form of GCaMP is fluorescent, whereas in the absence of calcium it is weakly fluorescent [154]. Genetically encoded calcium indicators are compatible with *in vivo* studies, where their fluorescence can be viewed with fluorescence microscopes. Afterwards, neuronal activity can be investigated by analyzing fluorescence intensity changes at a specific wavelength.

In this work, to investigate possible sensing functions of immature *P. dumerilii* MBs, animals were injected with GCaMP6s mRNA (a modified GCaMP6) to promote the ubiquitous expression of this calcium indicator. One larva at 6 dpf was introduced and immobilized in a microfluidics device developed for this purpose [155,156]. There, the larva was exposed to chemicals in a rigorous concentration and time-controlled manner. The transparency and small size of the larvae at 6 dpf allowed whole head imaging with a 63x microscope objective.

All these preliminary investigations can be the beginning of a vast research, knowing the conserved functions of Arx and of the MBs in several phylogenetically distant animals [106,113,115,117].

2 Materials and methods

2.1 *Platynereis dumerilii* culture

Platynereis dumerilii used in this study descended from a laboratory culture at the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, maintained as described elsewhere [72]. Maturing animals were cultured in plastic boxes with natural sea water (NSW; collected on the North Sea) and fed once a week with *Tetraselmis marina* until three months old, or spinach, fish food, *Tetraselmis marina* and *Dunaliella tertiolecta* algae from then on. To synchronize sexual maturation, the animals were kept at 18 °C on an artificial moon cycle (3 weeks of darkness, 1 week of moonlight).

Batches were set up by joining one adult female and one or two adult males in filtered natural sea water (FNSW; filtered with Millipore Express PLUS bottle top filter with a pore size of 0.22 µm) in a plastic cup. After spawning, adult worms were discarded and the eggs kept at 18 °C in an incubator (KB53, Binder), under a 16 h light - 8 h dark cycle. At 24 hpf, water was changed by fresh FNSW to remove the oocyte jelly protection layer.

2.2 PCR amplifications

PCR amplifications with several purposes were performed with Ex Taq DNA polymerase (hot start version, TaKaRa). The temperature program used was 95 °C for 3 min, followed by 36 cycles of 95 °C for 10 s, annealing temperature for 30 s and 72 °C for 30 s, and a final step of extension at 72 °C for 10 min. Each primer was added to a final concentration of 0.25 µM. Amplifications from *P. dumerilii* genomic DNA (gDNA) were performed with an annealing temperature of 62 °C, using primers according to each gRNA target locus: the gRNAs A, B and C target region was amplified with primers 5'-GGGGTTTGCAGAAGGTCATA and 5'-GGCCGTTATCCCTACAGACG, the gRNAs D and E target region was amplified with primers 5'-TCCGTTAAAAACCTGACATAATCGC and 5'-CTGAAAGCTTCCG GCTACGA. All the primers were from Sigma-Aldrich. Amplifications from plasmids pDR274 [157] and pCRII-TOPO (Invitrogen) were performed with M13 Forward (−20) and Reverse primers (5'-GTAAAACG ACGGCCAG and 5'-CAGGAAACAGCTATGAC; Invitrogen), using an annealing temperature of 57 °C.

To produce GCaMP6s, plasmid pUC57-T7-RPP2-GCaMP6 (offered by Gáspár Jékely lab) was amplified with Phusion DNA polymerase (Thermo Scientific). The temperature program was 98 °C for 1 min 30 s, followed by 35 cycles of 98 °C for 10 s and 72 °C for 45 s, and a final step of extension at 72 °C for 10 min. Each primer was added to a final concentration of 0.5 µM. The primers used were from Sigma-Aldrich and had the sequences 5'-GATCCCCCTCGGATCCTAATACGACTCACTATAGGG AGATTTGATGTTTACAGGGC and 5'-GAATTCGAGCTCGGTACCTCGCGAATGCATCTAGATCCAAT TTTCTCTTAAACAACTCC.

PCR mixes were assembled following the manufacturer protocols of each DNA polymerase. PCRs were done in a ProFlex thermal cycler (Thermo Fisher Scientific) and the amplicons obtained were analyzed by electrophoresis on a 1% (w/v) agarose gel (Sigma-Aldrich) prepared with TAE buffer. PCR

products were purified using QIAquick PCR Purification Kit (QIAGEN). Whenever the amplicons obtained were to be sequenced, their concentrations were measured in a NanoDrop 8000 spectrophotometer (Thermo Scientific) and final concentrations were adjusted with DNase-free water. Sanger sequencing was done in GATC Biotech with M13-FP primer (TGTAAAACGACGGCCAGT).

2.3 Design of the guide RNAs

Some steps had to be followed before designing gRNAs specifically targeting the *Arx* gene, considering that *P. dumerilii* genome is still a draft.

The human homeobox protein ARX was obtained from the UniProt database (uniprot.org/) and mapped against *P. dumerilii* transcriptome, using tblastn (jekely-lab.tuebingen.mpg.de). The transcriptome contig with the highest similarity to the human ARX was retrieved and the longest ORF in this transcriptome contig was found with ExPASy (expasy.org/). This ORF was aligned to NCBI database with blastp (blast.ncbi.nlm.nih.gov/) to check that the most similar sequences are other Arx protein sequences. Next, the sequence of the longest ORF in the transcriptome contig was mapped against *P. dumerilii* genome (restricted access) to find the genome scaffolds where *Arx* is located. Exons and introns were then predicted based on the genomic and transcriptomic data and Arx homeodomain was predicted with CATH (cathdb.info). Afterwards, crRNAs targeting each of the exons were identified and ranked according to their predicted efficiency with sgRNA Scorer 2.0 (crispr.med.harvard.edu/) [158], choosing the *S. pyogenes* CRISPR system, a spacer length with 20 nt, NGG as PAM sequence and PAM with 3' orientation. The choice of the crRNA sequences and the off-target analysis is explained in 3.1.1. The online tool NEBcutter V2.0 (nc2.neb.com/NEBcutter2/) was used to analyze DraI restriction sites.

To genotype the animals injected with the crRNAs used in this work, primers were designed to PCR-amplify the target genomic loci, using default settings in Primer BLAST (ncbi.nlm.nih.gov/tools/primer-blast/). The primers chosen (see 2.2) bind at least 300 nt away of the crRNAs PAM sequences in the genome of wild-type animals, considering that large deletions may happen.

Before producing the gRNAs chosen, genomic DNA (gDNA) of adult worms was extracted using the QIAamp DNA Micro Kit (QIAGEN). The target loci were amplified by PCR and sequenced to confirm their sequences. Sequencing results were analyzed with the ApE software (available at biologylabs.utah.edu/jorgensen/wayned/ape/).

2.4 Production of the gRNAs, Cas9 and GCaMP6s mRNA

To produce the sgRNAs, oligos with the sequences of the crRNAs chosen and their reverse complements were designed to be cloned into plasmid pDR274 (plasmid map in Appendix B.1), as previously described [157], and were ordered from Sigma-Aldrich. These oligos had the following sequences: TAGGTAATTCTTCTAACTGGAAGG and AAACCCTTCCAGTTAGAAGAATTA to produce the sgRNA A, TAGGTTTCTCCTTCTTCCTCCACT and AAACAGTGGAGGAAGAAGGAGAAA to

produce the sgRNA D, TAGGAGGAAGAAGGAGAAAGTG and AAACCACTTTCTCCTTCTTCCT to produce the sgRNA E. Afterwards, pDR274 was linearized with BsaI (New England Biolabs) in CutSmart Buffer (New England Biolabs) at 37 °C and purified using the QIAquick PCR Purification Kit (QIAGEN). The oligos to produce sgRNAs A, D and E were annealed in T4 DNA ligase reaction buffer (New England Biolabs), after incubating at 96 °C for 10 min and cooling down to room temperature. The annealed oligos were cloned into pDR274, using a molar ratio in the ligation mixture of 4 moles of annealed oligos to 1 mol of linearized pDR274, with T4 DNA Ligase (Promega) and Rapid Ligation Buffer (Promega), at 25 °C. One Shot TOP10 Chemically Competent *E. coli* cells (Invitrogen) were then transformed with the ligation mixture, following the longest transformation procedure proposed by the manufacturer protocol. The bacteria were spread on 30 µg/mL kanamycin plates and incubated at 37 °C overnight. Colonies were picked and used as PCR template (PCR performed with M13 Forward and Reverse primers) and part of the PCR product was sent to sequence. The remaining PCR sample was stored at −20 °C. After analyzing the sequencing results with the ApE software, amplicons with the desired sequence were cut with DraI (New England Biolabs) in CutSmart Buffer (New England Biolabs) at 37 °C and *in vitro* transcribed with MEGAscript T7 Transcription Kit (Ambion) at 37 °C, following the manufacturer protocol.

gRNAs B and C were ordered from Integrated DNA Technologies as crRNAs (TCCAGTTAGAAGAATTAGAA and AGGGTAGTGCGTCTTCTGGA, respectively), together with the tracrRNA, because the cloning of the respective oligos was not successful.

To produce the Cas9 mRNA, plasmid pMLM3613 [157] (plasmid map in Appendix B.2) was linearized with PmeI (New England Biolabs) in CutSmart Buffer (New England Biolabs) at 37 °C and purified with the QIAquick PCR Purification Kit (QIAGEN). Then it was *in vitro* transcribed and poly(A)-tailed at 37 °C with the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Ambion), following the manufacturer protocol.

The Cas9 protein was produced from plasmid pMJ915 [160] and purified at EMBL protein core facility. Appendix B.3 presents the protein sequence.

To produce the GCaMP6s mRNA, a PCR was performed from plasmid pUC57-T7-RPP2-GCaMP6 (contains the GCaMP6s ORF placed downstream of a 169-bp 5′ UTR from the *Platynereis* 60S acidic ribosomal protein P2 [159]; offered by Gáspár Jékely lab). After purification, the PCR product was *in vitro* transcribed and poly(A)-tailed at 37 °C with the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Ambion), following the manufacturer protocol.

The sgRNAs, Cas9 mRNA and GCaMP6s mRNA were purified with the RNeasy Mini Kit (QIAGEN), following the manufacturer protocol for RNA cleanup, and concentrations were measured in a NanoDrop 8000 spectrophotometer (Thermo Scientific). After purification and quantification, 2-µL aliquots were stored at −80 °C.

2.5 Cas9 *in vitro* cleavage assay

A Cas9 *in vitro* cleavage assay was performed to test the activity of the gRNAs. Following the recommendations of the PNA Bio protocol (received by email), mixtures were prepared containing the Cas9 protein at 25 ng/µL, gRNA at 0.6 pM, amplicon with the gRNAs target sites at around 135 nM (90

ng/µL of the amplicon with the target sites of the gRNAs A, B and C, or 61 ng/µL of the amplicon with the target sites of the gRNAs D and E), NEBuffer 3 (New England Biolabs), BSA (New England Biolabs) and RNase-free water. The NEBioCalculator online tool (nebiocalculator.neb.com) was used to convert mass to moles. The gRNAs were tested individually, except for one assay where the crRNAs B and C were used together, each of them at around 0.3 pM. Three assays were done with the sgRNA E. One of them contained dextran labeled with rhodamine B isothiocyanate (Sigma-Aldrich) at 1.66 µg/µL. In another assay with the sgRNA E, the incubation temperature used was 18 °C. All the other assays were done at 37 °C. The incubation period was 3 h and afterwards the Cas9 protein was heat inactivated at 65 °C for 10 min. In the end, samples and negative controls (amplicons from gDNA with the target sites) were analyzed by electrophoresis on a 1% (w/v) agarose gel (Sigma-Aldrich) prepared with TAE buffer, containing 0.06 µL of ethidium bromide (Sigma-Aldrich) per µL of agarose gel. Electrophoresis was performed using a power supply EPS 301 (Amersham Pharmacia Biotech) at 400 mA, 110 V for 35 min.

2.6 Delivery of the gRNAs, Cas9 and GCaMP6s mRNA

The delivery of Cas9, gRNAs and GCaMP6s mRNA was performed by microinjection into zygotes at the one-cell stage.

To establish the *Arx* knockout, multiple injection mixtures were tested before detecting the first mutations in the *Arx*. The injection mixtures contained different combinations and concentrations of the Cas9 protein or Cas9 mRNA, the gRNAs and dextran labeled with rhodamine B isothiocyanate (Sigma-Aldrich). The concentrations of the mixtures were adjusted with RNase-free water and dextran was added to 1.66 µg/µL. When the Cas9 protein was used, the mixture was incubated at 37 °C for 5 min to promote the assembly of the Cas9 with the gRNAs. Injection mixtures were prepared fresh when possible.

When performing calcium imaging experiments, the injection mixture was prepared fresh and contained the GCaMP6s mRNA at 1000 ng/µL and dextran labeled with rhodamine B isothiocyanate (Sigma-Aldrich) at 1.66 µg/µL. Concentrations were adjusted with RNase-free water.

The microinjection protocol used was previously described [54]. Briefly, zygotes jelly layer was removed 50 min after fertilization by rinsing with FNSW. Zygotes were treated with 47 µg/mL of proteinase K (purified from *Tritirachium album*, Merck) in FNSW for 45 s to soften the vitelline envelope. Afterwards, zygotes were rinsed once more with FNSW to stop the enzymatic reaction and 50-100 eggs were transferred onto a stage made of 2% (w/v) of agarose in FNSW, previously prepared from a mold. The remaining eggs were kept at 4 °C to delay the first cell division. Injections were carried out at 16 °C with a microinjector (FemtoJet express, Eppendorf) equipped with a micromanipulator, using an injection needle as shown in Figure 2.1 and a microscope (Axiovert 40 C, Zeiss) with a 10x objective. To pellet any precipitates and prevent the clogging of the needle during injection, injection mixtures were centrifuged (in a Prism R Refrigerated Microcentrifuge, Labnet) at 13500 rpm for 5 min prior to loading into the needle. The needles were pulled in advance in a micropipette puller (P-97 Flaming/Brown, Sutter Instrument), using glass capillaries (GC100TF-15, Harvard Apparatus). The parameter settings of the

needle puller had to be adjusted by trial and error, as these parameters varied along time, due to wear of the needle puller heating filament, temperature and humidity.

Throughout each injection session, the injection and compensation pressures were continuously adjusted (70-2000 hPA and 5-50 hPa, respectively) to inject as much volume as possible, while avoiding bursting the eggs. These pressure intensities depend on several factors, including the shape of the needle, the fluidity of the injection mix and the characteristics of each batch of embryos. The volume injected in each egg was not controlled. Each injection session stopped before the first cell division (which happens before 2 hpf) since injecting dividing embryos leads to mosaic animals. The embryos were then transferred to a 6-well plate (Nunc non-treated multidish, Thermo Scientific). The embryos injected with a higher volume were sorted out on a fluorescent microscope equipped with an RFP wavelength filter (SZX16, Olympus), as the injection mixtures contained dextran labeled with rhodamine B isothiocyanate. Afterwards, the injected embryos were incubated at 18 °C in FNSW. The animals were daily monitored and transferred to new wells with fresh FNSW.

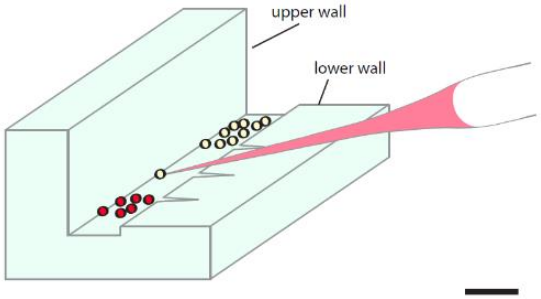


Figure 2.1. Schematic representation of the microinjection apparatus used to inject *P. dumerilii* zygotes. The stage is represented in blue, the injection mix in the needle is in pink, injected eggs are in red and non-injected eggs are in white. There is an upper and a lower wall on the stage. The latter has the same height as the zygote diameter. Scratches on the right side of the stage are manually made with the tip of a needle and are used to release the egg after injection. The scheme is in scale. Scale bar indicates 1 mm. Adapted from [54].

2.7 Validation of the *Arx* knockout

Taking advantage of dextran labeled with rhodamine B isothiocyanate) fluorescence, the most fluorescent animals were selected between 24 to 48 hpf on a fluorescent microscope equipped with an RFP wavelength filter (SZX16, Olympus), and transferred to 1.5 mL tubes. Most of the water in the tubes was removed and the animals were kept at −80 °C until extraction of gDNA. After unfreezing, gDNA of single or pooled animals was extracted with the QIAamp DNA Micro Kit (QIAGEN), following the manufacturer protocol for isolation of gDNA from tissues. Incubation with proteinase K was done overnight and 1 µg of dissolved carrier RNA was added to 200 µL of Buffer AL.

To sequence CRISPR-Cas9 target regions, gRNAs target loci were amplified by PCR and cloned into pCRII-TOPO (Invitrogen). One Shot TOP10 Chemically Competent *E. coli* cells (Invitrogen) were transformed with the ligation mixture, following the longest transformation procedure proposed by the manufacturer protocol. The bacteria were spread on 100 µg/mL ampicillin plates supplemented with 50 µg/µL of X-gal and incubated at 37 °C overnight. Only white colonies were picked and used as PCR

template (PCR performed with M13 Forward and Reverse primers). After purification, the PCR product was sent to sequence. Sequencing results were analyzed with the ApE software.

