

Regulatory mechanisms of planarian regeneration



Jakke Neiro

The Queen's College
Department of Biology
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term
2024

Abstract

Regeneration of missing tissues requires regulatory mechanisms that ensure the coordinated behaviour of cells in both spatial and temporal dimensions. Planarians are highly regenerative flatworms capable of regenerating any missing body part, providing a suitable model system to understand these regulatory mechanisms. Planarian whole-body regeneration is enabled by a population of adult somatic pluripotent stem cells and a dynamic positional information system that instructs cell differentiation following injury. Although planarian regeneration and its molecular underpinnings have been studied for decades, the mechanisms of gene regulation during regeneration remain poorly understood. In this thesis, we investigate system-wide gene regulatory networks of planarian stem cells, while focusing on the role of TALE homeobox genes in planarian regenerative patterning. To understand the gene regulatory networks of planarian neoblasts, we comprehensively reannotate the genome of the planarian model species *Schmidtea mediterranea*, compile a list of its transcription factors, and discover enhancer-like regions active in stem cells by combining epigenomic evidence from histone modifications and chromatin accessibility. By integrating footprints of transcription factors in these enhancer-like regions, we construct emergent gene regulatory networks that provide testable hypotheses for how transcription factors maintain stemness and drive cell lineage differentiation in planarian stem cells. As our compilation of transcription factors identified multiple uncharacterised TALE homeobox genes within the genome of *S. mediterranea*, we continued to study the phylogeny of the TALE gene repertoire. Our analysis revealed that *S. mediterranea* possesses both evolutionarily conserved and recently evolved lineage-specific TALE genes, including both spiralian- and platyhelminth-specific genes. Among the conserved TALE genes, *pbx-1* and *prep* had been previously implicated in axial regeneration, and we discovered that an additional TALE factor of the Meis family, *meis-2*, is required for posterior regeneration. To infer the targets regulated by these TALE factors, we performed transcriptomic and epigenomic profiling following knockdowns. Our analysis of sequencing data, in combination with preliminary functional experiments and existing planarian literature, suggests that *pbx-1* and *meis-2* regulate as a dimer the posterior pole marker *wnt1* and other genes of the posterior specification program, while *prep* and *pbx-1* regulate pole markers and FGFR family members in the anterior blastema. In conclusion, these findings provide one of the first insights into the gene regulatory networks of planarian regenerative abilities, offering perspectives for understanding the evolution of animal stem cells and body axis patterning, and potential implications for regenerative medicine.

Declaration and statement of authorship

I hereby declare that this thesis entitled *Regulatory mechanisms of planarian regeneration* has been conducted by myself under the supervision of Professor Aziz Aboobaker at the Department of Biology, University of Oxford. The contributions of other individuals are transparently and explicitly stated, and the work of others has been clearly cited.

The research presented in Chapter 2 has been published as an original research article in *eLife* (2022) with myself as first and corresponding author. All the work for the manuscript has been conducted during my time as a DPhil student at the University of Oxford. The contributions of coauthors are stated at the beginning of this chapter.

I confirm that this thesis has not been submitted for any other qualification or degree at this university or any other institution. The particulars given in the thesis are true and correct to the best of my knowledge.

A handwritten signature in black ink, appearing to read 'Jakke Neiro', with a long horizontal stroke extending to the right.

Jakke Neiro
The Queen's College
12th October 2024

Contents

Abstract	2
Declaration and statement of authorship.....	3
Contents	4
List of publications	8
Chapter 1. General introduction	9
1.1 Principles of planarian regeneration and gene regulation	9
1.1.1 Regenerative biology is set for a renaissance	9
1.1.2 Planarians remain a <i>primus motor</i> of regenerative biology	11
1.1.3 Planarians have adult pluripotent stem cells that drive regeneration.....	12
1.1.4 Sequencing technologies are used to infer gene regulatory networks.....	14
1.2 TALE factors in metazoans.....	16
1.2.1 TALE factors are a subclass of homeobox genes.....	16
1.2.2 TALE factors possess conserved functional domains	18
1.2.3 MEINOX and PBC proteins form dimers	19
1.2.4 Nuclear localisation controls TALE activity.....	20
1.2.5 TALE proteins are Hox cofactors.....	20
1.2.6 TALE factors also interact with non-HOX components.....	22
1.2.7 TALE proteins are modulators of the epigenomic landscape	24
1.2.8 The DNA binding sites of TALE factors are well characterised	24
1.2.9 TALE factors regulate the cell cycle and cell physiology.....	25
1.2.10 TALE factors play diverse roles in metazoan development	26
1.2.11 TALE factors have diversified in spiralian.....	29
1.2.12 Summary of TALE factors	31
1.3 Patterning in planarian regeneration.....	31
1.3.1 Pattern formation in the context of regeneration	31
1.3.2 <i>Entwicklungsmechanik</i> of planarians informs on pattern formation.....	32
1.3.3 A molecular coordinate system provides positional information in planarians	33
1.3.4 Signalling pathways in planarian regeneration	34
1.3.5 Regeneration cascade	38
1.3.6 Planarian eye regeneration is self-organising.....	43
1.3.7 Summary of planarian regeneration.....	43
1.4 Aims of thesis.....	44
Chapter 2. Identification of putative enhancer-like elements predicts regulatory networks active in planarian adult stem cells	47
2.1 Introduction	47