Amplicons from animals injected with sgRNA E were also run in an automated electrophoresis system (2100 Bioanalyzer, Agilent), after measuring their concentrations in a fluorometer (Qubit 2.0, Invitrogen). The same amplicons were additionally digested with HpyAV (New England Biolabs). Digestion mixtures contained 400 ng of amplicons, HpyAV in a volume ratio of 4:13 in the final mix and CutSmart Buffer (New England Biolabs). After incubation at 37 °C, the samples and the negative controls (amplicons non-digested) analyzed by electrophoresis on a 3% (w/v) Phor Agarose gel (Biozym Scientific) prepared with TAE buffer, containing 0.06 µL of ethidium bromide per µL of agarose gel. Electrophoresis was performed using a power supply EPS 301 (Amersham Pharmacia Biotech) at 400 mA and 90 V for 2 h 20 min.

2.8 Immunostaining of larvae at 6 dpf

Animals injected with the Cas9 protein at 62 ng/µL and the sgRNA E at 13 ng/µL were fixed at 6 dpf, together with negative controls (animals injected with a similar mix where the sgRNA E was replaced by RNase-free water, and wild-type non-injected animals). The animals were subsequently stained to analyze their internal neuroanatomy, adapting a protocol previously described [62].

Larvae at 6 dpf were collected with custom-made nets (100 mm mesh size), which fit into individual wells of 6-well plates. Muscles were relaxed with 3.5% of $MgCl_2$ (magnesium chloride hexahydrate; Merck) in FNSW for at least 1 min, followed by fixation in 4% of paraformaldehyde (PFA; from 16% paraformaldehyde aqueous solution, Electron Microscopy Sciences) in PTW (0.1% of Tween 20 from Sigma-Aldrich in 1x PBS) for 1 h to crosslink the proteins. Larvae were then rinsed 5x 10 min in PBS 1x and stored at 4 °C if the protocol was not continued immediately. Fixed larvae were digested with proteinase K in PTW at 100 µg/mL for 1 min to digest the cuticle of the animals and permeabilize cells. Larvae were then rinsed 2x 2 min in fresh glycine (Merck) at 2 mg/mL in PTW and rinsed 2x 2 min in PTW to stop the enzymatic reaction. Larvae were again fixed in 4% of PFA in PTW for 20 min, rinsed 5x 5 min in PTW and transferred to 1.5 mL tubes.

Fixed larvae were then used in immunostaining experiments. With this purpose, they were blocked with 5% of sheep serum in PTW with shaking for 1 h at room temperature. The primary antibody (monoclonal anti-acetylated tubulin antibody produced in mouse from Sigma-Aldrich) was then added to the blocking solution at a 1:250 volume ratio and incubation continued overnight at 4 °C with shaking. After this, the primary antibody solution was removed and the animals were rinsed 5x 5 min in PTW with shaking at room temperature. 5% of sheep serum in PTW was added with shaking for 1 h at room temperature. The secondary antibody (Alexa Fluor 488-conjugated AffiniPure Goat Anti-Mouse IgG (H+L); Jackson ImmunoResearch Laboratories) and DAPI (Sigma-Aldrich) were then added to the blocking solution at a 1:250 volume ratio and 1 µg/mL, respectively. From then on, the tubes were protected from the light with aluminum foil to prevent degradation of DAPI and of the secondary antibody fluorophore. After 2 h of incubation at room temperature with agitation, the antibody solution was

removed and the larvae were rinsed 5x 5 min in PTW at room temperature with shaking. Throughout this process, shaking was done on a thermomixer (Compact 5350, Eppendorf) at 450 rpm.

2.9 Imaging of stained larvae

Pure glycerol (AppliChem) was added to the stained larvae, shaking briefly, before mounting the larvae between a slide (Menzel microscope slide; Thermo Scientific) prepared with two layers of adhesive tape on each end and a coverslip (1871, Carl Roth). Stained larvae were imaged with a confocal microscope (Leica TCS-SP8), using a 40x glycerol immersion objective with a 1.3 numerical aperture. Detection of the Alexa Fluor 488 fluorophore was done using a 488-nm argon laser with a hybrid detector (HyD). DAPI was detected using a 405-nm diode laser with a photomultiplier tube (PMT). Image size was set to 512x512 pixels, scan frequency was 600 Hz with bidirectional scanning in the x axis and the optical/digital zoom was between 0.75-1.25. Images were processed using Fiji [161], available at https://fiji.sc/. Brightness and contrast were adjusted for each channel individually.

2.10 Calcium imaging

Calcium imaging experiments were based on a protocol extensively described elsewhere [55]. With this purpose, a 6-day-old larva ubiquitously expressing GCaMP6s was trapped in a PDMS (polydimethylsiloxane) microfluidics chip and exposed to 1-butanol (10 µM), dopamine (prepared from dopamine hydrochloride), phenylacetaldehyde and putrescine (prepared from putrescine dihydrochloride; all compounds were from Sigma-Aldrich) at 1 mM in FNSW. A scheme of the chip used is shown in Figure 2.2. It comprises one introduction channel for the animal, three inlet channels and one outlet channel. The larva is trapped in the center of the chip, usually with the ventral or dorsal side up. Once trapped, the animal has its head freely exposed.

The chip was always operated in laminar flow, with a total flow rate in the chamber constant (50 µL/min). For each stimulus tested, the protocol consisted of an initial habituation period of 30 s (in which the animal is just exposed to FNSW, panel B in Figure 2.2), six stimulation periods of 15 s (exposures to the stimulus, panel C, alternated with exposures to the control, panel D), and “resting intervals” of 20 s (situation B again). Figure 2.3 illustrates the stimulus protocol.

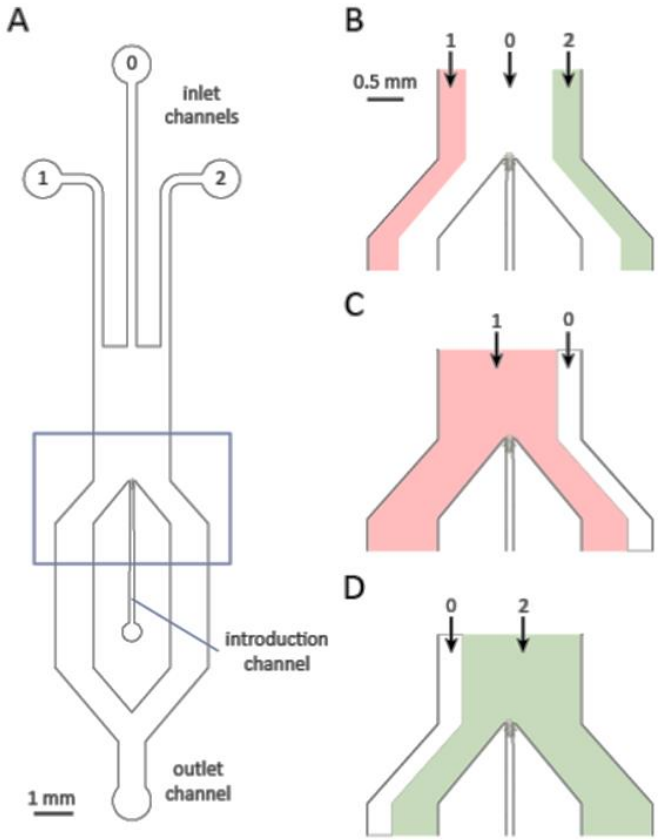


Figure 2.2. Chip design (A) and principle of stimulus delivery (B-D). The chip height was 60 µm. The introduction channel starts with a width of 150 µm that linearly decreases to reach 75 µm at its end, which constitutes the trap. The chamber upstream of the animal has a width of 2 mm, and then splits into two lateral chambers with a width of 1 mm each. The inlet channels are numbered from 0 to 2 and had FNSW, the stimulus and FNSW as a control for flow detection, respectively. Flow direction is from top to bottom. The animal is trapped at the end of the introduction channel. Figure adapted from [55].

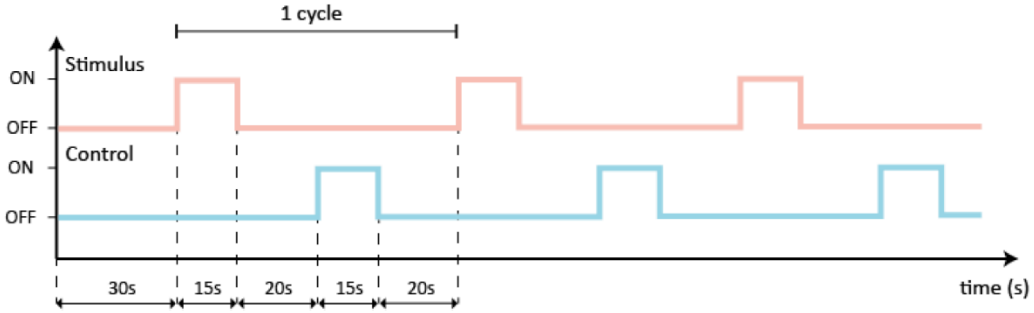


Figure 2.3. Stimulus protocol. When the stimulus/control is ON, the larva is exposed to the stimulus/control. Stimulus ON corresponds to panel C in Figure 2.2. Control ON corresponds to panel D in Figure 2.2. Stimulus and control OFF correspond to panel B in Figure 2.2. Figure adapted from [55].

An overview of the setup is presented in Figure 2.4. Each inlet channel was supplied by a dedicated 5 mL syringe (Luer Plastipak, BD). Polytetrafluoroethylene tubing (TW24, Adtech Polymer Engineering) was used to connect the chip and the needles (Microlance 3, BD) attached to the syringes. The outflow was collected in a beaker. Each syringe was operated by a separate pump (Aladdin Double Syringe Pump, AL4000-220Z, World Precision Instruments) and controlled via Micro Manager software. A customized metallic chip holder was used to support the chip.

Figure 2.4. Picture of the setup used for imaging, showing the three syringe pumps on the left side of a confocal microscope. Each pump operated a syringe connected to an inlet channel of the chip. The syringe visible on the right side was used for the introduction of the animal and the beaker for collection of the outflow. Figure adapted from [55].

During the experiments, images of the calcium signal were taken through the chip's coverslip with a confocal microscope (Leica TCS-SP8), using a 63x water immersion objective with a 1.2 numerical aperture. GCaMP6s was excited at 488 nm with an argon laser, using its main power setting at 10%. The fluorescent signal was detected with a HyD in photon-counting mode. Transmitted light images were recorded with a PMT, from the same excitation light as GCaMP6s. Images of 12 planes were acquired at 1 fps. For each plane, the xy spatial resolution was 504x504 pixels with a pixel size of 0.3 µm. The z spatial resolution between two consecutive confocal planes was less than 5 µm. The laser scanning speed was 8000 Hz (resonant mode) with a phase X correction of 1.32, a pinhole opening of 6.4 airy units and an optical plane thickness with pinhole wide open of 5 µm. Subsequently, images were processed using Fiji software. Brightness and contrast were adjusted and mean fluorescence intensity was determined with z project tool.

3 Results and discussion

3.1 Knockout of the *Arx* gene

To knockout the *Arx* gene, a strategy based on the use of CRISPR-Cas9 system was developed, involving the design, production and testing of gRNAs.

3.1.1 Design of the gRNAs

Some steps had to be followed before designing gRNAs that specifically target the *Arx* gene, considering that *P. dumerilii* genome is still a draft.

The protein sequence of the well annotated human homeodomain protein ARX (Appendix A.1) was retrieved from the UniProt database and mapped against the *P. dumerilii* transcriptome available at the moment in http://jekely-lab.tuebingen.mpg.de. Transcriptome contig #18590 (TC 18590; Appendix A.2) had the highest score (149.83) and an e-value of 6.25×10^{-37}. Using ExPASy, the contig sequence was translated into the amino acid sequence and the longest ORF was found (Appendix A.2 and Figure 3.1). These bioinformatics procedures led to identify the putative *P. dumerilii Arx* as being located between nucleotides 160-1626 of TC 18590. The most similar sequences to the putative *P. dumerilii* Arx protein (Appendix A.3) are Arx proteins from other organisms, based on BLAST results (data not shown). To know *Arx* gene sequence, the sequence correspondent to the longest ORF in TC 18590 (nucleotides 160-1626 of TC 18590) was mapped against *P. dumerilii* genome. Score and e-value parameters indicated that *Arx* is located in genome scaffolds #57705 (GS 57705; score of 1380 and e-value of 0.0) and #522890 (GS 522890; score of 452 and e-value of 10^{-124}).

By analyzing the genome scaffolds and the longest ORF in TC 18590, the coding regions (exons) and non-coding regions (introns) of *Arx* were predicted. Continuous sequences in the genome that matched the ORF in the transcriptome were classified as exons. The DNA segments between these exons were considered to be introns, assuming the transcriptome contains the whole *Arx* mRNA and does not have any intron. Four different exons could be identified (Figure 3.1).

As the last step before designing the crRNAs, the Arx homeodomain was identified using CATH database. The 3cmyA00 domain (Appendix A.4) had the smallest e-value (5.2×10^{-22}). This domain is the homeodomain of the human Pax-3, a paired box homeodomain protein, and corresponds to nucleotides 595-771 of the longest ORF in TC 18590. Appendix A.5 presents the alignment of the human Pax-3 homeodomain with the predicted Arx homeodomain.

To design the crRNAs, a ranked list with all possible crRNAs targeting the coding regions was obtained with sgRNA Scorer 2.0, choosing the most studied CRISPR system (from *S. pyogenes*), its appropriate spacer length, PAM sequence and PAM orientation (20, NGG and 3', respectively). Each coding region was individually analyzed with this software since single exons should be targeted rather than introns or segments corresponding simultaneously to exons and introns. Mutations in introns are

in most cases removed from the mRNA due to alternative splicing and are not likely to produce an NHEJ knockout.

crRNA sequences obtained with sgRNA Scorer 2.0 were sorted from largest to smallest predicted efficiencies, as indicated by the computationally calculated score. Table 3.1 contains the most relevant crRNA sequences.

Figure 3.1. The longest ORF in transcriptome contig #18590, predicted to be the *Arx* coding sequence. Each color corresponds to a different exon. Bigger letters correspond to the homeobox. Targets of the crRNAs are in bold with the PAM sequences underlined.

Table 3.1. crRNA sequences chosen (indicated by letters on the left side) and crRNA sequences with the highest predicted efficiencies, according to sgRNA Scorer 2.0. Target sequence column contains the crRNA sequences plus the PAMs (underlined). The order indicates the position of each sequence in the overall ranking of the score values, containing 233 elements. Strand and position indicate the DNA strand and the place in the longest ORF of TC 18590 where the sequence is located.

	Order	Target sequence (crRNA plus PAM)	Score	Strand	Position
	1	CACAGGGGGTGAGTGGGGGGAGG	1.676546645	−	1274 - 1296
	2	TCCCCGGGGGAGAAGTAGGGGGG	1.614119516	−	132 - 154
	3	GCATCCGGGGAGTCACGGAGCGG	1.411978736	+	1035 - 1057
	4	GGGGGAAGGGGGATGCTCCGGGG	1.375873747	−	203 - 225
	5	TAACACCACAAAATTTACGGCGG	1.283079052	+	1217 - 1239
E	6	GGAGGAAGAAGGAGAAAGTGGGG	1.241670897	+	761 - 783
	7	TTTGGTGGTCCAGGTATGGGTGG	1.18617048	+	931 - 953
	8	TGCTGAGGGTACAACACAGGGGG	1.132491492	−	1288 - 1310
	9	GGTACAACACAGGGGGTGAGTGG	1.082498228	−	1281 - 1303
	10	ATGTCGAATCCCGTCCCCGGTGG	1.076014333	+	1099 - 1121
	11	CTCCCCGGGGGAGAAGTAGGGGG	1.038317181	−	133 - 155
	12	GCCCCCCTACTTCTCCCCCGGGG	1.036330201	+	131 - 153
	13	GTGGAGGAAGAAGGAGAAAGTGG	1.001614063	+	759 - 781
	14	GTTAGGTTCGGCGAACCCTGGGG	0.993728154	−	1198 - 1220
	15	GAGAGTGAGGTTGTCCACCGGGG	0.96941855	−	1113 - 1135
	16	GTCATGAGCCAGATGCATTGTGG	0.947408025	−	263 - 285
	17	CGAGGTTGGAACTTGCACAGGGG	0.925743588	−	1154 - 1176
A	18	TAATTCTTCTAACTGGAAGGAGG	0.916494487	−	620 - 642
B	96	TCCAGTTAGAAGAATTAGAAAGG	0.023795579	+	626 - 648
C	176	AGGGTAGTGCGTCTTCTGGAAGG	-0.581274328	−	650 - 672
D	200	TTTCTCCTTCTTCCTCCACTTGG	-0.913548612	−	755 - 777

The last crRNA to be designed was chosen based on some considerations.

First, crRNAs can target either the coding or the complementary DNA strand (Cas9 creates DSBs), but their orientation is important, as most of the mutations occur at the gRNA-DNA binding region [157].

Furthermore, one must opt to target either an upstream region of the gene (and upstream of the homeobox, in this case) or to directly target a functionally important domain. The first option increases the probability of a completely non-functional protein if a frameshift mutation occurs. In this work we chose to directly target the homeobox because even non-frameshift indels are likely to alter protein function when they occur in essential protein domains [162]. Thinking of a transcription factor such as Arx, a non-frameshift mutation far from the homeodomain in principle does not prevent the homeodomain to bind to DNA. In addition to this, Arx homeodomain is roughly in the middle of the gene. Targeting this sequence and not upstream of it reduces the probability of transcription of the gene from alternative start sites. This choice is also supported by recent investigations, that used a similar strategy, resulting in a higher proportion of loss-of-function mutations [162].

Lastly, if a sgRNA is to be produced by cloning into an appropriate plasmid and subsequent transcription, constrains related to restriction digestions done before transcription and compatibility with the promoter used for the transcription must be considered. In this work, following an approach described elsewhere [157], the crRNA sequence had to start with GG to be compatible with the T7 promoter and could not originate any additional DraI restriction sites after being cloned, knowing that this enzyme is used to prepare the template for the transcription.

All the five crRNAs predicted to be the most efficient (Table 3.1) target sequences far from the homeobox (correspondent to nucleotides 595-771 of the longest ORF in TC 18590). crRNAs #2 and #4 target sequences upstream of the homeobox and only the crRNA #4 starts with GG. This one could be an option, but it was not selected because it does not target directly the homeobox. Therefore, the crRNA #6 was chosen. This crRNA targets the downstream part of the homeobox and its orientation indicates that mutations most likely will affect this domain (Figure 3.1). It also starts with GG, making it appropriate for the production method chosen. None of the subsequent crRNAs seems to be a better candidate, either for the same reasons discussed before, or for the lower scores.

Before choosing this crRNA, four other crRNAs had been designed and tested, as it will be discussed further ahead in this thesis. From here on, all crRNA sequences, correspondent sgRNAs and target sequences will be designated by a letter, corresponding to the chronological order of their conception (Table 3.1). Thus, crRNA #6 will be called crRNA E.

crRNAs A, B, C and D (crRNAs A-D) were designed following a similar procedure. However, these crRNAs were chosen from an unsorted list containing all the possible crRNAs. The list was used in the form as computationally produced by sgRNA Scorer 2.0. We considered that the crRNAs were already ranked according to their efficiency and that the score value was an indication of off-target likelihood (similarly to the e-value). Actually, this last assumption does not make sense. The program did not have access to the whole *P. dumerilii* genome to be able to predict off-targets. Due to this mistake, crRNAs B, C and D have considerably low predicted efficiencies. Luckily, crRNA A does not have such a low score (Table 3.1) and it is the second-best candidate, as deduced by its score and target region.

Additionally, the imposition of choosing crRNAs starting with GG was not considered when selecting

crRNAs A-D. The oligos used to clone their sequences in plasmid pDR274 were adapted by adding an extra GG at their 3', assuming that an extra GG would not affect the assembly of the sgRNAs with the Cas9 and the cleavage efficiency. Indeed, some studies indicate that adding one or two Gs still leads to sgRNAs capable of inducing DSBs [163,164].

Finally, crRNAs A-D target regions within *Arx* homeobox (Figure 3.1). crRNAs A, B and C target the same exon, which contains the upstream part of the homeobox. crRNA D targets the same exon as crRNA E, which contains the downstream part of the homeobox. Table 3.1 summarizes all the information about the crRNAs selected to use in this work.

The next step was to check if any of the crRNAs was predicted to have off-targets. The sequences shown in Table 3.1 contain the crRNA sequence plus the PAM sequence expected to be present in the genome. To check for off-targets, each crRNA sequence plus a generic NGG PAM was mapped against *P. dumerilii* genome, bearing in mind that there are many single nucleotide polymorphisms (SNPs) in the DNA of this organism (unpublished results from the Arendt lab). This was repeated considering a NAG PAM, as recent works have shown that SpCas9 also cleaves target sites followed by this PAM, even though with just 20% of the efficiency for target sites with NGG PAMs [137,165]. The results were then analyzed considering that SpCas9 can still cleave DNA when there are some mismatches in the gRNA-DNA binding region, depending on the quantity, position, distribution and even identity of mismatching bases [165]. Generally speaking, mismatches at a distance to the PAM are more easily tolerated than mismatches within 10-12 bp of the PAM [137,165,166]. Therefore, sequences similar to the chosen crRNAs were ruled out as possible off-targets if they were not followed by either an NGG or NAG PAM, if there was a total of four or more mismatches or if at least two mismatches lied within 12 bp of the PAM sequence [165].

In principle, crRNAs A and C do not have off-targets in *P. dumerilii* genome. As shown in Figure 3.2, the crRNA B might have one off-target, but it may not be relevant. There are two mismatches, one of them just 4 bp away of a NAG PAM. Possible off-targets of the crRNA E are similar situations: two or three mismatches and a NAG PAM. Yet, these mismatches are slightly far from the PAM sequence, compared with the off-target of the crRNA B. The crRNA D may have a more serious off-target. There is a NGG PAM and two mismatches far from the PAM. The closest mismatch to the PAM is 9 bp away. Nevertheless, all of these crRNAs were used in this work. Other crRNAs gave comparable or worse results (results not shown). In addition, there is always a certain degree of uncertainty due to *P. dumerilii* being a draft genome.

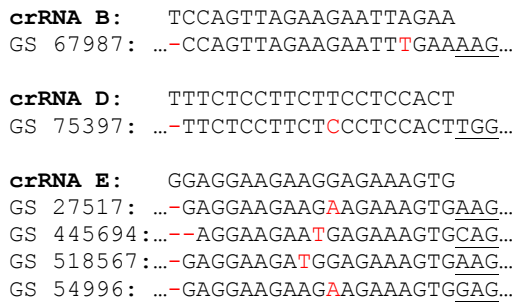


Figure 3.2. Possible off-targets of the crRNAs chosen. The genome scaffolds (GS) where the off-targets are located are indicated. The PAM sequences are underlined and mismatches are in red.

In this work, following an approach described elsewhere to produce sgRNAs [157], oligos to clone the chosen crRNAs sequences into BsaI-linearized pDR274 (Appendix B.1) were determined. Table 3.2 presents the oligos necessary to produce the sgRNAs designed, taking into account that the BsaI restriction sequence is GGTCTC(1/5). Each pair of oligos was annealed before ligation.

Table 3.2. Oligos used to produce the sgRNAs designed. The sequences correspondent to the crRNAs are in bold. The underlined nucleotides were added to the crRNA sequences to make transcription compatible with the T7 promoter of plasmid pDR274. Nucleotides not in bold are required to clone into BsaI-linearized pDR274.

crRNA	Oligos
A	TAGG**TAATTCTTCTAACTGGAAGG**
	AAAC**CCTTCCAGTTAGAAGAATTA**
B	TAGG**TCCAGTTAGAAGAATTAGAA**
	AAAC**TTCTAATTCTTCTAACTGGA**
C	TAGG**AGGGTAGTGCGTCTTCTGGA**
	AAAC**TCCAGAAGACGCACTACCCT**
D	TAGG**TTTCTCCTTCTTCCTCCACT**
	AAAC**AGTGGAGGAAGAAGGAGAAA**
E	TA**GGAGGAAGAAGGAGAAAGTG**
	AAAC**CACTTTCTCCTTCTTCCT**

Prior to transcribing the sgRNAs, DraI was used to prepare the template. It is easy to verify that no additional DraI restriction sites appear after cloning the oligos from Table 3.2 into pDR274, since DraI restriction sequence is TTT/AAA.

After determining the oligos required to produce the desired sgRNAs, primers were designed to amplify the target genomic loci by PCR. These amplifications are important to validate the genetic modifications caused by the sgRNAs. Figure 3.3 presents the loci amplified with these primers. There are more than 300 nt at 3' of each PAM sequence to allow genotyping of large deletions.

Before ordering the oligos indicated in Table 3.2, gDNA of adult worms was extracted and the target loci were amplified and sequenced. The sequencing results revealed that the target loci sequences considered to design the crRNAs were correct.

GS 522890: 96-1172 **1077 bp**

GGGGTTTGCAGAAGGTCATATTATTATGCAAATATTTTGAAGACAATATAAAATTAAAAGGTTTTGTCTATTTATCTTTGCAGATGACCAATACAAAAAAC
TGATGTCACCCACAATGCATCTGGCTCATGACGTCACAGCTCGAATCATTCTAGCCAATCACAGCCAAGAGTTGCAGTCACGTGATTATGCTTATGACCAATCACAAAG
GAGGCAAAGTGTGATGTCATCAGACAGTGCTGCCACCAGCCTGAGAAATGATGATATAGATGATGATGAACAGTTATGTATTGTGGATGATGATGAACCACTAAGTCCT
GTAAATACCAGCCCAGTGCAGCCCTCGGCCTCGATCCAGGCACCGAAAGAAGTCAAAGTCCAGGAAGACAACTTATCGATCTCCGATGGCGATGGCGACTTGGATGA
AATGGGAAAACGCAAACAAAGACGATATCGAACAACGTTCA**CCTCCTTCCAGTTAGAAGAATTAGAAAGG**G**CCTTCC**
AGAAGACGCACTACCCTGATGTGTTCACTAGGTATGTACCTAAAAAGTGTGTCTATGCCTGTGCATGTAGTAAATATATCCTTTTTAGTCGTTA
AAAGGACATTATTAGGGCACCAAAAAAACTTGGTCCCGTTAAATATCCAGACGGGTGTAGGGGGACCCCTCTGATCTACCCGACTATTTCTGTTGAAAAAATGCTACTA
TTTTGTTGAGTGTTAGTTCTTAGTTCAATATTTGAACCATTGGTCCATTAGAAAGCATTCATCATTACTTTTTTGAAGCAGCTTCACCAATTTGAAGCCATGTAGAAGATCC
GCCTAAATTTCTCAAAGATATTTTGGAATATTTCTCTGCAGCCAGAATGATGACTTACACGTCAGTTTGATCTGTAATTTGTGAAAATACAAAATCAGAATCGCTAACTGA
CTGGCTGAAGATACATTGCTCTAAATGATAGTATTTCTGTAAAAAAAATCCTATACATTGTAAATGAGAAAATTATTTGAAGACATTTACATTTCCTTCCATCTTTACATAG
AGATGGTGATTTCATA**CGTCTGTAGGGATAACGGCC**

GS 57705: 272272-28003 **731 bp**

TCCGTTAAAAACCTGACATAATCGCTTTAATCTCTCATACTCGTTCAATGATTGTAACCTTTTGTGGTGCCTATCTGTTTATATGGTTAAATAATCTTT
CTATAAATGTGAAGTTTTTTTTTTTAACAAGTGTAGTTTTAGTAGCCGTAACTTTTTGATGACTTATGTGTCGCCGACATTACGCGAGTCATTGAAACGGAAACAGAAAGAA
TCAAATCAAATCGATCAGCCATATCTTTACCAAATTGGTGCAGAAATACTGGATCATATATACCTTTCATATGATATAACATTACTATAGGCTTCTTATGCAAAGTCCATGT
GACACAAACTAATAAAATATTAATACTTATAAATATGAATTTTCTCTTGCTGTTTTCAGGTGTGGTTTCAAAACAGAAGAG**CCAAGTGGAG**
GAAGAAGGAGAAAGTGGGGCCACAATCGCATCCCTACGGCCCTTTTGGGGGCGGCCCTCCGACCGGGCCATTGGGCATCCCAGGGTTGGGCCAAAGACCCC
TCGGCCCACACACCACTTACACAGACTTGTTGATCAAATCTTACGAAAACCATCTCGCTCAGAAGTTTGGTGGTCCAGGTATGGGTGGATTCTATCCACCAGGATTAGG
GGGACCGGTCGCTGCCATGGGAGTCTTTCCCGGATTGCCCTATTCCGGATTGGGTCATCTGCATCCGGGGAGTCACGGAGCGGCCCAG**TCGTAGCCGGAA**
GCTTTCAG

Figure 3.3. Target genomic loci. The binding sites of the primers used to amplify by PCR the target loci are in bold and underlined. Non-coding regions of the genome are in black. The two exons targeted by the crRNAs designed are in red and yellow. Bigger letters not underlined correspond to the homeobox. The targets of the chosen crRNAs are in bold and not underlined.

3.1.2 Cas9 *in vitro* cleavage assay

From the five gRNAs designed, three of them were successfully produced in the lab as sgRNAs (A, D and E), while the other two were ordered from a company as crRNAs, since their cloning failed.

Before *in vivo* testing, an *in vitro* cleavage assay was carried out to test the activity of each gRNA. This step can identify gRNAs not efficient enough to produce DSBs in DNA and may save time. *In vivo* tests are time-consuming and demanding.

Amplicons from gDNA of wild-type animals (Figure 3.3) were used as template and were incubated with the Cas9 protein and the gRNAs. The resulting DNA fragments were separated according to their sizes by electrophoresis (Figure 3.4). When the Cas9 is able to cut the template, the band of the original amplicon disappears (at least in part) and yields two other bands of lower molecular weight.

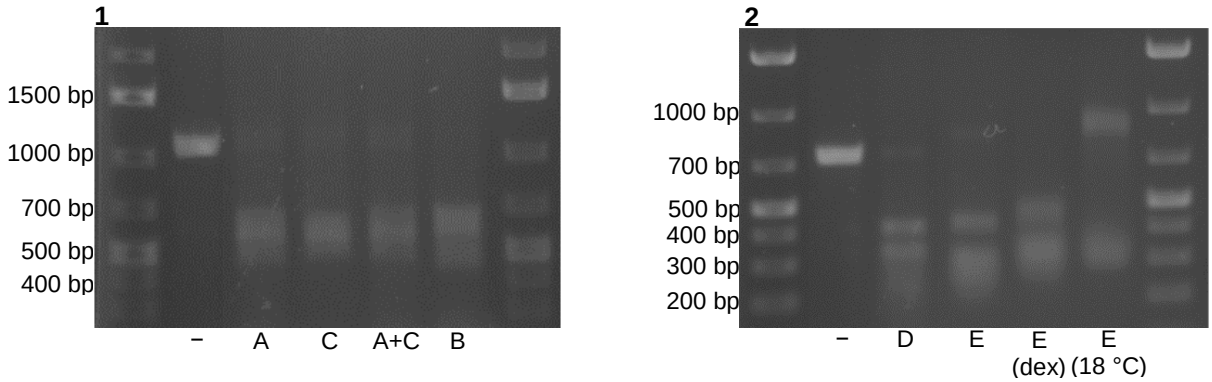


Figure 3.4. *In vitro* cleavage assay results. The left and right columns of each picture contain the DNA ladder with the length of the bands indicated. Each picture corresponds to a different DNA template, as seen by the different lengths of the bands of the controls lacking the Cas9 protein (−). Letters indicate the gRNAs tested. A+C is a simultaneous test of gRNAs A and C at an equimolar concentration in the same mixture. (dex) means that dextran labeled with rhodamine B isothiocyanate was added to the mix. All assays were done at 37 °C, except for the last one.

Attending to the results, all gRNAs induced the cleavage of DNA in the conditions tested. The bands of the cut amplicons are smears, which may indicate that, as expected, DSBs can happen in slightly different positions close to the PAM. In some of the assays, a faint band is also visible in the region of the non-cut amplicons, corresponding to a very low percentage of amplicons not successfully cleaved. This is probably due to loss of activity of the Cas9 enzyme during the incubation or to an incubation time not long enough to promote the complete cleavage of DNA.

One essay was done at 18 °C with sgRNA E to test if the temperature at which *P. dumerilii* lives is a factor that reduces the Cas9 activity. On this assay, a band correspondent to a molecular weight higher than the molecular weight of the non-cut amplicon was observed (E (18 °C) in panel 2 of Figure 3.4). A possible explanation is that the lower temperature might reduce the Cas9 activity [167] and at the same time promote a stronger bound of the CRISPR complex to the DNA template. This can cause a delay of the DNA in the gel compared to its normal migration (see the control), despite the heat-inactivation step. Nevertheless, this would need to be further investigated and to be tested with the other gRNAs, as well.

The addition of dextran labeled with rhodamine B isothiocyanate does not seem to affect DNA cleavage (E (dex) in panel 2 of Figure 3.4). The presence of dextran was tested because the Cas9 and

the gRNAs were always delivered to the animals together with this compound, to later allow the sorting out of the animals injected.

Inspired by a previous study [168], gRNAs A and C were tested together at an equimolar concentration (the total molar concentration of gRNA was the same as in the other assays). The result seems to be at least as good as using each of them separately, looking at A+C in panel 2 of Figure 3.4.

Overall, these *in vitro* assays indicate that the chosen gRNAs can lead to DSBs in DNA. However, these positive results do not warrant *in vivo* success. The environment where these cleavages occurred is very simple and controlled compared to the interior of cells.

3.1.3 Validation of the *Arx* knockout

After the *in vitro* cleavage assay, hundreds of embryos were injected with different concentrations and combinations of Cas9 (either as protein or as mRNA) and gRNAs (Table 3.3). CRISPR-Cas9-mediated knockouts depend on a multitude of factors, including crRNA efficiency and epigenetic modifications related to the target locus [137].

Table 3.3. Concentrations of the Cas9 and the gRNAs in each injection mixture tested. Letters indicate the gRNAs. Letters in bold indicate the mixtures that led to mutations in *Arx*. A+B+C+D refers to an equimolar concentration of gRNAs A to D. * means that the effects could not be assessed due to some problems during genotyping.

	Injection mixtures				
Cas9 protein (ng/µL)	24	62	298	-	-
Cas9 mRNA (ng/µL)	-	-	-	333	383
gRNA (ng/µL)	5	13	100	13	100
	A, B, C, D, (A+C+D)*, E	A*, **E**	A, **E**	A, B, E	A, E

To validate the genetic modifications generated, the target loci were sequenced following the Sanger method. Even when genotyping single animals, the amplicons from the gDNA could not be sequenced directly after the PCR. Usually, if Cas9 is able to induce mutations in DNA, cells of the same organism will have different genotypes. Some cells can be wild-type, whereas the others can carry mutations. These animals are mosaic due to a late action of Cas9, which cuts DNA independently in different cells after the first cell division. This, in turn, can be caused by a non-ideal concentration of the Cas9 and the gRNAs, low efficient or unspecific gRNAs or the additional time to translate Cas9 mRNA (when Cas9 is delivered as mRNA and not as protein). Therefore, the amplicons from gDNA were cloned into a vector and latter individually sequenced. This approach only provides a semi-quantitative assessment of the mutation frequency because smaller amplicons are more easily cloned, and provides the knowledge of the exact sequences.