2.2 Results	49
2.2.1 Annotation with a full range of transcriptome samples identifies more than 3000 new protein coding genes in the <i>Schmidtea mediterranea</i> genome.....	49
2.2.2 Comprehensive annotation of planarian transcription factors highlights a diversity of unknown zinc fingers	52
2.2.3 Histone modifications and chromosome accessibility mark enhancer-like regions	55
2.2.4 Gene regulatory networks involving fate-specific transcription factors in neoblasts	55
2.2.5 Multiple enhancers are linked to unstudied transcription factors	61
2.2.6 Elements of the planarian positional gene regulatory network are active in neoblasts.....	63
2.2.7 Available transcriptomic data supports putative gene regulatory networks	66
2.3 Discussion.....	68
2.4 Methods	70
2.4.1 Reference assembly and annotations	70
2.4.2 Expression values, homology, and GO.....	70
2.4.3 Identification and characterization of transcription factors	71
2.4.4. Motifs of transcription factors	71
2.4.5 ATAC-seq library preparation.....	71
2.4.6 ChIP-seq library preparation and sequencing.....	72
2.4.7 ChIP-seq data analysis	72
2.4.8 ATAC-seq data analysis and motif footprinting	72
2.4.9 Enhancer targets and gene regulatory networks	72
2.4.10 Functional transcriptomics and validation of gene regulatory network predictions	73
2.4.11 Code availability.....	73
2.4.12 Data availability	73
2.5 Supplementary figures.....	74
2.6 Supplementary material	79
2.7 References.....	80
Chapter 3. Phylogenetic analysis of TALE homeobox genes in <i>Schmidtea mediterranea</i> reveals conservation and lineage-specific evolution	91
3.1 Introduction	91
3.2 Results	92
3.2.1 <i>S. mediterranea</i> possesses 15 TALE genes from both conserved and divergent families	92
3.2.2 Analysis of the SPILE clade reveals the presence of SPILE-E and SPILE-Z orthologues in <i>S. mediterranea</i>	97
3.2.3 Lineage-specific TALE factors are widely present in platyhelminth species.....	99
3.2.4 TALE factors of <i>S. mediterranea</i> contain both conserved and novel protein domains	101
3.2.5 TALE factors are expressed during both early development and organogenesis.....	103

3.2.6 Functional characterisation of the platyhelminth-specific <i>PALF</i> and spiralian-specific <i>SPLE-Z1</i> uncovers roles in regenerative patterning and adult biology.....	106
3.3 Discussion.....	108
3.4 Methods	109
3.4.1 Phylogenetic tree construction.....	109
3.4.2 Functional domain analysis.....	110
3.4.3 Expression analysis	110
3.4.4 Planarian culture	110
3.4.5 Cloning.....	110
3.4.6 RNA interference	111
3.4.7 In situ hybridisation	112
3.5 Supplementary material	113
Chapter 4. A functional genomic approach infers TALE-dependent and TALE-independent regulatory mechanisms in planarian axial regeneration	135
4.1 Introduction	135
4.2 Results	136
4.2.1 <i>meis-2</i> is required for posterior regeneration	136
4.2.2 Anterior pole markers and FGFR1 genes are downregulated after knockdown of <i>pbx-1</i> and <i>prep</i>	138
4.2.3 The posterior specification program is downregulated after knockdown of <i>pbx-1</i> and <i>meis-2</i>	142
4.2.4 Chromatin accessibility and footprinting analysis predicts direct targets of planarian TALE genes.....	145
4.2.5 Functional characterisation and RNAi screening informs on the regulatory mechanisms of TALE factors.....	147
4.2.6 Footprinting analysis infers direct regulatory interactions in the anterior blastema.....	151
4.2.7 Footprinting analysis infers elements of the posterior gene regulatory network.....	154
4.3 Discussion.....	157
4.4 Methods	158
4.4.1 Planarian culture, cloning, and RNA interference	158
4.4.2 RNA sequencing.....	158
4.4.3 ATAC-seq.....	158
4.4.4 RNA seq analysis.....	160
4.4.5 ATAC-seq analysis	160
4.4.6 In situ hybridisation and immunostaining.....	162
4.5 Supplementary material	163
Chapter 5. General discussion	171
5.1 Synopsis.....	171

5.2 Bioinformatics for the planarian community.....	171
5.3 Evolution of animal stem cells.....	172
5.4 TALE factors and their tale on molecular evolution.....	173
5.5 TALE factors in the shadow of the Wnt gradient.....	174
5.6 The enigma of anterior regeneration – does <i>non-darake</i> regulate self-organising cell adhesion?.....	174
5.7 Coda.....	175
Acknowledgements.....	176
References	177

List of publications

Work for the following chapter has resulted in a publication:

Chapter 2

Neiro, J., Sridhar, D., Dattani, A., & Aboobaker, A. (2022). Identification of putative enhancer-like elements predicts regulatory networks active in planarian adult stem cells. *Elife*, 11, e79675.