Animals that did not seem to be developing normally (morphologically atypical, not swimming or just spinning) were genotyped in parallel with healthy animals, to screen more injected animals, since it is common to have few survivors after injections. This is also an unbiased approach. Mutations in *Arx* may cause unexpected effects that lead to early death or nonstandard development.

In the beginning, only the gRNAs A-D were tested (the gRNA E was developed later). These gRNAs were first used at a low concentration (5 ng/µL of gRNA and 24 ng/µL of Cas9 protein). Considering that

animals could be wild-type, mosaic mutants or non-mosaic mutants, animals were individually sequenced. No mutations were detected with any of the gRNAs. An additional mixture containing gRNAs B, C and D at equimolar concentrations (5 ng/µL in total) was used, following the results described elsewhere [168]. In this study, the use of multiple gRNAs was described as having a synergistic effect (although the aim was to upregulate genetic expression, using a distinct Cas9). However, a problem with the extraction of the gDNA did not allow to sequence the target loci. This might have occurred due to a non-successful extraction of gDNA, PCR amplification, cloning or transformation.

In subsequent experiments, we decided to not test again the gRNAs C and D due to their lower scores and time constraints. In addition, single animals were never again genotyped, as it requires that the amplicons of each animal are individually cloned, greatly increasing the cost of this process.

Afterwards, instead of the Cas9 protein, Cas9 mRNA was used at 333 ng/µL together with the gRNA A or B at 13 ng/µL. This combination and has been very successful in generating another knockout (the *Dbx* knockout) [48], but did not induce any mutations in the *Arx* gene. A Cas9 mRNA quality control was carried out before testing another injection mixture. The Cas9 mRNA had been produced in the lab and could be degraded (it is difficult to assess its quality by electrophoresis, as it always produces a smear in an agarose gel due to the poly(A) tail). Thus, it was injected on the same conditions as previously to produce the *Dbx* knockout, using the appropriate gRNA. Genotyping was done with a restriction digestion approach and revealed that the *Dbx* knockout was successful (results not shown), meaning that the Cas9 mRNA was not degraded.

The gRNA A was then tested at higher concentrations (100 ng/µL of gRNA) with the Cas9 mRNA at 383 ng/µL or with the Cas9 protein at 298 ng/µL, but mutations were never detected. The sgRNA A was injected also injected at 13 ng/µL with the Cas9 protein at 62 ng/µL. However, the genotyping process failed and it was not possible to assess the effects of this mixture.

Considering it was not worth trying additional conditions with these gRNAs, the sgRNA E was designed and produced in the lab. The sgRNA E had a higher predicted efficiency and was tested in all the conditions formerly used with the other gRNAs. After genotyping, the first mutations in the *Arx* were detected in animals injected with the sgRNA E at 100 ng/µL and the Cas9 protein at 298 ng/µL and with the sgRNA E at 13 ng/µL and the Cas9 protein at 62 ng/µL.

Figure 3.5 shows the mutations detected using the sgRNA E. No large indels were detected. Four of the mutations caused frameshifts. Only the largest deletion (deletion of 9 base pairs) directly affected the predicted homeobox sequence, which is the result of sgRNA E targeting the 3' end of *Arx* homeobox. Nonetheless, even frameshift mutations that do not affect the homeobox are likely to significantly disrupt the Arx function, considering that the homeodomain is located roughly in the middle of the protein. Interestingly, the largest deletion and the largest insertion (insertion of 21 base pairs) do not cause frameshifts. This insertion of 21 base pairs is also curious as it occurred at 3' of the PAM and it is a repetition of the adjacent sequence immediately at its 3'.

Regardless of the amplicons sizes biasing the cloning and therefore some mutations being more easily or difficultly detected than others, a mutation rate can be estimated by the proportion of amplicons with mutations to wild-type amplicons. In this case, an injection mixture with the sgRNA E at 100 ng/µL and the Cas9 protein at 298 ng/µL yielded a mutation rate of 31% (4 out of 13 amplicons had mutations

in *Arx*), whereas the sgRNA E at 13 ng/µL and the Cas9 protein at 62 ng/µL led to 29% (2 out of 7). It is worth mentioning that mutations were only detected in maldeveloped animals.

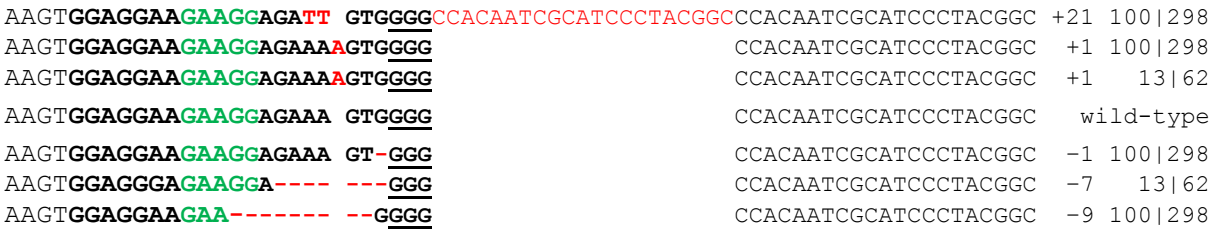


Figure 3.5. Mutations in the *Arx* gene. The size of each indel is indicated. `100|298` and `13|62` indicate in ng/µL the concentrations of the sgRNA E and Cas9 protein used to inject the animals. Insertions (above wild-type) and deletions (below wild-type) are in red. Part of the homeobox sequence is in bigger letters (`AAGTGGAGGAAGAAG`). The target region of the sgRNA E is in bold with the PAM sequence underlined. HpyAV recognition site is in green.

None of the animals injected with the sgRNA E at 100 ng/µL and the Cas9 protein at 298 ng/µL was healthy and therefore only organisms with an atypical development were genotyped. Among the animals injected with the sgRNA E at 13 ng/µL and the Cas9 protein at 62 ng/µL, gDNA was also extracted from healthy animals. Three amplicons from gDNA with this origin were sequenced and were wild-type. Four amplicons came from anomalous animals and included the two amplicons with the mutations presented in Figure 3.5.

To rule out if maldevelopment was an effect of mutations in the *Arx*, pools of animals non-injected and others injected with the Cas9 protein at 62 ng/µL (without gRNA; a control to account for microinjection perturbations and Cas9 effects) were also genotyped. No mutations were detected in 3 amplicons from maldeveloped non-injected, 3 from maldeveloped injected with the Cas9, 5 from healthy non-injected and 1 from healthy animals injected with the Cas9. Still, one should not conclude that mutations in the *Arx* trigger an unexpected development, as few animals were genotyped and few amplicons from each pool of animals were sequenced. Usually, the pools of animals genotyped contained less than 10 larvae. In addition, it is common that only a few animals develop normally after microinjection [62]. Moreover, the control animals should belong to the same batch as those injected with the Cas9 and the gRNA to discard any batch effect.

The fact that mutations were only detected in animals injected with two of the multiple mixtures tested does not mean that the other concentrations or the other gRNAs cannot induce mutations. With a Sanger sequencing approach, it is not possible to know if all the different amplicons were sequenced. However, it is probable that all the other injection mixtures do not generate mutations or generate mutations at a very low frequency, since in most cases more than 5 amplicons were sequenced per condition.

Additionally, the amplicons from animals injected with the sgRNA E were also analyzed on a Bioanalyzer, an automated electrophoresis system that can provide a much higher resolution than a normal electrophoresis on an agarose gel. Theoretically, by separating amplicons with very similar sizes, it could provide another perspective on the distribution of the indels lengths. Still, a single peak correspondent to the wild-type amplicon size was detected in each of the samples (data not shown). This may be the result of very short indels that cannot be distinguished even in such a sensitive system, being consistent with the mutations shown before (Figure 3.5).

Although the gRNA concentration was approximately the same as the one used to generate the very efficient *Dbx* knockout (12 ng/µL), it is curious to verify that *Arx* mutations were obtained only with the Cas9 protein and not with the Cas9 mRNA used to produce the *Dbx* knockout [48]. The Cas9 protein can have some pros and cons compared to the mRNA. The protein is ready to work *in vivo*, whereas the mRNA still needs to be translated. However, the protein can be more toxic, according to unpublished results from the Arendt lab.

Finally, comparing the results from these *in vivo* tests and the *in vitro* assays (3.1.2), it can be hypothesized that epigenetic modifications (such as chromatin compaction) do not allow some gRNAs to efficiently target *in vivo* the *Arx* gene. With a simple amplicon as template, the target loci might be more easily accessible than in the cell nucleus.

3.1.4 Genotyping with a restriction digestion

Genotyping with a restriction enzyme has been suggested as a fast and cheap method to evaluate the gRNAs efficiency [48]. The methodology involves the use of a restriction enzyme with a unique recognition site within the amplicons to genotype, as close as possible to the PAM sequence and in the gRNA-DNA binding region [48]. Since the Cas9 preferentially cuts DNA within the gRNA-DNA binding region close to the PAM, many mutations will remove the restriction site and the amplicons from the target locus will be resistant to the digestion.

This approach was tested with the amplicons from animals injected with the sgRNA E using HpyAV, a restriction enzyme that recognizes the motif GAAGG, unique in the wild-type amplicons and only 7 nt away from the PAM (Figure 3.5). Attending to the results in Figure 3.6, there does not seem to be any big difference between digesting wild-type amplicons (from non-injected animals) or digesting amplicons from animals injected with the sgRNA E. In some cases, a very faint band correspondent to HpyAV-resistant amplicons is visible in some columns of Figure 3.6, including the positive control (column 2).

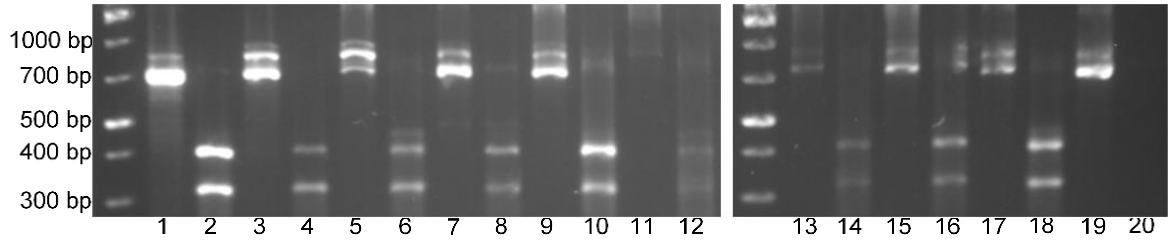


Figure 3.6. Results of digesting with HpyAV the amplicons from gDNA of animals injected with the sgRNA E. The left column of each picture contains the DNA ladder with the length of the bands indicated. Columns with an odd number correspond to negative controls (amplicons not incubated with HpyAV). Columns with an even number correspond to amplicons incubated with HpyAV. Numbers indicate the origin of the amplicons. 1, 2: wild-type; 3, 4: maldeveloped animals injected with 100 ng/µL sgRNA E and 383 ng/µL Cas9 mRNA; 5, 6: normal animals injected with 100 ng/µL sgRNA E and 383 ng/µL Cas9 mRNA; 7, 8: maldeveloped animals injected with 100 ng/µL sgRNA E and 298 ng/µL Cas9 protein; 9, 10: maldeveloped animals injected with 13 ng/µL sgRNA E and 62 ng/µL Cas9 protein; 11, 12: normal animals injected with 13 ng/µL sgRNA E and 62 ng/µL Cas9 protein; 13, 14: maldeveloped animals injected with 13 ng/µL sgRNA E and 333 ng/µL Cas9 mRNA; 15, 16: normal animals injected with 13 ng/µL sgRNA E and 333 ng/µL Cas9 mRNA; 17, 18: maldeveloped animals injected with 5 ng/µL sgRNA E and 24 ng/µL Cas9 protein; 19, 20: healthy animals injected with 5 ng/µL sgRNA E and 24 ng/µL Cas9 protein.

Although SNPs were never detected in the target sequence in wild-type animals, the faint band correspondent to HpyAV-resistant amplicons in the positive control can indicate SNPs in the target region in wild-type animals. This band may be also the result of a contamination when loading the agarose gel before the electrophoresis or the result of an incomplete digestion, caused by any problem during the incubation with HpyAV. Therefore, one should not conclude that similar bands present in the other samples indicate mutations induced by the CRISPR-Cas9 system which could have removed the HpyAV recognition site and could have not been detected by sequencing.

Moreover, by looking at Figure 3.6, there are samples that seem to do not have as much DNA as the others. However, the mass of DNA in each sample digested with HpyAV was carefully adjusted to the same amount. This genotyping method was repeated several times, changing incubation times, DNA concentrations and the amount of enzyme (results not shown), leading to similar results and indicating that HpyAV is very sensitive to the incubation conditions (such as DNA purity). Indeed, according to the company that provides this enzyme (New England Biolabs), HpyAV has star activity in glycerol concentrations above 5%. Even taking this into account, the results did not change (results not shown). One can hypothesize that HpyAV might have the same unspecific activity in other conditions, for instance, caused by residual concentrations of organic solvents from the DNA preparation.

Despite these results, this genotyping method has given reproducible and consistent results when validating the *Dbx* knockout with AvaII [48]. This can suggest that the success of this method depends on the enzyme itself. In fact, AvaII does not have star activity, according to the company that provides this enzyme (New England Biolabs), which can make a difference.

Although this technique is a fast way of genotyping when the enzyme used has a robust and reproducible behavior, there is always a considerable level of uncertainty. Assuming that the enzyme does not have unspecific activity, any problem with the digestion can indicate false positive results because of an incomplete digestion. In addition, there might be many mutations not detected. Attending to results from Figure 3.5, even with a recognition site only 7 nt away of 5' of the PAM sequence, only 1 out of the 6 mutations sequenced would be detected.

3.1.5 Morphological analysis of *Arx* mutants

To study the morphology of *Arx* mutant animals at 6 dpf, embryos were injected with a solution containing 13 ng/µL of sgRNA E and 62 ng/µL of Cas9 protein. Among the two solutions that have successfully induced mutations in *Arx*, this one requires less amount of the Cas9 and sgRNA E. Several hundreds of animals had to be injected to overcome a series of problems that lead to a reduced survival rate at 6 dpf. As mentioned before, microinjection is a very invasive technique that usually causes the death of 75% of the injected animals, due to rupture of the cell membrane and release of cytoplasmic content, while many of the remaining animals do not develop normally [62]. The survival rate is especially low after 4 dpf, when the body starts to elongate [169]. In addition, the percentage of survivors depends on each batch and even on the weather conditions, such as atmospheric depressions that cause the general death of embryos. It is also common to lose animals throughout the whole process of fixing, staining and mounting. Lastly, many animals that receive the Cas9 and the gRNA may not be mutant, or may be mosaic, increasing the need to analyze as many organisms as possible.

During the time available, ten rounds of injections were performed. All animals from the first five rounds died before 6 dpf, which could be caused by some of the factors described before or could be due to any effect produced by the mixture injected. On the subsequent five rounds of injections, a few 6-day-old larvae survived. At the end of the whole process, there were only 1 non-injected wild-type larva (the others were lost during the fixation) and 16 larvae injected with the sgRNA E at 13 ng/µL and the Cas9 protein at 62 ng/µL. There were no survivors injected with only the Cas9 at 62 ng/µL, a control to account for microinjection perturbations and Cas9 effects, such as toxicity.

Even though the main objective was to characterize the brain and more specifically the MBs morphology, the whole body was imaged. This is a less biased approach since not only the region expected to be affected by *Arx* mutation is analyzed. Moreover, organisms with an overall atypical morphology caused by microinjection can be identified.

Figure 3.7 presents images from the only wild-type survivor (panels A1 and A2) and four of the injected animals (B to E). Unfortunately, no good quality anti-acetylated tubulin staining images could be obtained. This staining (here in green) is not visible on the internal side of the animals in panels B, D and E (as well as in all the other 12 animals injected not shown here) and it is not strong in the animals in panels A1 and C. Results suggest that the antibodies could not penetrate the tissues and mainly labelled the cilia on the exterior, probably because of an ineffective digestion of the cuticle by proteinase K. Staining efficiency has been reported to vary across animals and batches [62]. Indeed, the animals where anti-acetylated tubulin staining is visible on the internal side (panels A1 and C) belonged to a different batch from the others.

This bad quality staining greatly hampers the analysis, since DAPI staining on its own does not allow the analysis of the neuronal circuitry. A staining with the quality of that shown in Figure 1.5 would be required to look in detail at the MBs region. To give an idea of the resolution and quality required, the region of the MBs ventral peduncles is highlighted in panel A1 in Figure 3.7. It clearly does not have enough resolution to see the MBs peduncles.