In addition, the reannotation of the genome of *Schmidtea mediterranea* presented in this publication, along with the reannotation of the genome of *Dugesia japonica*, also contributed to the following publication.

García-Castro, H., Kenny, N. J., Iglesias, M., Álvarez-Campos, P., Mason, V., Elek, A., Schönauer, A., Sleight, V. A., **Neiro, J.**, Aboobaker, A., Permanyer, J., Irimia, M., Sebé-Pedrós, A., & Solana, J. (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome biology*, 22(1), 89.

Chapter 1. General introduction

This thesis studies gene regulatory networks in planarian stem cells, the evolution of planarian TALE homeobox genes, and the regulatory function of the TALE genes *prep*, *pbx-1*, and *meis-2* in planarian regenerative patterning. First, the field of regenerative biology, the planarians as a model system, planarian adult pluripotent stem cells, and sequencing technologies used to infer gene regulatory networks are introduced (1.1 Principles of planarian regeneration and gene regulation). This serves as background to all the result chapters and the published second chapter that examines the gene regulatory networks active in planarian stem cells. Second, the literature on TALE factors is reviewed to summarise our knowledge of their evolution and their role as transcription factors and regulators of gene activity in metazoan development (1.2 TALE factors in metazoans). This serves as background to the third chapter, which studies the phylogeny of planarian TALE genes, and the fourth chapter, which investigates the regulatory roles of TALE factors in regenerative patterning. To further understand the regulatory context of TALE factors in the fourth chapter, the planarian positional information system and the molecular underpinnings of the regeneration process from amputation to healing are reviewed in the third part (1.3 Patterning in planarian regeneration). The research objectives are clarified in the final section of the introduction (1.4 Aims of thesis).

1.1 Principles of planarian regeneration and gene regulation

1.1.1 Regenerative biology is set for a renaissance

Regeneration is the replacement or functional restoration of injured or lost structures in adult multicellular organisms. It can occur at different structural levels: axons repair at the cellular level, the intestinal epithelium renews at the tissue level, the heart in zebrafish heals at the organ level, and planarians are capable of whole-body regeneration (Bely and Nyberg, 2010). Specifically, restorative or injury-induced regeneration refers to events of traumatic injury or sudden tissue loss, while the continuous turnover of cells and tissues throughout the life of an organism is sometimes included in the definition of regeneration and denoted as homeostatic or physiological regeneration. Historically, regenerative events were classified as either epimorphosis or morphallaxis. In this classification, epimorphic regeneration is driven by cell proliferation, whereas morphallactic regeneration occurs through the remodeling of pre-existing cells and tissues without an increase in cell number (Morgan, 1901b). However, since many regenerative events include both cell proliferation and remodeling such as in planarians, modern regenerative biology classifies regenerative mechanisms based on the origin of regenerated cells: stem cells, dedifferentiation, and transdifferentiation (Poss and Tanaka, 2024).

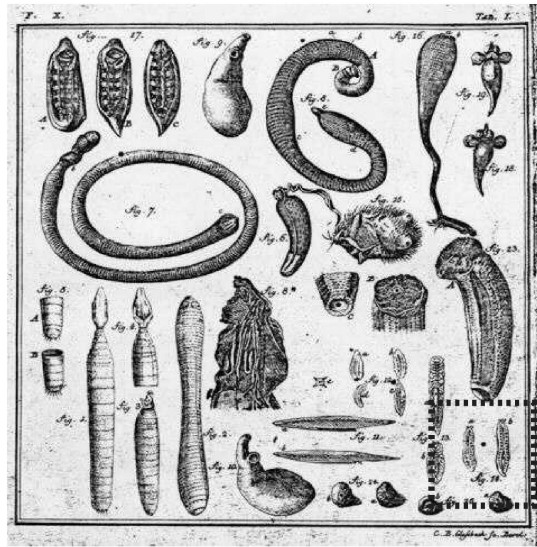
The conundrum of regeneration has intrigued biologists since the dawn of biology itself; the first humans likely observed which injuries regenerate and which do not. In Ancient Greece and Rome, Aristotle and Pliny described some cases of regeneration, but it was Trembley (1744) who led the rediscovery of the field by describing regeneration in hydra. Trembley's findings sparked a series of further studies: Bonnet (1745) demonstrated that some annelid worms regenerate, while Spallanzani's *Prodromo* (1768) described regeneration in tadpoles, salamanders, and snails. As for planarians, old Chinese manuscripts from the 800s hint at regeneration by describing fission in flatworms (Elliott and Sánchez Alvarado, 2013). In the West, Pallas (1774) encountered planarians in the Ural mountains (Figure 1.1B), and to Dalyell (1814) planarians were "almost to be called immortal under the edge of the knife". However, it was T. H. Morgan (Figure 1.1D) who established planarians as a scientific model system in his influential work from 1898 to 1905 before famously turning to fly genetics.

Regeneration remains one of the greatest unanswered questions in biology. Does regeneration recapitulate development? How has regeneration evolved, and why do some animals regenerate, while others do not? Are regenerative abilities adaptive traits in themselves or emergent epiphenomena of other underlying

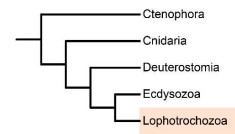
A *Schmidtea mediterranea*



B Pallas 1774



C

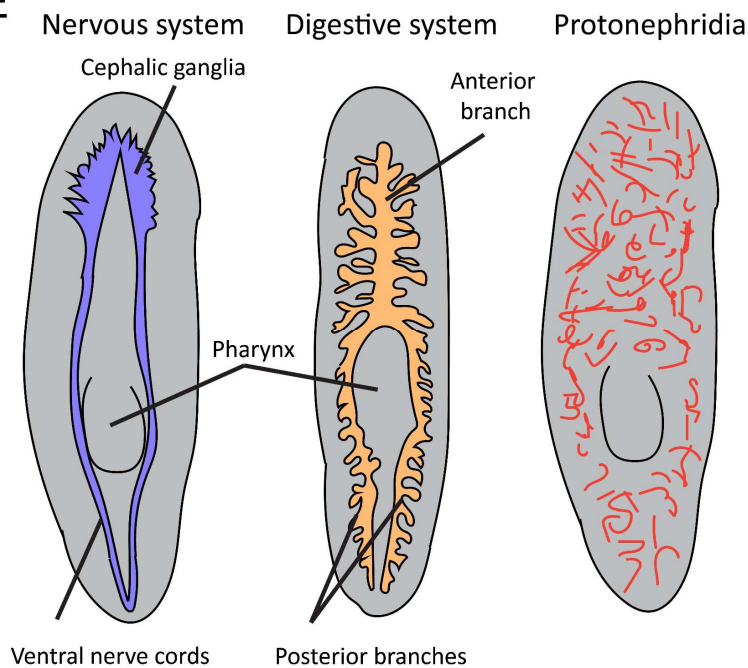


D

T. H. Morgan



E



F

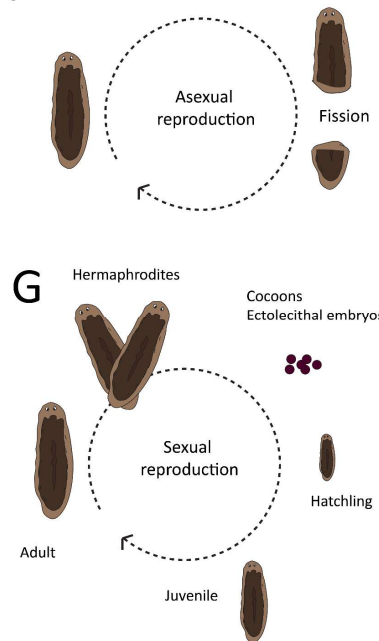


Figure 1.1 Introduction to planarians. (A) Image of the main modern planarian model species, *Schmidtea mediterranea*. (B) Illustration from Pallas (1774) *Spicilegia Zoologica*. Two planarians are depicted in the lower right corner. (C) Planarians belong to the lophotrochozoans or spiralian in the metazoan phylogeny. (D) T. H. Morgan established planarians as a modern model system between 1898 and 1905 (Shine and Wrobel 2014). (E) Schematic illustrations of the planarian nervous system, the digestive system, and protonephridia. (F) Asexual reproduction in planarians occurs through fission. (G) In sexual strains, hermaphrodites copulate and lay cocoons.

processes? What are our limits to regenerative abilities? The existence of regenerative abilities in all metazoan basal taxa suggests that regeneration is an ancestral trait of all animals (Bely and Nyberg, 2010; Lai and Aboobaker, 2018). Recently, there has been a renaissance in regenerative biology from an evolutionary or comparative perspective, as new model species are emerging or re-emerging (Lai and Aboobaker, 2018). In the past two decades, planarians have entered the age of molecular genetics and remain as a primary driving force of the field of developmental and regenerative biology. In this thesis, I aim to bring the planarians into the age of “regulomics” and understand the gene regulatory networks of the two robust aspects of planarian whole-body regeneration: adult pluripotent stem cells and the positional information system that patterns regenerating structures.