As far as DAPI staining allowed, the overall morphology was compared among the larvae. 13 out of the 16 injected animals (12 of them not shown) had a normal wild-type-like morphology, as exemplified in panel B of Figure 3.7. In contrast, panel C presents a body which is not internally and externally symmetrical and which does not have a regular shape. Strikingly, the head is not round, the ventral nerve chord has at some point three columns (indicated by the arrows), which is not normal, compared with the larva in panel A2. The whole body of the animal in panel C is also shorter and not so elongated. This is sometimes seen in maldeveloped wild-type animals. The larva in panel D presents a wild-type appearance, except for half of the head that is smaller. It can be a mosaic animal with one side of the head normal and the other half lacking some structures. The animal in panel E is wider (as it is the one in C) and has a deformed head, particularly on the right side, where there seems to be a big protrusion. This deformation could have been caused during the immunostaining process.

Only with these results, much can be speculated but very little concluded. It was not possible to characterize any morphological phenotype associated with the *Arx* mutations, since we do not know if any of the animals is mutant, not enough animals were imaged, the anti-acetylated tubulin staining failed and there were not enough controls (only one wild-type non-injected larva). Animals injected with only

the Cas9 would be an important control, revealing the percentage of non-mutant larvae with an atypical development caused by microinjection. Likewise, more wild-type non-injected animals would indicate the proportion of organisms with a natural maldevelopment at this stage.

Aiming at genotyping the imaged larvae to know if they were mutant (even if mosaic), attempts were made to recover the animals from the slides where they were mounted. However, this process failed because it is almost impossible to remove the cover slide without dispersing the glycerol that contains the animals.

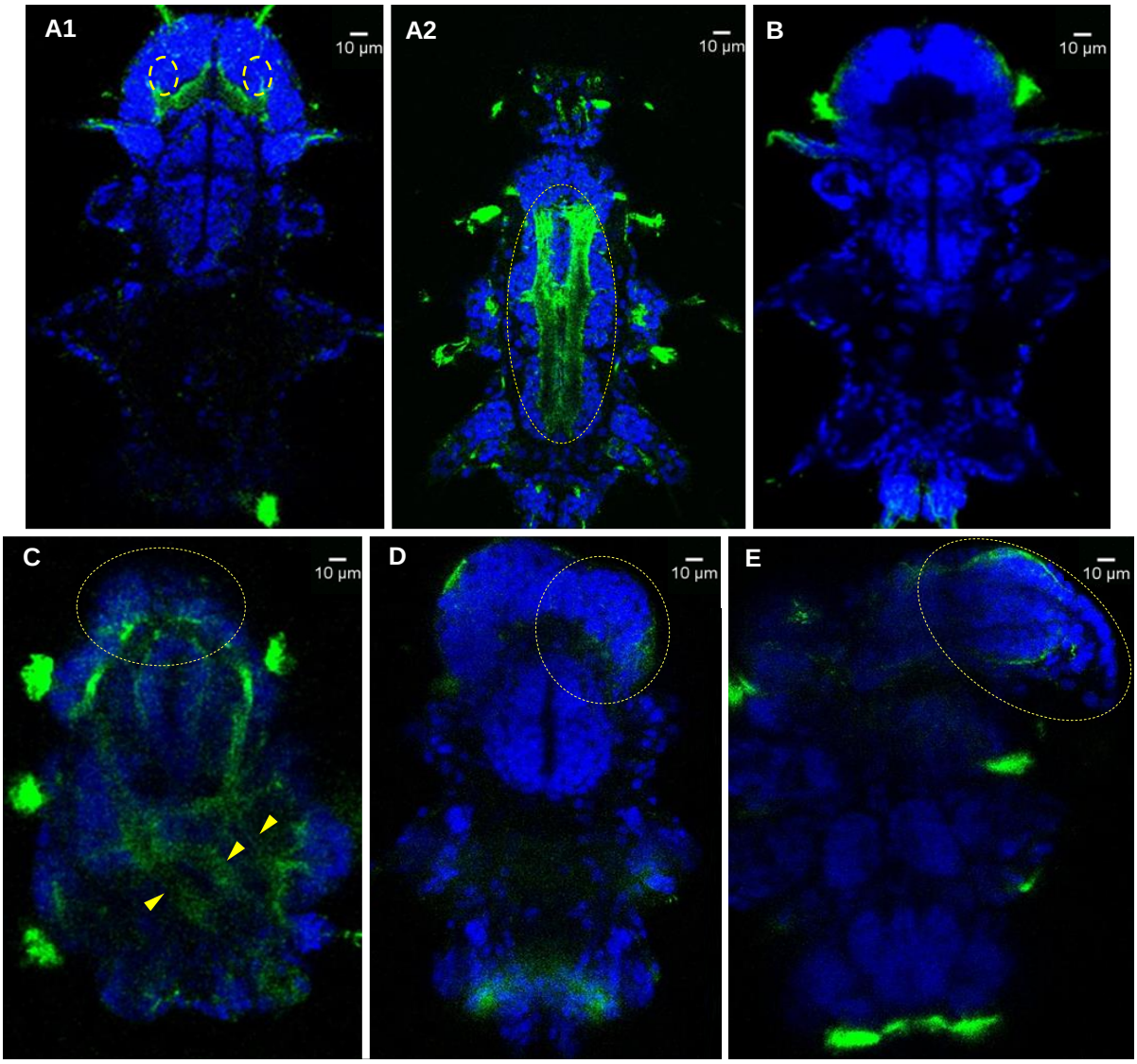


Figure 3.7. *P. dumerilii* larvae at 6 dpf stained with DAPI (in blue) and anti-acetylated tubulin antibody (in green). Anterior side is up and posterior down. A1 and A2: Two planes of a non-injected wild-type larva; dashed lines indicate the region of the ventral peduncles of the MBs (A1) and the ventral nerve chord (A2). B-E: Larvae injected with the sgRNA E at 13 ng/µL and the Cas9 protein at 62 ng/µL. B: Wild-type-like larva. C: Larva with an atypical and asymmetrical form; dashed line indicates the head and the arrows indicate the three columns visible in the ventral nerve chord. D: Larva with a smaller half of the head (indicated by the dashed line). E: Larva with a protrusion in one side of the head (indicated by the dashed line).

3.2 Imaging of chemically-evoked neuronal responses

A recent work has described several neuronal responses in the head of *P. dumerilii* at 6 dpf to a set of compounds [55]. In this work, additional chemicals were tested with the aim of triggering responses in the MBs region.

Putative receptor proteins have been identified in the MBs region based on protein phylogenetic studies (unpublished results from the Arendt lab). These receptors might be associated with the hypothetical chemosensory role of the MBs at early stages of development. Therefore, the ligands of these receptor proteins are good candidates to fire the MBs. These ligands are expected to be the same as the ones of the proteins with the closest sequences to the putative receptors.

For several reasons, including time constraints and availability of the chemicals in the lab, dopamine and phenylacetaldehyde were the only chemicals suggested by the analysis of protein phylogenetic trees (data not shown) that were tested. Additionally, putrescine (1,4-diaminobutane) was tested. Being one of the major chemicals responsible for the smell of decaying flesh, it is a strongly repulsive or attractant odor in many species [170]. It might elicit a strong behavioral response in *P. dumerilii*, associated with a fierce neuronal activation (perhaps of the MBs). All these chemicals were tested at 1 mM.

Only one animal injected with GCaMP6s survived until 6 dpf. Nonetheless, calcium imaging was done on this larva following a recently developed method that takes advantage of a microfluidics setup. For each chemical, the larva was exposed 3 times 15 s to the stimulus and to the negative control (FNSW). In one of the experiments, FNSW was used as a stimulus. In another experiment, 1-butanol at 10 µM was a positive control, since it triggers clear responses in a large number of cells of different organs [55]. However, the results of this last exposure could not be analyzed (data not shown) because of the position in which the larva was in that moment.

Figure 3.8 presents a summary of the results. There was no difference between the neuronal responses to the three exposures to each stimulus. Therefore, Figure 3.8 only presents the fluorescence detected for one of the three exposures to each stimulus (A-D) and to the FNSW controls (A''-D''). The fluorescence images obtained when the animal was not exposed to either the stimulus or the control are also shown (A'-D').

Considering that the aim of this preliminary work was to give a general idea of the regions activated by each stimulus (paying special attention to the MBs), GCaMP6s fluorescence intensity was not quantified. The images in Figure 3.8 present the mean fluorescence obtained during 5 s. The mean fluorescence shown in panels A'-D' results from the average of the fluorescence signal over the 5-s period immediately before one of the exposures to each stimulus. The mean fluorescence shown in panels A-D and A''-D'' results from the average of the fluorescence signal over the 5-s period 2 s after the beginning of one of the exposures to each chemical stimulus or control.

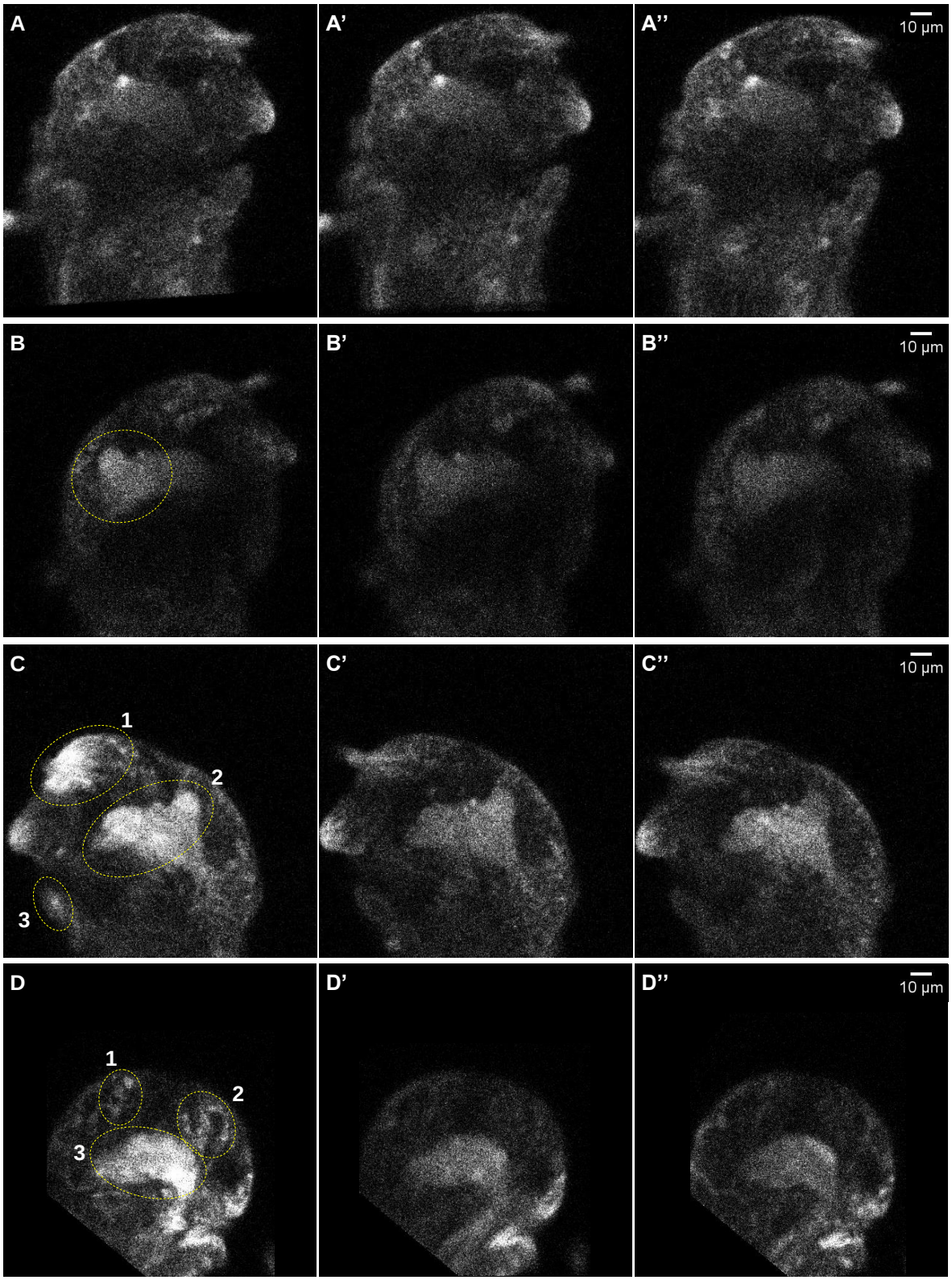


Figure 3.8. Calcium signal in the head of a 6-day-old *P. dumerilii* larva injected with GCaMP6s mRNA. The brighter the higher the calcium signal fluorescence. A-D: larva exposed to FNSW, dopamine, phenylacetaldehyde and putrescine, respectively. A'-D': larva not exposed to the stimulus nor the FNSW control ("resting interval"). A''-D'': larva exposed to the FNSW control. Images are an average of the fluorescence over 5 s. In all images, the anterior side is up and posterior down. In A-A'' and B-B'' the dorsal side is on the left and the ventral on the right. In C-C'', dorsal side is on the right and ventral on the left. In D-D'', the larva was imaged in dorsal view. Regions highlighted: neuropil and the apical plexus (B), antennal muscle (C 1), the whole neuropil (C 2), palps (C 3), antennal nerve (D 1), antennal nerve and possibly the dorsal peduncle of the MBs (D 2), and the neuropil (D 3).

The microfluidics setup involves three inflow currents: one contains the stimulus, another the FNSW control and a third one that also contains FNSW. This third current (channel 0 in Figure 2.2) is continuously active to avoid mixing of the two other currents, as shown in Figure 2.2. Therefore, when using FNSW as a stimulus, the larva was exposed to FNSW from three different channels (A-A'' in Figure 3.8). Considering that the calcium signal is similar in panels A-A'', it is possible to conclude that there is no response to the flow switching from either the stimulus or the control channels. This control is also important to detect cells that spontaneously fire. In this experiment, no such cells were seen (results not shown used to generate the images in Figure 3.8).

Attending to images B-D and comparing them with B''-D'', all chemical stimuli tested generated neuronal responses of some sort. These responses are difficult to compare because the larva moved between the experiments, being imaged in different positions. By comparing these results with previous calcium imaging data [55], some regions activated by the chemical stimuli tested can be distinguished. Dopamine activates the neuropil and the apical plexus close to the apical organ (B-B''). Phenylacetaldehyde triggers a very strong retraction of the antenna (activation of the antennal muscle, region 1 in C) and a clear response of the whole neuropil (region 2 in C), including the apical plexus and probably the region where the MBs are developing, and the palps (region 3 in C). Putrescine activates the antennal nerve (regions 1 and 2 in D) and the neuropil (region 3 in D). Putrescine may also activate the MBs, as in region 2 more than one nerve bundle seems to be active. To draw stronger conclusions about the regions activated by the chemical stimuli tested, the same experiments will have to be repeated in more larvae.

Based on these results, it is impossible to surely say whether the MBs responded to any stimulus. Not only is this structure hard to identify at this developmental stage, but also their characteristic peduncles are anatomically variable across animals and ambiguous to identify [62]. In addition, it is hard to say whether a certain cell belongs to the MBs other than by the proximity to the peduncles MBs, which are rarely seen based only on GCaMP6s signal [55]. The insights provided by these calcium imaging experiments on the putative chemosensory role of the MBs are limited.

It is also difficult to distinguish between primary and secondary responses, even by analyzing fluorescence images not averaged over time (results not shown used to generate the images in Figure 3.8). Assuming that MBs are involved in chemosensation, it would be important to determine whether these structures are involved in directly sensing chemical cues from the exterior, possibly using the projections to the head surface previously described [62], or they have interneuron functions, as recently suggested for the MBs ventral lobes [55].

Concerning the responses to dopamine, it is questionable if it makes sense to test a compound known as a neurotransmitter and/or a neuromodulator in invertebrates [171], unlikely to be present in FNSW. If dopamine is able to reach the tissues inside the head, we can think that the neuronal activation could be caused by a direct interference of this compound with the synapses.

Based on the results presented here, one can speculate that the tested concentrations were too high, leading to unspecific responses that would explain such strong responses with phenylacetaldehyde, for instance. Hence, the concentration at which each compound is tested has to be optimized to promote the activation of only the most sensitive structures [55].

The larva used in these calcium imaging experiments was most of the time in an unusual position. The chip was designed to trap animals in a dorsal-ventral position, as in D-D''. Imaging the animal from the side seems to give another perspective and might contribute to clarify some doubts on certain cell-specific activations.

The larva used in this study was injected with GCaMP6s mRNA together with dextran labeled with rhodamine B isothiocyanate. It is not a common procedure to coinject this dextran derivative in these experiments, but it can make the process of sorting out the injected from non-injected animals much easier. These experiments also show that the procedure did not affect the calcium signal, compared to previous results [55].

Lastly, the larva used in these experiments was subject to exposures with different chemicals, proving that it is possible to screen various chemicals in a single animal. There are strong responses presented in panels B-D, but the neuronal activity when exposed to the FNSW is invariable (panels A-A'', B'-D' and B''-D''). This indicates that there was no contamination of the two FNSW channels with the chemical stimuli. Usually, only FNSW and one stimulus were tested per larva and per microfluidics chip [55].

4 Conclusion and future perspectives

Platynereis dumerilii is a relatively simple neurobiological model organism that can provide insights into the nervous system architecture and function. For its phylogenetic position and slow-evolving characteristics, *P. dumerilii* is a good system to explore the nervous system origin, a current field of research in evo-devo. This organism has mushroom bodies, a sensory-associative brain structure homologous to the vertebrate pallium [37] that might be chemosensory, according to putative chemoreceptor proteins expressed in that region and neuronal projections to the head surface at early stages of development [37,62]. Likely involved in the development of this head structure is the transcription factor Arx, specifically expressed in the MBs at early stages [38,62]. Motivated by this, the main objective of this thesis was to establish a CRISPR-Cas9-mediated knockout of *Arx*.