1.1.2 Planarians remain a *primus motor* of regenerative biology

Planarians are flatworms (order Tricladida, phylum Platyhelminthes, superphylum Lophotrochozoa, Figure 1.1A, C) that inhabit both marine and freshwater ecosystems, as well as terrestrial habitats (Sluys and Riutort, 2018). They are bilateral, triploblastic and acoelomate, with a complex internal anatomy (Figure 1.1E) (Sluys and Riutort, 2018). The central nervous system consists of the bilobed, arch-shaped cephalic ganglia (i.e. the brain) and two ventral nerve cords (VNCs) (Ross et al., 2017). The cephalic ganglia are connected by an anterior commissure, and lateral branches project towards the margin (Ross et al., 2017). The VNCs extend to posterior end and are interconnected via transversal commissures along the length of the worm (Ross et al., 2017; Pascual-Carreras, 2020). The brain includes different neuronal subpopulations: serotonergic, dopaminergic, glutamatergic, GABAergic, and cholinergic neurons (Ross et al., 2017; Pascual-Carreras, 2020). Photoreceptors are located dorsally and include rhabdomeric photoreceptor neurons and pigmented optic cup cells (Reddien et al., 2021). Neuronal projections from the photoreceptors form an optic chiasm (Scimone et al., 2020).

A three-branched digestive system with an anterior branch and two posterior branches distributes nutrients ingested by the muscular pharynx in the worm's central region (Figure 1.1E) (Forsthöfel et al., 2011). The pharynx is a cylindrical muscle that connects to the intestine through the esophagus and serves as both mouth and anus (Forsthöfel et al., 2011). The monostratified epithelium of the gut includes absorptive phagocytes and secretory goblet cells that are surrounded by the enteric musculature (Forsthöfel et al., 2011; Pascual-Carreras, 2020). The excretory and osmoregulatory system is composed of protonephridia that are scattered throughout the worm (Figure 1.1E) (Scimone et al., 2011; Pascual-Carreras, 2020). Protonephridia consist of flame cells, tubule cells, and the collecting duct (Scimone et al., 2011; Pascual-Carreras, 2020).

The planarian epidermis secretes protective mucus and contains non-ciliated and ventral ciliated cells used for locomotion (Cheng et al., 2018; Pascual-Carreras et al. 2020). The three-layered body wall musculature contains longitudinal, circular, and diagonal fibers and is important for locomotion but also for providing positional information during regeneration and secreting extracellular matrix (ECM) components (Scimone et al., 2017; Cote et al., 2019). Planarians lack circulatory and respiratory systems, and oxygen diffuses through the epidermis to internal organs (Lewallen and Burggren, 2020). Planarians have both asexual and sexual strains; obligate asexual worms lack reproductive organs and reproduce through fission, while sexual strains are cross-fertilising hermaphrodites with testes and ovaries (Figure 1.1F, G) (Newmark et al., 2008).

Today, the main planarian model species is *Schmidtea mediterranea* (Figure 1.1A), followed by *Dugesia japonica*, *Girardia tigrina*, and *Schmidtea polychroa* (Reddien and Sánchez Alvarado, 2004; Agata et al., 2003; Wurtzel et al., 2015; Booth et al., 2024), but less regenerative planarian species such as *Dendrocoelum lacteum* and *Phagocata kawakatsui* have also been used for comparative studies of regeneration competence (Liu et al., 2013; Sikes et al., 2013; Umesono et al., 2013). The advantages of planarians as a model system include easy husbandry and an expanding molecular toolkit: the genome of *S. mediterranea* (800 Mb, 4 chromosomes) has been sequenced and annotated (Grohme et al., 2018), robust functional knockdowns are possible through RNAi

(Sánchez Alvarado and Newmark, 1999), and gene expression can be revealed by *in situ* hybridisation and protein localisation by immunohistochemistry (King et al., 2013; Forsthoefel et al., 2018). However, planarians lack tools readily available in other systems: transgenesis, genetic lineage tracing, gene editing, overexpression approaches and stem cell culture systems are still unavailable.

1.1.3 Planarians have adult pluripotent stem cells that drive regeneration

Planarian regeneration is based on adult somatic pluripotent stem cells called neoblasts (Figure 1.2A, B). Here, pluripotency is defined as the capacity to give rise to all somatic cell types and germ cells (Hanna et al. 2010; Varley et al., 2023). Neoblasts are the only cycling cells of the adult, enable indefinite self-renewal, and can differentiate to all somatic cell types of the adult body plan, and most likely also the germ cells in sexual strains (Figure 1.2B) (Baguña et al., 1989; Newmark and Sánchez Alvarado, 2000; Aboobaker, 2011; Wagner, 2011). The ability to differentiate to all somatic lineages was demonstrated when a single transplanted neoblast from an asexual donor could rescue regeneration in a neoblast-depleted sexual host (Wagner et al., 2011). The ability to give rise to germ cells has been inferred from the regenerative capacity of head fragments in sexual worms, showing that fragments lacking germ cells can regenerate the sexual body along with the germline (Morgan, 1901a; Sato et al., 2006).

Neoblasts are driving both homeostatic cell turnover and injury-induced regeneration. The cells are abundantly scattered in the mesenchyme and constitute 20-30% of all planarian cells (Figure 1.2A) (Baguña and Romero, 1981; García-Castro et al., 2021). The neoblast cytoplasm contains chromatoid bodies that are intricate ribonucleoprotein complexes (Aboobaker, 2011). Additionally, neoblasts have a distinct morphology and a large nucleus and a high nuclear:cytoplasmic ratio (Figure 2.1B). This ratio is used in fluorescence-activated cell sorting (FACS) to separate planarian cells into three compartments: X1 cells that are neoblasts in the S/G2/M phases of the cell cycle, X2 cells that are neoblasts in the G1 phase and post-mitotic progeny, and Xins cells (insensitive to irradiation) that are differentiated cells (Hayashi et al., 2006; Labbé et al., 2012; Önal et al., 2012; Romero et al., 2012; Solana et al., 2012). Thus, X1 cells are commonly used to represent neoblasts for experimental reasons. Since neoblasts are more sensitive to ionising radiation than differentiated cells, this property is routinely used to specifically ablate stem cells in planarians.

Neoblasts express a set of markers mostly linked to the cell cycle, chromatoid bodies or germ plasm, and chromatin remodelers (Shibata et al., 1999; Reddien et al., 2005b; Salvetti et al., 2005; Guo et al., 2006; Rossi et al., 2006; Oviedo and Levin, 2007; Rossi et al., 2007; Yoshida-Kashikawa et al., 2007; Eisenhoffer et al., 2008; Palakodeti et al., 2008; Solana et al., 2009; Rouhana et al., 2010; Scimone et al., 2010). The canonical markers linked to the chromatoid bodies in planarians or other animals include e.g. *piwi-1*, *piwi-2*, *piwi-3*, *vasa*, *tudor*, *bruli*, *pumilio*. These genes are associated with RNA binding and metabolism in the germline of many organisms; the Piwi/Argonaute gene family is required for the synthesis of piRNAs that mediate transposon silencing, while *bruli* and *pumilio* are RNA-binding proteins required for germline development in flies (Lin and Spradling, 1997; Aboobaker, 2011). Therefore, neoblasts express animal germ plasm components but are not germline stem cells, as these are a separate planarian cell type expressing unique markers such as *nanos* (Wang et al., 2007). This has inspired the idea of a common germline multipotency program (GMP) shared between the germline and adult stem cells (Juliano et al., 2010). According to the primordial stem cell hypothesis, adult stem cells are also part of the germline, implying that the Weissmann barrier would not be positioned between the germline and soma, but rather between “maintainers” in general (adult stem cells and germline) and differentiated cells (Solana, 2013).

Originally, neoblasts were defined as a morphologically uniform cell population. However, with the advent of transcriptomics, individual neoblasts were found to be highly heterogeneous in their gene expression profiles. The neoblast population contains subpopulations of neoblasts expressing transcription factors