Using the Cas9 endonuclease from *Streptococcus pyogenes*, the *Arx* knockout was successfully implemented. Based on the transcriptomic and genomic data available, exons and introns of the *Arx* were predicted and crRNAs targeting the homeobox were designed. After producing some of the gRNAs as sgRNAs, an *in vitro* assay was done to discard eventually non-efficient gRNAs. Since the results indicated that all of them were able to induce *in vitro* DSBs in DNA, sgRNAs were *in vivo* tested in different combinations and concentrations with the Cas9, provided as protein or mRNA. After sequencing, a few indels were detected in animals injected with two different mixtures containing the gRNA predicted to be the most efficient. The small sizes of the six mutations sequenced were consistent with the results of analyzing all the amplicons from the animals mutated on a Bioanalyzer, meaning that all other mutations eventually not sequenced were also small indels. Together with the fact that the successful gRNA targeted the downstream end of the *Arx* homeobox, this explains why only one of the mutations directly affected the homeobox. A different genotyping approach that makes use of a restriction digestion was unsuccessfully used, and does not seem to be an alternative to sequencing.

Afterwards, 6-day-old larvae injected with one of the successful mixtures were stained with DAPI and with an antibody targeting acetylated tubulin. However, no conclusions about the involvement of Arx in the MBs development could be drawn due to a generalized bad staining of the axon bundles and lack of control animals.

Finally, a newly developed method to perform calcium imaging in a microfluidics device was employed to assess the head neuronal activity triggered by three chemical stimuli, exploring the putative chemosensory role of the MBs at 6 dpf. MBs might be activated by putrescine and phenylacetaldehyde, but further work is required to take stronger conclusions. Only one larva was imaged, and the MBs are hard to identify at 6 dpf, especially with only GCaMP6s signal [55]. Nonetheless, this experiment showed that it is possible to screen various chemicals in a single animal and in a single chip, without contamination of the two FNSW inflow currents of the chip, and that dextran labeled with rhodamine B isothiocyanate apparently does not affect GCaMP6s fluorescence.

Much work still needs to be done to unveil the role of Arx in the MBs development and in the sensory capacities of these structures.

Next, it will be necessary to scale up and continue the work done in this project. Now that injection solutions capable of inducing indels in *Arx* have been established, it might be useful to repeat these

experiments and genotype single animals. In this work, most of the times pools of animals were genotyped to screen a broader range of conditions in a fast and cheap manner. Genotyping single animals would give a better notion of how many injected larvae are still wild-type, mosaic or non-mosaic mutant. Moreover, if next-generation sequencing is used, the mutation rate can be more exactly calculated. Off-targets should also be explored by sequencing the off-target candidate regions shown in Figure 3.2, or by analyzing the whole genome in an unbiased manner with next-generation sequencing.

It must be stressed that it has not been proved that none of the other injection solutions or gRNAs are completely ineffective. Besides sequencing more amplicons with the Sanger method, other genotyping approaches such as next-generation sequencing or T7 endonuclease I assay [172] could detect mutations that may have gone unnoticed. As a suggestion that might enhance CRISPR-Cas9 efficiency, injected eggs can be incubated at temperatures above the usual temperature (18 °C) immediately after injection. A very recent investigation revealed that at least in some cell lines this gene editing system is more effective at higher temperatures [167].

It will be important to determine if mutations in *Arx* cause an overall anomalous development, as all detected mutations belonged to animals with an atypical development. This is unexpected, knowing that all animals were genotyped before 48 hpf and *Arx* only starts to be expressed after 3 dpf (unpublished results from the Arendt lab). Additionally, *Arx* is specifically expressed in the MBs at least from 4 dpf to 12 dpf [37,38,62] (and unpublished data from the Arendt lab) and therefore a localized effect is expected. Yet, whole mount *in situ* hybridizations at different developmental stages should be done to rule out *Arx* ectopic expression caused by the knockout. This can be combined with single-cell RNA-sequencing to analyze alterations in the expression levels of other genes. Single-cell RNA-sequencing results can then be map against the available 3D gene expression atlas with cellular resolution for an entire *P. dumerilii* larva at 6 dpf [38].

Morphological analysis of the mutant larvae should be performed using as many animals as possible to statistically take into account mosaicism and microinjection perturbations. Ideally, control and non-control animals of the same batches should be used. Being aware of how difficult this can be, more than one person should perform microinjections in parallel, maximizing the number of animals from the same batch injected with the Cas9 and gRNA mixture and the Cas9 without gRNA (a control for microinjection perturbations and Cas9 toxicity).

Considering that the peduncles of the MBs are difficult to spot in larvae at 6 dpf, not only DAPI and anti-acetylated tubulin antibody should be used, but also a membrane dye such as FM 4-64FX (from Thermo Fisher Scientific) [62]. This might increase the resolution and allow an easier analysis of the immunostainings. In addition, slightly latter stages can be investigated. Transmission electron microscopy data can help to compare developmental alterations caused by the *Arx* knockout.

Genotyping single animals after immunostaining is also of interest. This procedure would allow discarding from the analysis some of the animals with natural or microinjection-induced malformations. Although this was tried in this project without success, once the larvae are recovered from the mounting, it should be feasible as a common extraction of gDNA.

As soon as a phenotype associated with *Arx* knockout is found, other independent knockouts of this transcription factor should be developed using different gRNAs. Recalling that in this work it was

preferred to directly target the homeodomain rather than targeting more upstream parts of *Arx*, crRNAs #2 and #4 from Table 3.1 could be used, aiming at confirming and strengthening a phenotype-genotype association.

In all the experiments carried out during this work, a knock in of the *Arx* for a fluorescent protein gene (such as *GFP*) would help to identify mutant cells in mutant animals, despite the lower efficiency of the knockout with this approach [48]. An *Arx* knockout line would also be valuable. For instance, to further validate phenotype-genotype associations, the wild-type *Arx* gene could be heterologously expressed in an *Arx* mutant line of *P. dumerilii* to try to rescue wild-type phenotypes. If the *Arx* knockout results in the selective ablation of the MBs (due to the localized expression of *Arx*) and an *Arx* mutant line is possible to establish, it would be a tremendous advantage over methods such as laser ablation. This method requires the ablation of the MBs, a structure hard to identify at 6 dpf and deep inside the head. Moreover, laser ablation needs to be performed in every single animal to test [62].

Another aspect that might be considered is a gene knockdown instead of a knockout. Activation of a compensatory network can happen in response to deleterious mutations, whereas it may not occur with knockdowns [173]. A knockdown can also be helpful if the knockout is lethal and does not allow the study of the Arx role in the MBs development, let alone the establishment of a line. To generate a knockdown of Arx, maybe targeting a site downstream of the homeodomain would reduce Arx activity without completely disrupting its function. Besides, there are modified CRISPR systems that can specifically (and transiently) downregulate genes [174], as well as different RNA interference methods [175].

Additional work is required to test the hypothesis that the MBs possess chemosensory features. In this project, only one larva was imaged. Therefore, this investigation needs to be expanded, testing again the same chemicals used in this work and others that are likely to activate the MBs. Different concentrations should be used to trigger specific responses and several chemicals can be tested on a single animal. Dextran labeled with rhodamine B isothiocyanate might be used to facilitate the sorting out of injected animals. Furthermore, other chip designs may give a complementary perspective on the neural responses, by imaging the larvae from the side, for instance.

Calcium imaging experiments can be performed in the *Arx* mutants and neuronal responses of wild-type and mutants can be compared. To do such an experiment, GCaMP6s mRNA could be injected together with the CRISPR solution, or animals from an *Arx* mutant line could be injected solely with GCaMP6s. More difficultly to establish, but theoretically appealing would be a line where the *Arx* is knocked in for *GCaMP6s*.

It will always be hard to surely attribute neuronal firing to the MBs and to classify these activations between primary or secondary responses, due to the MBs morphology at 6 dpf. Nevertheless, some suggestions may facilitate this task. Instead of whole head imaging, only looking into the MBs region can increase the imaging spatiotemporal resolution. Moreover, a genetically encoded membrane marker can be coinjected with GCaMPs mRNA, which would help to distinguish different cells/structures inside the head. Such an approach has been recently employed (unpublished results from Luis Bezares-Calderón).

Knowing that dye-filling has been described as a characteristic of chemosensory neurons in *C. elegans* [176], live dye-filling experiments with MitoTracker fluorescent dyes (from Thermo Fisher

Scientific) have been used together with transmission electron microscopy data to identify chemosensory neurons [58]. Similar experiments can be added to the calcium imaging results to determine whether responses are primary or secondary. Activity associated with synaptic inputs may indicate secondary responses of an interneuron, while dendritic microvillar structures can be linked to primary responses of a sensory neuron [58]. Similar experiments can be used to understand the function of the MBs neuronal projections to the head surface at 6 dpf [62].

All the experiments suggested here can be coupled to behavior and learning investigations. These studies may reveal a participation of *P. dumerilii* Arx in cognition, knowing the specific expression of Arx in the MBs (considered to be a sensory-associative brain center [37]), and its involvement in the nervous system ontogeny of different animals [106,113,115,117] and in the human cognitive development [120,122].

Some general considerations should also be taken into account.

Microinjection is a widely employed delivery method for whoever works on *P. dumerilii*. Usually, the concentrations of the injection mixes are adjusted by diluting with RNase- or DNase-free water. However, in other organisms it is common to use a microinjection buffer. This should be considered when working on *P. dumerilii*. Introducing solutions with low osmotic pressure into embryos at the one-cell stage may reduce even more the survival rate of animals likely harmed by the microinjection procedure and which develop in sea water, a high osmotic pressure environment [177,178].

To do gene editing in the future, other CRISPR nucleases recognizing other PAM sequences may be considered to target any desired genome location [179–182]. Moreover, algorithms have been developed to estimate the frequency of in-frame mutations (undesired to produce a knockout) based on the target site [183]. The use of modified CRISPR systems, such as the double nickase system, can reduce the number of off-targets [184]. Furthermore, different tools to design gRNAs can be used and the predicted efficiencies compared. The gRNAs with higher scores across different programs are likely to be the best.

Although the work developed in this project is not enough to answer many of the questions investigated, it is an important step towards understanding Arx role in the MBs development, opening the door to multiple experiments. Ultimately, these advances can contribute to characterize the MBs and Arx functions in Lophotrochozoa and further unveil the origin of the nervous system in Bilateria. Additionally, the homology of the MBs to the vertebrate pallium [37], the conserved functions of Arx among distantly related animals [106,113,115,117], together with the importance of Arx mutations to some neurologic human disorders [120,122,123], make this a highly relevant field of study. This research can eventually provide insights into the conserved basic neurologic mechanisms, who knows contributing to the understanding of mental disorders caused by Arx mutations.

Bibliography

1. Raff, R. A. Evo-devo: the evolution of a new discipline. *Nat Rev Genet* **1,** 74–79 (2000).
2. Fröbius, A. C. & Funch, P. Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. *Nat Commun* **8,** 9 (2017).
3. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A Homologous Protein-Coding Sequence in *Drosophila* Homeotic Genes and Its Conservation in Other Metazoans. *Cell* **37,** 403–408 (1984).
4. Subramoniam, T. *Sexual Biology and Reproduction in Crustaceans*. *Academic Press*, USA (2016).
5. Müller, G. B. Evo–devo: extending the evolutionary synthesis. *Nat Rev Genet* **8,** 943–949 (2007).
6. Sommer, R. J. The future of evo–devo: model systems and evolutionary theory. *Nat Rev Genet* **10,** 416–422 (2009).
7. Love, A. C. & Raff, R. A. Knowing your ancestors: Themes in the history of evo-devo. *Evol Dev* **5,** 327–330 (2003).
8. Minelli, A. *et al.* The changing role of the embryo in evolutionary thought: roots of Evo-Devo. *Heredity* **96,** 419–420 (2006).
9. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* **9,** 868–882 (2008).
10. Willmer, E. N. *Cytology and Evolution*. *Academic Press*, USA (1970).
11. Salvini-Plawen, L. V. & E., M. *Evolutionary Biology*. *Plenum*, USA (1977).
12. Raff, R. A. Written in stone: fossils, genes and evo–devo. *Nat Rev Genet* **8,** 911–920 (2007).
13. Davidson, E. H. Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science* **311,** 796–800 (2006).
14. Adoutte, A. *et al.* The new animal phylogeny: reliability and implications. *PNAS* **97,** 4453–6 (2000).
15. Rokas, A. & Carroll, S. B. Bushes in the tree of life. *PLoS Biology* **4,** 1899–1904 (2006).
16. Koch, C. Complexity and the Nervous System. *Science* **284,** 96–98 (1999).
17. Ohtaki, H. & Shioda, S. *Neuroanesthesia and Cerebrospinal Protection. Springer*, Japan (2015).
18. Arendt, D. *et al.* The evolution of nervous system centralization. *Anim Evol Genomes, Foss Trees* **363,** 1523–1528 (2008).
19. Moroz, L. L. On the independent origins of complex brains and neurons. *Brain Behav Evol* **74,** 177–190 (2009).
20. Swanson, L. W. *Brain architecture. Understanding the basic plan*. *Oxford University Press*, UK (2011).
21. Brusca, R. C., Moore, W. & Shuster, S. M. *Invertebrates*. *Sinauer Associates*, USA (2016).
22. Young, J. Z. Structure and function in the nervous systems of invertebrates. *Electroencephalogr Clin Neurophysiol* **20,** 538 (1966).
23. Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* **22,** 1246–1253 (2005).
24. Aguinaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387,** 489–493 (1997).
25. Bullock, T. H. & Horridge, G. A. *Structure and function in the nervous systems of invertebrates. W. H. Freeman*, USA (1965).
26. Ayala, F. J., Rzhetsky, A. & Ayala, F. J. Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *PNAS* **95,** 606–611 (1998).
27. Bromham, L., Rambaut, A., Fortey, R., Cooper, A. & Penny, D. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *PNAS* **95,** 12386–12389 (1998).
28. Bard, J. *Principles of Evolution: Systems, Species, and the History of Life*. *Garland Science*. USA (2016).
29. Hoffecker, J. F. *Modern Humans: Their African Origin and Global Dispersal*. *Columbia University Press*, USA (2017).
30. Knoll, A. H. Early Animal Evolution: Emerging Views from Comparative Biology and Geology. *Science* **284,** 2129–2137 (1999).
31. Brunet, T., Lauri, A. & Arendt, D. Did the notochord evolve from an ancient axial muscle? The axochord hypothesis. *BioEssays* **37,** 836–850 (2015).
32. Arendt, D. & Wittbrodt, J. Reconstructing the eyes of Urbilateria. *Philos Trans R Soc B Biol Sci* **356,** 1545–1563 (2001).
33. Hirth, F. *et al.* An urbilaterian origin of the tripartite brain: developmental genetic insights from *Drosophila*. *Development* **130,** 2365–2373 (2003).
34. Raible, F. Vertebrate-Type Intron-Rich Genes in the Marine Annelid *Platynereis dumerilii*. *Science* **310,** 1325–1326 (2005).
35. Raible, F. & Arendt, D. Metazoan Evolution: Some Animals Are More Equal than Others. *Curr Biol* **14,** R106–R108 (2004).
36. Denes, A. S. The evolution of dorsoventral patterning and of neuron types in the trunk nervous system of bilaterian animals. (2007).
37. Tomer, R., Denes, A. S., Tessmar-Raible, K. & Arendt, D. Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium. *Cell* **142,** 800–809 (2010).
38. Vergara, H. M. *et al.* Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*. *PNAS* **114,** 5878–5885 (2017).
39. Arendt, D., Tosches, M. A. & Marlow, H. From nerve net to nerve ring, nerve cord and brain: evolution of the nervous system. *Nat Rev Neurosci* **17,** 61–72 (2016).
40. Strausfeld, N. J. & Hirth, F. Deep Homology of Arthropod Central Complex and Vertebrate Basal Ganglia. *Science* **340,** 157–161 (2013).
41. Holland, N. D. Nervous systems and scenarios for the invertebrate-to-vertebrate transition. *Philos Trans R Soc Lond B Biol Sci* **371,** 20150047 (2016).
42. Holland, L. Z. *et al.* Evolution of bilaterian central nervous systems: a single origin? *Evodevo* **4,** 27 (2013).
43. Fischer, A. H., Henrich, T. & Arendt, D. The normal development of *Platynereis dumerilii* (Nereididae, Annelida). *Front Zool* **7,** 31 (2010).

44. Tessmar-Raible, K. & Arendt, D. New animal models for evolution and development. *Genome Biol* **6,** 303 (2005).

45. Bonhomme, F. Evolutionary Relationships in the Genus Mus. *Curr Top Microbiol Immunol* **1,** 19–34 (1986).

46. Morris, S. C., Peel, J. S., Conway, S. & Uk, M. The earliers annelids: Lower Cambrian polychaetes from the Sirius Passet Lagerstatte, Peary Land, North Greenland. *Acta Palaeontol Pol* **53,** 137–148 (2008).

47. Shu, D. Cambrian explosion: Birth of tree of animals. *Gondwana Res* **14,** 219–240 (2008).

48. Vergara, H. M. Descriptive and functional approaches for a system-level understanding of *Platynereis dumerilii* and the evolution of locomotor circuits in Bilateria. *University of Heidelberg,* Germany (2016).

49. Raible, F. & Tessmar-Raible, K. *Platynereis dumerilii*. *Curr Biol* **24,** R676–R677 (2014).

50. Tessmar-Raible, K. *et al.* Conserved Sensory-Neurosecretory Cell Types in Annelid and Fish Forebrain: Insights into Hypothalamus Evolution. *Cell* **129,** 1389–1400 (2007).