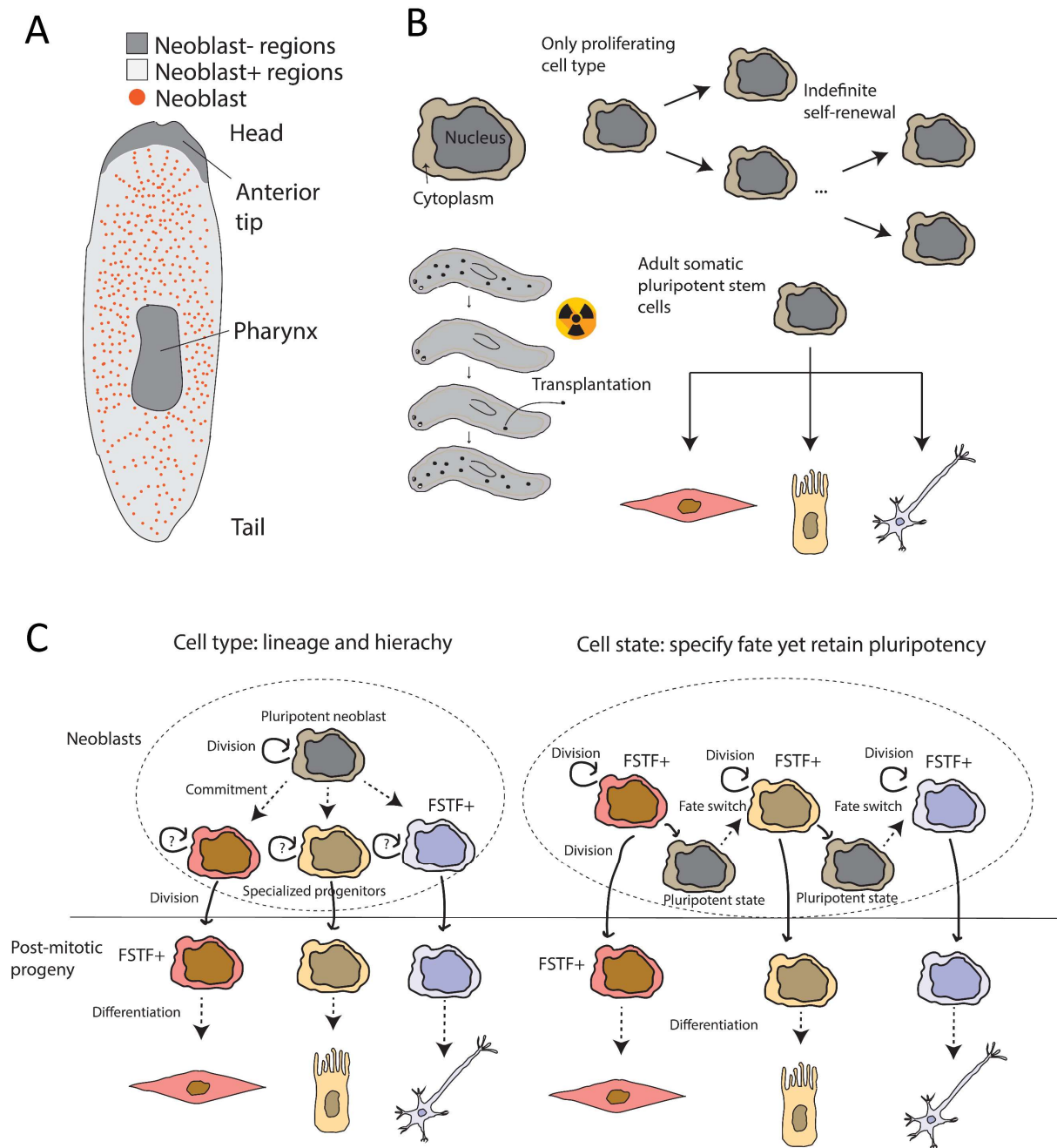


Figure 1.2 Introduction to neoblasts. (A) Neoblasts are scattered throughout the mesenchyme of the worm, except in the most anterior region and the pharynx. (B) Schematic illustrations of neoblast properties: neoblasts have a high nuclear:cytoplasmic ratio, are the only proliferating cell type, and are capable of indefinite self-renewal. Transplantation of a single neoblast can rescue irradiated (neoblast-depleted) worms, and neoblasts are able to differentiate into all cell types in the adult worm. (C) Two main models of neoblast pluripotency are depicted as simplified schematics. In the “hierarchical model”, a subset of neoblasts constitutes a pluripotent subpopulation that gives rise to specialised neoblasts expressing fate-specific transcription factors. In the single-step fate model, specialised neoblasts divide asymmetrically and give rise to a fate-specified daughter and a daughter cell that undergoes fate switch into another lineage.

specific for differentiated cell fates (FSTFs, fate-specific transcription factors), including those for eyes (Lapan and Reddien, 2011; 2012), neural subtypes (Cowles et al., 2013; Currie and Pearson, 2013; März et al., 2013; Cowles et al., 2014; Scimone et al., 2014b, Molinaro and Pearson, 2016; Roberts-Galbraith et al., 2016; Ross et al., 2017, 2018), pharyngeal cells (Adler et al., 2014), protonephridial cells (Scimone et al., 2011), muscle cells (Scimone et al., 2017, 2018), anterior pole cells (Scimone et al., 2014a, Vázquez-Doorman and Petersen, 2014; Vogg et al., 2014), and parenchymal cells (Fincher et al., 2018). Recently, these transcriptomic profiling efforts have been further augmented by an ever-increasing collection of single cell RNA-sequencing atlases (Fincher et al., 2018; Plass et al., 2018; Zeng et al., 2018; Benham-Pyle et al., 2021; García-Castro et al., 2021). Thereby, the presence of neoblasts specialised for certain fates has raised the question of whether any of the subclasses is at the top of the differentiation hierarchy and constitutes a population of naïve stem cells that give rise to all other subclasses. An earlier pioneering single-cell multiplexed qPCR approach subdivided neoblasts into σ , ζ , and γ ; ζ -neoblasts are epidermis-specialised, γ -neoblasts are intestine-specialised, while σ -neoblasts were hypothesised to harbor the naïve or pluripotent neoblasts (van Wolfswinkel et al., 2014). In a subsequent scRNA-seq study, neoblasts enriched for the cell surface marker tetraspanin-1 (TSPAN-1) and the tetraspanin group specific gene 1 (*tgs-1*) were observed to have higher rescue potential in a neoblast-depleted background, suggesting that these could constitute the source of pluripotency (Zeng et al., 2018).