51. Dray, N. *et al.* Hedgehog Signaling Regulates Segment Formation in the Annelid *Platynereis*. *Science* **329,** 339–342 (2010).

52. Tessmar-Raible, K. & Arendt, D. Emerging systems: Between vertebrates and arthropods, the Lophotrochozoa. *Curr Opin Genet Dev* **13,** 331–340 (2003).

53. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* **33,** 503–509 (2015).

54. Tosches, M. A. Development and function of brain photoreceptors in the annelid *Platynereis dumerilii*. *University of Heidelberg,* Germany (2013).

55. Chartier, T. Chemosensation in the marine annelid *Platynereis dumerilii*: anatomy, physiology, behaviour. *University of Heidelberg,* Germany (2017).

56. Bannister, S. *et al.* TALENs mediate efficient and heritable mutation of endogenous genes in the marine annelid *Platynereis dumerilii*. *Genetics* **197,** 77–89 (2014).

57. Zantke, J. *et al.* Genetic and genomic tools for the marine annelid *Platynereis dumerilii*. *Genetics* **197,** 19–31 (2014).

58. Conzelmann, M. *et al.* Conserved MIP receptor-ligand pair regulates *Platynereis* larval settlement. *PNAS* **110,** 8224–9 (2013).

59. Backfisch, B. *et al.* Stable transgenesis in the marine annelid *Platynereis dumerilii* sheds new light on photoreceptor evolution. *PNAS* **110,** 193–198 (2013).

60. Tessmar-Raible, K., Steinmetz, P. R. H., Snyman, H., Hassel, M. & Arendt, D. Fluorescent two-color whole mount *in situ* hybridization in *Platynereis dumerilii* (Polychaeta, Annelida), an emerging marine molecular model for evolution and development. *Biotechniques* **39,** 460–462 (2005).

61. Gazave, E. *et al.* Posterior elongation in the annelid *Platynereis dumerilii* involves stem cells molecularly related to primordial germ cells. *Dev Biol* **382,** 246–267 (2013).

62. Dürichen, W. Neuroanatomy and Function of Mushroom Bodies in 6- and 12- days old *Platynereis dumerilii* larvae. *University of Heidelberg,* Germany (2016).

63. Tomer, R. The evolution of mushroom body and telencephalic cell types, studied by single cell expression profiling of *Platynereis dumerilii* larvae. *University of Heidelberg,* Germany (2008).

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

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

66. Denes, A. S. *et al.* Molecular Architecture of Annelid Nerve Cord Supports Common Origin of Nervous System Centralization in Bilateria. *Cell* **129,** 277–288 (2007).

67. Tosches, M. A., Bucher, D., Vopalensky, P. & Arendt, D. Melatonin Signaling Controls Circadian Swimming Behavior in Marine Zooplankton. *Cell* **159,** 46–57 (2014).

68. Jékely, G. *et al.* Mechanism of phototaxis in marine zooplankton. *Nature* **456,** 395–399 (2008).

69. Steinmetz, P. R. H., Kostyuchenko, R. P., Fischer, A. & Arendt, D. The segmental pattern of otx, gbx, and *Hox* genes in the annelid *Platynereis dumerilii*. *Evol Dev* **13,** 72–79 (2011).

70. Arendt, D. Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain. *Science* **306,** 869–871 (2004).

71. Lauri, A. *et al.* Development of the annelid axochord: Insights into notochord evolution. *Science* **345,** 1365–1368 (2014).

72. Fischer, A. & Dorresteijn, A. The polychaete *Platynereis dumerilii* (Annelida): A laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *BioEssays* **26,** 314–325 (2004).

73. Hauenschild, C. & Fischer, A. Platynereis dumerilii: Mikroskopische Anatomie, Fortpflanzung, Entwicklung. *Großes Zool Prakt H* 1–55 (1969).

74. Daly, J. M. Behavioural and Secretory Activity during Tube construction by *Platynereis dumerilii* Aud M. Edw. [Polychaeta: Nereidae]. *J Mar Biol Assoc United Kingdom* **53,** 521 (1973).

75. Ackermann, C., Dorresteijn, A. & Fischer, A. Clonal domains in postlarval *Platynereis dumerilii* (Annelida: Polychaeta). *J Morphol* **266,** 258–280 (2005).

76. Farris, S. M. Evolutionary convergence of higher brain centers spanning the protostome-deuterostome boundary. in *Brain, Behavior and Evolution* **72,** 106–122 (2008).

77. Dujardin, F. Mémoire sur le système nerveux des insectes. *Ann Sci Nat é* **14,** 195–205 (1850).

78. Strausfeld, N. J., Hansen, L., Li, Y., Gomez, R. S. & Ito, K. Evolution, Discovery, and Interpretations of Arthropod Mushroom Bodies. *Learn Mem* **5,** 11–37 (1998).

79. Vowles, D. M. Olfactory learning and brain lesions in the wood ant (*Formica rufa*). *J Comp Physiol Psychol* **58(1),** 105–111 (1964).

80. Heisenberg, M. What Do the Mushroom Bodies Do for the Insect Brain? An Introduction. *Learn Mem* **5,** 1–10 (1998).

81. Menzel, R., Erber, J. & Masuhr, T. Learning and Memory in the Honeybee. *Exp Anal Insect Behav* **282**, 195–217 (1974).

82. Mizunami, M., Weibrecht, J. M. & Strausfeld, N. J. Mushroom bodies of the cockroach: Their participation in place memory. *J Comp Neurol* **402,** 520–537 (1998).

83. de Belle, J. & Heisenberg, M. Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science* **263,** 692–695 (1994).

84. Han, P. L., Levin, L. R., Reed, R. R. & Davis, R. L. Preferential Expression of the *Drosophila rutabaga* Gene in Mushroom Bodies, Neural Centers for Learning in Insects. *Neuron* **9,** 619–627 (1992).

85. Heisenberg, M., Borst, A., Wagner, S. & Byers, D. *Drosophila* Mushroom Body Mutants are Deficient in Olfactory Learning. *J Neurogenet* **2,** 1–30 (1985).

86. Joiner, W. J., Crocker, A., White, B. H. & Sehgal, A. Sleep in *Drosophila* is regulated by adult mushroom bodies. *Nature* **441,** 757–760 (2006).

87. Pitman, J. L., McGill, J. J., Keegan, K. P. & Allada, R. A dynamic role for the mushroom bodies in promoting sleep in *Drosophila*. *Nature* **441,** 753–756 (2006).

88. Martin, J. R., Ernst, R. & Heisenberg, M. Mushroom bodies suppress locomotor activity in *Drosophila melanogaster*. *Learn Mem* **5,** 179–191 (1998).

89. Tang, S. Choice Behavior of *Drosophila* Facing Contradictory Visual Cues. *Science* **294,** 1543–1547 (2001).

90. Wolff, G. H. & Strausfeld, N. J. Genealogical Correspondence of Mushroom Bodies across Invertebrate Phyla. *Curr Biol* **25,** 38–44 (2015).

91. Wolff, G. H., Thoen, H. H., Marshall, J., Sayre, M. E. & Strausfeld, N. J. An insect-like mushroom body in a crustacean brain. *Elife* **6,** e29889 (2017).

92. Farris, S. M. Structural, functional and developmental convergence of the insect mushroom bodies with higher brain centers of vertebrates. *Brain Behav Evol* **72,** 1–15 (2008).

93. Farris, S. M. & Roberts, N. S. Coevolution of generalist feeding ecologies and gyrencephalic mushroom bodies in insects. *PNAS* **102,** 17394–17399 (2005).

94. Farris, S. M. Developmental organization of the mushroom bodies of *Thermobia domestica* (Zygentoma, Lepismatidae): Insights into mushroom body evolution from a basal insect. *Evol Dev* **7,** 150–159 (2005).

95. Gronenberg, W. & Hölldobler, B. Morphologic representation of visual and antennal information in the ant brain. *J Comp Neurol* **412,** 229–40 (1999).

96. Strausfeld, N. J., Strausfeld, C. M., Stowe, S., Rowell, D. & Loesel, R. The organization and evolutionary implications of neuropils and their neurons in the brain of the onychophoran *Euperipatoides rowelli*. *Arthropod Struct Dev* **35,** 169–196 (2006).

97. Strausfeld, N. J. & Barth, F. G. Two visual systems in one brain: Neuropils serving the secondary eyes of the spider *Cupiennius salei*. *J Comp Neurol* **328,** 43–62 (1993).

98. Farris, S. M. Evolution of insect mushroom bodies: Old clues, new insights. *Arthropod Struct Dev* **34,** 211–234 (2005).

99. Wolff, G. H. & Strausfeld, N. J. Genealogical correspondence of a forebrain centre implies an executive brain in the protostome-deuterostome bilaterian ancestor. *Philos Trans R Soc Lond B Biol Sci* **371,** 20150055 (2016).

100. Arendt, D. Genes and homology in nervous system evolution: Comparing gene functions, expression patterns, and cell type molecular fingerprints. *Theory Biosci* **124,** 185–197 (2005).

101. Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principles of Neural Science*. *McGraw-Hill*, USA (2013).

102. Müller, W. A. Autoradiographic Studies on the Synthetic Activity of Neurosecretory Cells in the Brain of *Platynereis Dumerilii* During Sexual Development and Regeneration. *Zeitschrift fur Zellforsch und mikroskopische Anat* **139,** 487–510 (1973).

103. Evans, S. M. & Downie, P. J. Decision-making processes in the polychaete *Platynereis dumerilii*. *Anim Behav* **34,** 472–479 (1986).

104. Evans, S. M. Non-associative avoidance learning in nereid polychaetes. *Anim Behav* **14,** 102–106 (1966).

105. Heisenberg, M. Mushroom body memoir: from maps to models. *Nat Rev Neurosci* **4,** 266–275 (2003).

106. Miura, H., Yanazawa, M., Kato, K. & Kitamura, K. Expression of a novel *aristaless* related homeobox gene '*Arx*' in the vertebrate telencephalon, diencephalon and floor plate. *Mech Dev* **65,** 99–109 (1997).

107. Friocourt. Mutations in *ARX* result in several defects involving GABAergic neurons. *Front Cell Neurosci* **4,** 1–11 (2010).

108. Hombría, J. C.-G. & Lovegrove, B. Beyond homeosis—HOX function in morphogenesis and organogenesis. *Differentiation* **71,** 461–476 (2003).

109. Banerjee-Basu, S. & Baxevanis, A. D. Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res* **29,** 3258–69 (2001).

110. Töhönen, V. *et al.* Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun* **6,** 8207 (2015).

111. Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125,** 497–521 (2016).

112. Galliot, B., de Vargas, C. & Miller, D. Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. *Dev Genes Evol* **209,** 186–97 (1999).

113. Melkman, T. *et al.* Regulation of chemosensory and GABAergic motor neuron development by the *C. elegans Aristaless*/*Arx* homolog *alr-1*. *Development* **132,** 1935–1949 (2005).

114. Gehring, W. J., Affolter, M. & Burglin, T. Homeodomain Proteins. *Annu Rev Biochem* **63,** 487–526 (1994).

115. Meijlink, F. *et al.* Vertebrate *aristaless*-related genes. *Int J Dev Biol* **43,** 651–63 (1999).

116. Wolanski, M., KhosrowShahian, F., Kelly, L. E., El-Hodiri, H. M. & Crawford, M. J. *xArx2*: An aristaless homolog that regulates brain regionalization during development in *Xenopus laevis*. *Genesis* **47,** 19–31 (2009).

117. Wigle, J. T. & Eisenstat, D. D. D. D. Homeobox genes in vertebrate forebrain development and disease. *Clin Genet* **73,** 212–226 (2008).

118. Chelly, J. & Mandel, J.-L. Monogenic causes of X-linked mental retardation. *Nat Rev Genet* **2,** 669–680 (2001).

119. Reish, O., Fullston, T., Regev, M., Heyman, E. & Gecz, J. A novel de novo 27 bp duplication of the *ARX* gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations. *Am J Med Genet Part A* **149A,** 1655–1660 (2009).

120. Bienvenu, T. *ARX*, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. *Hum Mol Genet* **11,** 981–991 (2002).

121. Kitamura, K. *et al.* Mutation of *ARX* causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat Genet* **32,** 359–369 (2002).

122. Strømme, P. *et al.* Mutations in the human ortholog of *Aristaless* cause X-linked mental retardation and epilepsy. *Nat Genet* **30,** 441–445 (2002).

123. Strømme, P., Mangelsdorf, M. E., Scheffer, I. E. & Gécz, J. Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in *Aristaless* related homeobox gene, *ARX*. *Brain Dev* **24,** 266–8 (2002).

124. Sayeed, O. & Benzer, S. Behavioral genetics of thermosensation and hygrosensation in *Drosophila*. *PNAS* **93,** 6079–84 (1996).

125. Manning, A. Antennae and Sexual Receptivity in *Drosophila melanogaster* Females. *Science* **158,** 136–137 (1967).

126. Göpfert, M, H. & Robert, D. The mechanical basis of Drosophila audition. *J Exp Biol* **202,** 2727–38 (2002).

127. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60,** 174–182 (2005).

128. Jansen, R., Van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43,** 1565–1575 (2002).

129. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide Sequence of the *iap* Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in *Escherichia coli*, and Identification of the Gene Product. *J Bacteriol* **169,** 5429–5433 (1987).

130. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).

131. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8,** 2281–2308 (2013).

132. Vassena, R. *et al.* Genome engineering through CRISPR/Cas9 technology in the human germline and pluripotent stem cells. *Hum Reprod Update* **22,** 411–419 (2016).

133. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for genome editing, regulation and targeting. *Nat Biotechnol* **32,** 347–355 (2014).

134. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).

135. Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat Rev Drug Discov* **16,** 387–399 (2017).

136. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339,** 823–826 (2013).

137. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

138. Du, D. & Qi, L. S. CRISPR technology for genome activation and repression in mammalian cells. *Cold Spring Harb Protoc* **2016,** 40–50 (2016).

139. Cho, S. W., Kim, S., Kim, J. M. & Kim, J. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31,** 230–232 (2013).

140. Ramos-Vara, J. A. & Miller, M. A. When Tissue Antigens and Antibodies Get Along. *Vet Pathol* **51,** 42–87 (2014).

141. Siddiqui, S. S., Aamodt, E., Rastinejad, F. & Culotti, J. Anti-tubulin monoclonal antibodies that bind to specific neurons in *Caenorhabditis elegans*. *J Neurosci* **9,** 2963–72 (1989).

142. Jackson, A. R., MacRae, T. H. & Croll, R. P. Unusual distribution of tubulin isoforms in the snail *Lymnaea stagnalis*. *Cell Tissue Res* **281,** 507–515 (1995).

143. Garcia-Arrarfis, J. E. & Viruet, E. Enteric nerve fibers of holothurians are recognized by an antibody to acetylated -tubulin. *Neurosci Lett* **157,** 153–156 (1993).

144. Wolf, N., Regan, C. L. & Fuller, M. T. Temporal and spatial pattern of differences in microtubule behaviour during *Drosophila* embryogenesis revealed by distribution of a tubulin isoform. *Development* **102,** 311–324 (1988).

145. Harzsch, S., Anger, K. & Dawirs, R. R. Immunocytochemical detection of acetylated alpha-tubulin and Drosophila synapsin in the embryonic crustacean nervous system. *Int J Dev Biol* **41,** 477–84 (1997).

146. Kapuscinski, J. DAPI: a DNA-Specific Fluorescent Probe. *Biotech Histochem* **70,** 220–233 (1995).

147. Jamur, M. C. & Oliver, C. in *Immunocytochemical Methods and Protocols. Methods in Molecular Biology* 63–66 (Humana Press, 2010).

148. Butt, D., O'Connor, S. J., Kuchel, R., O'Connor, W. A. & Raftos, D. A. Effects of the muscle relaxant, magnesium chloride, on the Sydney rock oyster (*Saccostrea glomerata*). *Aquaculture* **275,** 342–346 (2008).

149. Tian, L., Hires, S. A., Looger, L. L., Andrew Hires, S. & Looger, L. L. Imaging neuronal activity with genetically encoded calcium indicators. *Cold Spring Harb Protoc* **7,** 647–656 (2012).

150. Kerr, R. *et al.* Optical imaging of calcium transients in neurons and pharyngeal muscle of *C. elegans*. *Neuron* **26,** 583–594 (2000).

151. Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433,** 597–603 (2005).

152. Baker, P. F., Hodgkin, A. L. & Ridgway, E. B. Depolarization and calcium entry in squid giant axons. *J Physiol* **218,** 709–55 (1971).

153. Chen, T.-W. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).

154. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nat Biotechnol* **19,** 137–41 (2001).

155. Qi, W., Shui, B., Kotlikoff, M. I. & Sondermann, H. Structural Basis for Calcium Sensing by GCaMP2. *Structure* **16,** 1817–1827 (2008).

156. Akerboom, J. *et al.* Crystal Structures of the GCaMP Calcium Sensor Reveal the Mechanism of Fluorescence Signal Change and Aid Rational Design. *J Biol Chem* **284,** 6455–6464 (2009).

157. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31,** 227–229 (2013).

158. Chari, R., Yeo, N. C., Chavez, A. & Church, G. M. SgRNA Scorer 2.0: A Species-Independent Model to Predict CRISPR/Cas9 Activity. *ACS Synth Biol* **6,** 902–904 (2017).

159. Guhmann, M. *et al.* Spectral Tuning of Phototaxis by a Go-Opsin in the Rhabdomeric Eyes of *Platynereis*. *Curr Biol* **17,** 2265–2271 (2015).

160. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3,** e04766 (2014).

161. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol Reprod Dev* **82,** 518–529 (2015).

162. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33,** 661–667 (2015).

163. Hwang, W. Y. *et al.* Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS One* **8,** e68708 (2013).

164. Ansai, S. & Kinoshita, M. Targeted mutagenesis using CRISPR/Cas system in medaka. *Biol Open* **3,** 362–371 (2014).

165. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31,** 827–832 (2013).

166. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).

167. Xiang, G., Zhang, X., An, C., Cheng, C. & Wang, H. Temperature effect on CRISPR-Cas9 mediated genome editing. *J Genet Genomics* **44,** 199–205 (2017).

168. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods* **13,** 563–567 (2016).

169. Janeschik, M. Characterization of cholinergic and serotonergic neurons in the marine annelid *Platynereis dumerilii*. *University of Heidelberg,* Germany (2017).

170. Hussain, A. *et al.* High-affinity olfactory receptor for the death-associated odor cadaverine. *PNAS* **110,** 19579–19584 (2013).

171. Blenau, W. & Baumann, A. Molecular and pharmacological properties of insect biogenic amine receptors: Lessons from *Drosophila melanogaster* and *Apis mellifera*. *Arch Insect Biochem Physiol* **48,** 13–38 (2001).

172. Vouillot, L., Thélie, A. & Pollet, N. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3* **5,** 407–415 (2015).

173. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524,** 230–233 (2015).

174. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2013).

175. Rao, D. D., Vorhies, J. S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: Similarities and differences. *Adv Drug Deliv Rev* **61,** 746–759 (2009).

176. Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev Biol* **117,** 456–487 (1986).

177. Demayo, J. L., Wang, J., Liang, D., Zhang, R. & Demayo, F. J. Genetically Engineered Mice by Pronuclear DNA microinjection. *Curr Protoc Mouse Biol* **2,** 245–262 (2012).

178. Popova, E. A., Krivokharchenko, A. S. & Vil'yanovich, L. I. Investigation of the Effect of Various Buffer Solutions Used for Microinjection of Gene-Engineering Constructs on Preimplantation Development of Mouse Embryos. *Russ J Dev Biol* **33,** 110–112 (2002).

179. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10,** 1116–21 (2013).

180. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532,** 517–521 (2016).

181. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

182. Hirano, H. *et al.* Structure and Engineering of *Francisella novicida* Cas9. *Cell* **164,** 950–961 (2016).

183. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods* **11,** 705–706 (2014).

184. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154,** 1380–1389 (2013).

Appendix A

A.1 Human ARX protein sequence (Q96QS3), taken from UniProt database:

MSNQYQEEGCSERPECKSKSPTLLSSYCIDSILGRRSPCKMRLLGAAQSLPAPLTSRADPEKAVQGS
PKSSSAPFEAELHLPPKLRRLYGPGGGRLLQGAAAAAAAAAAAAAAAATATAGPRGEAPPPPPTAR
PGERPDGAGAAAAAAAAAAAAWDTLKISQAPQVSISRSKSYRENGAPFVPPPPALDELGGPGGVTHP
EERLGVAGGPGSAPAAGGGTGTEDDEEELLEDEEDEDEEEELLEDDEEELLEDDARALLKEPRRCPV
AATGAVAAAAAAAVATEGGELSPKEELLLHPEDAEGKDGEDSVCLSAGSDSEEGLLKRKQRRYRTTF
TSYQLEELERAFQKTHYPDVFTREELAMRLDLTEARVQVWFQNRRAKWRKREKAGAQTHPPGLPFP
GPLSATHPLSPYLDASPFPPHHPALDSAWTAAAAAAAAAFPSLPPPPGSASLPPSGAPLGLSTFLGAA
VFRHPAFISPAFGRLFSTMAPLTSASTAAALLRQPTPAVEGAVASGALADPATAAADRRASSIAALRLK
AKEHAAQLTQLNILPGTSTGKEVC

A.2 *P. dumerilii* transcriptome contig #18590, with the longest ORF underlined:

GGTCCGGAATTCCGGGATTGAATATCTTAGGATTGGGAGGATATACCATTGTTCATATATAGTTGC
TGGAATCAGGGACAAGTTGATGAAAAAAGAAATCTTGCAAAATCTGCTAGTGGGAAGCATCACCT
TATAATTGGCTAAAGCAATTAAGACTACATGGTAATTGCTAATTGCCATTGTGACGTCACCATGGA
TCTATCAATTAAGAAACAACGCACAGCTTACGACATAGCAAGCCTAATAGGCCCCGAAGCTACAA
AAGTTGAAAAATGGACTTCATCAGAAAGCCCCCCTACTTCTCCCCCGGGGAGACCCAGCTCAATT
TCCCGGCAGGATGACTTCGTCGCCAACTCTTCAGCCCCGGAGCATCCCCCTTCCCCCCTTAGTG
CTGATGACCAATACAAAAAACTGATGTCACCCACAATGCATCTGGCTCATGACGTCACAGCTCGA
ATCATTCTAGCCAATCACAGCCAAGAGTTGCAGTCACGTGATTATGCTTATGACCAATCACAAAGG
AGGCAAAGTGTGATGTCATCAGACAGTGCTGCCACCAGCCTGAGAAATGATGATATAGATGATGA
TGAACAGTTATGTATTGTGGATGATGATGAACCACTAAGTCCTGTAAATACCAGCCCAGTGCAGC
CCTCGGCCTCGATCCAGGCACCGAAAGAAGTCAAAGTCCAGGAAGACAACTTATCGATCTCCGA
TGGCGATGGCGACTTGGATGAAATGGGAAAACGCAAACAAAGACGATATCGAACAACGTTCACCT
CCTTCCAGTTAGAAGAATTAGAAAGGGCCTTCCAGAAGACGCACTACCCTGATGTGTTCACTAGA
GAGGAACTGGCCATGAGAATCAACCTAACCGAAGCGAGAGTTCAGGTTTGGTTTCAAAACAGAC
GAGCCAAGTGGAGGAAGAAGGAGAAAGTGGGGCCACAATCGCATCCCTACGGCCCTTTTGGGG
GCGGCCCTCCGACCGGGCCATTGGGCATCCCAGGGTTGGGCCAAAGACCCCTCGGCCCACACA
CCACTTACACAGACTTGTTGATCAAATCTTACGAAAACCATCTCGCTCAGAAGTTTGGTGGTCCAG
GTATGGGTGGATTCTATCCACCAGGATTAGGGGGACCGGTCGCTGCCATGGGAGTCTTTCCCGG
ATTGCCCTATTCCGGATTGGGTCATCTGCATCCGGGGAGTCACGGAGCGGCCGCAGTCGTAGCC
GGAAGCTTTCAGAGTCTCCTGGCGGCCATGTCGAATCCCGTCCCCGGTGGACAACCTCACTCTC
ACTCTCATCCTCATGCGCCCCCTGTGCAAGTTCCAACCTCGCCAAACGTAAACGTTGCTATTCCC
CAGGGTTCGCCGAACCTAACACCACAAAATTTACGGCGGCCTGGTTCGGAGTCACCAGATTCAA
ACAAAACGCCCTCCCCCCACTCACCCCCTGTGTTGTACCCTCAGCAGGACACCCCCAGAAGTGA
CGTCATCAAATCGAATCCAGAATCCCCCTCTGAGCATGCTCAAGACAGACGATCCTCTAGCATAG
CAGCATTGCGCATGCGTGCCAGGGAGTATGCCCTCAAGTTGCAGATGGGCCAACCATGTGACAG
TGCAGTCTATTGAGCAACTCCAGAATCATACTAAAGTTCTCTGATAGACTTAGGCCAATTACAGGC
TATTAAAGACTTAAATTACCTATTCAATTACCTATAAATGGATACACTGTCTAGGTGAATGACTATG
ATATACCTCTTGTGCTTATGTATACACTGTTTGCCTTTATAAAATAATTAAATTAAATTGCAAGAAAC
TTCGCAATTGCTGGTACTTAATTTAATATTTATGTGTTAAAATTGGTGTATTTTGTATTTTCTTTATTT
AACTTTTTATTTGAATATGTATTTTCAATATTCTGATTGTTATAATTAATCAATAGTAAGAATATTAAT
TATTAAACAGTGCAAACAGACTTGTTACTTCTTGCAGACTGAATATATTCTTCATAAGAATTTATTAT
GAAAACGTAAATTATGAATTGTACGTCCTTTTTACTTTTTTAAGTCAGTTGCTACTGCTATGGGCCA
ATGCCTAATAATTGTAACCAATTTGTTAATTTGTATGTTAGTGTACATATCATGTGACCTGTGCTAC
TTGTAAATATGTAATTCCAATGTTATAAATGGAAAATGTACGCAATTTTTGTATAAAGACTGTTTAAT
ATGTATGTTATGTAATTTGTACAGATCGTTCAAACAATATATCTGACTTAAAATGAAAAAAAAAAAA

A.3 Protein sequence correspondent to the longest ORF in *P. dumerilii* transcriptome contig #18590:

MVIANCHCDVTMDLSIKKQRTAYDIASLIGPEATKVEKWTSSESPPTSPPGRPSSISRQDDFVANSSAP
EHPPSPLSADDQYKKLMSPTMHLAHDVTARIILANHSQELQSRDYAYDQSQRRQSVMSSDSAATSLR
NDDIDDDEQLCIVDDDEPLSPVNTSPVQPSASIQAPKEVKVQEDNLSISDGDGDLDEMGKRKQRRYR
TTFTSFQLEELERAFQKTHYPDVFTREELAMRINLTEARVQVWFQNRRAKWRKKEKVGPQSHPYGPF
GGGPPTGPLGIPGLGQRPLGPHTTYTDLLIKSYENHLAQKFGGPGMGGFYPPGLGGPVAAMGVFPG
LPYSGLGHLHPGSHGAAAVVAGSFQSLLAAMSNPVPGGQPHSHSHPHAPPVQVPTSPNVNVAIPQG
SPNLTPQNLRRPGSESPDSNKTPSPHSPPVLYPQQDTPRSDVIKSNPESPSEHAQDRRSSSIAALRM
RAREYALKLQMGQPCDSAVY

A.4 Domain 3cmyA00 from CATH (homeodomain of the human Pax-3):

GQRRSRTTFTAEQLEELERAFERTHYPDIYTREELAQRAKLTEARVQVWFSNRRARWRKQ

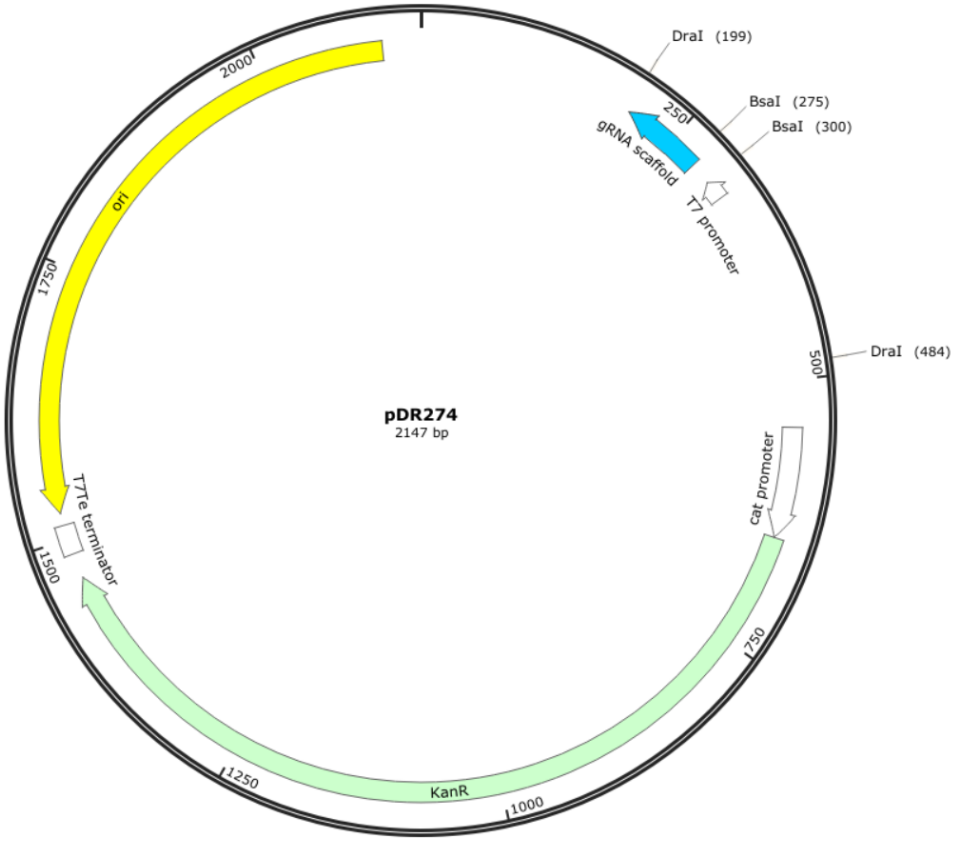
A.5 Alignment of the human Pax-3 homeodomain (Query; domain 3cmyA00 from CATH) with the predicted homeodomain of Arx (Sbjct):

Score	Expect	Method	Identities	Positives	Gaps
102 bits(255)	3e-32	Composition-based stats.	46/59(78%)	53/59(89%)	0/59(0%)

```
Query  2    QRRSRTTFTAEQLEELERAFERTHYPDIYTREELAQRAKLTEARVQVWFSNRRARWRKQ  60
            QRR RTTFT+ QLEELERAF++THYPD++TREELA R  LTEARVQVWF NRRA+WRK+
Sbjct  199  QRRYRTTFTSFQLEELERAFQKTHYPDVFTREELAMRINLTEARVQVWFQNRRAKWRKK  257
```

Appendix B

B.1 Scheme of plasmid pDR274, taken from snapgene.com, and the sequence of this plasmid encoding the sgRNA scaffold, from [157]:



+

TAATACGACTCACTATAGGAGAGACCGAGAGAGGGTCTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

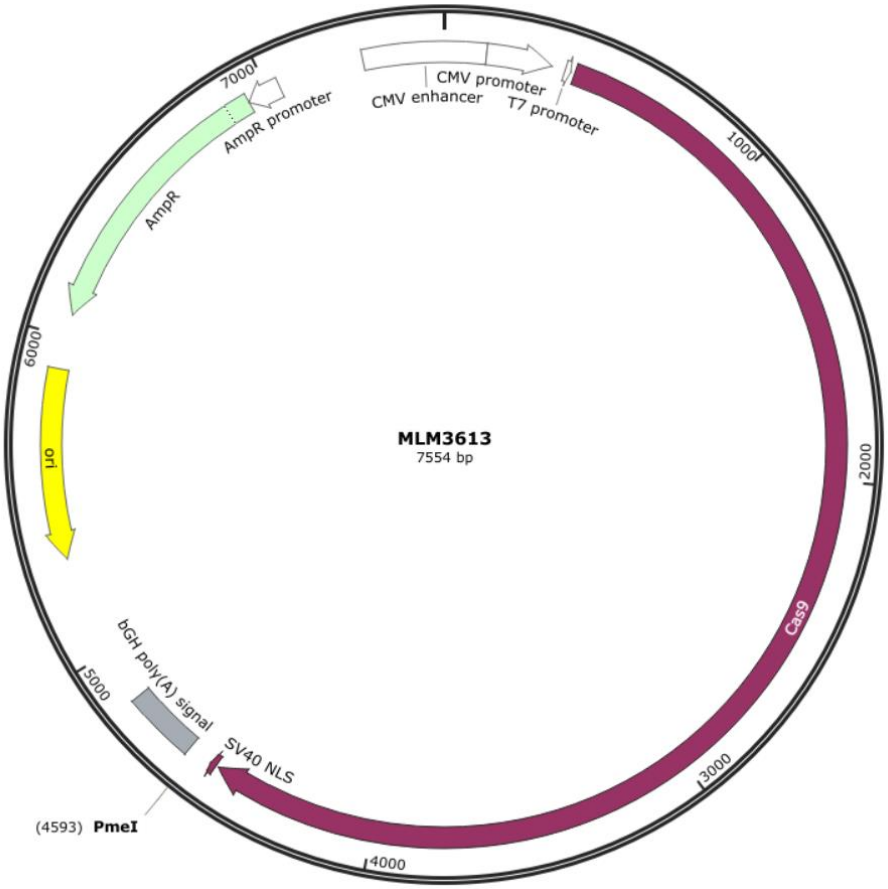
T7 promoter BsaI BsaI Guide RNA

CAACTTGAAAAAGTGGCACCGAGTCGGTGCT***TTTAAA***AAGCTTGGATCGACGAGAGCAGCGCGACTGGATCTGTCGCCCGTCTCAA

DraI

T7 promoter is underlined and the start site of transcription (+1) is marked with a + and highlighted in green. The pair of BsaI restriction sites used for cloning are indicated in bold text and the DraI restriction site used to linearize the plasmid for run-off transcription is shown in bold, italicized text. The “stuffer” sequence that is replaced with the annealed oligonucleotides to create customized sgRNAs is highlighted in yellow and the remainder of the full length sgRNA is shown as italicized underlined text.

B.2 Scheme of plasmid pMLM3613, taken from snapgene.com:



B.3 Cas9 protein sequence:

SNAMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY
FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSF
EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVEDPKKKR
KVD

In red is a double SV40 nuclear localization signal.

MW: 160947.3 Dalton