On the other hand, there might not be a single subclass to rule them all (Raz et al., 2021; Reddien et al., 2022). Indeed, Raz et al. (2021) discovered that different FSTF-expressing neoblast subpopulations likely harbour pluripotent neoblasts, or at least neoblasts not restricted to a single germ layer, i.e. FSTF-expressing cells can express different FSTFs in subsequent rounds of division (Raz et al., 2021). Furthermore, neoblasts expressing *tgs-1* were discovered to express many neural precursor markers (Raz et al., 2021). These experimental findings led to the single-step fate model, in which specialised neoblasts can divide asymmetrically and give rise to a fate-specified post-mitotic daughter cells and a daughter cell that is able to switch fate and specialise into another fate (Raz et al., 2021; Reddien et al., 2022). FSTFs are expressed mainly in the S/G2/M phases of the neoblast cell cycle, but less so in the G1 phase, suggesting that the fate-specific expression program becomes reset before the following cycle. Therefore, flexible and dynamic fate switching rather than a specific pluripotent subclass explains pluripotency of the whole neoblast population in the single-step fate model (Raz et al., 2021). However, how fate switching occurs, how the fate-specification apparatus is reprogrammed in neoblasts, and how the single-step fate model maintains stemness in the neoblast population over time are yet to be clarified.

Lastly, the epigenomic state and chromatin dynamics of neoblasts and their progeny are central to maintaining self-renewal and differentiation. Neoblasts are enriched for both histone-modifying enzymes and chromatin remodeling enzymes (Önal et al., 2012; Solana et al., 2012), and neoblast function is perturbed following knockdown of NuRD complex components (Scimone et al., 2010; Jaber-Hijazi et al., 2013; Vázquez-Doorman and Petersen, 2016), COMPASS and COMPASS-like families (Hubert et al., 2013; Duncan et al., 2015; Mihaylova et al., 2018), CBP/p300 homologues (Stelman, 2021), and BAF complex components (Wiggans et al., 2023). To avoid cellular senescence, asexual planarians have also evolved somatic telomerase activity to maintain telomeres (Tan et al., 2012). Altogether, it appears that many aspects of stem and germ cells are shared between planarians and other model systems in terms of RNA and genome biology, whereas transcription factors that confer pluripotency appear not to have been conserved.

1.1.4 Sequencing technologies are used to infer gene regulatory networks

Gene regulation is the regulation of gene expression in quantity, time, and space. In broad terms, gene regulation includes transcriptional regulation, pre-mRNA processing, translational regulation of mRNA, and post-translational modifications of the translated protein. Transcriptional regulation involves the coordinated activity of the transcription machinery, transcription factors, and non-coding sequences or cis-regulatory genomic regions (CREs), which include the promoter, enhancers, silencers or repressors, and

insulators (Kim and Wysocka, 2023). At the promoter, the enzyme RNA polymerase II (RNA Pol II) and basal or general transcription factors (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIH) form the basal transcriptional machinery, the preinitiation complex (PIC), which transcribes the gene (Soutourina, 2018). Specific transcription factors bind to transcription factor binding sites (TFBS) in enhancers and silencers, recruit co-regulators such as chromatin remodelers and histone-modifying enzymes, loop the chromatin to bring together the enhancer and the promoter, and modulate the activity of the PIC through the Mediator complex (Soutourina, 2018; Kim and Wysocka, 2023). Specific transcription factors can be classified with respect to their DNA binding domains, including homeodomain proteins, basic-helix-loop-helix factors, and zinc fingers (Lambert et al., 2018).

A system of transcription factors regulating their target genes is often presented and modelled as a gene regulatory network (GRN) (Karlebach and Shamir, 2008; Peter and Davidson, 2015; Badia-i-Mompel et al., 2023). Mathematically, GRNs can be modelled as discrete logical models (e.g. Boolean networks), continuous models (e.g. ordinary differential equations), and single-molecule level models (Karlebach and Shamir, 2008). In developmental biology, the concept of GRNs was pioneered by Eric H. Davidson, who utilised expression patterns, perturbation experiments and analysis of cis-regulatory elements to construct a logical GRN model capturing the specification of cell lineages in the sea urchin embryo (Britten and Davidson, 1969; Davidson et al., 2002; Davidson and Erwin, 2006; Davidson and Levine, 2008; Peter and Davidson, 2015). Overall, the pre-omic era recognised that constructing a GRN requires the identification of its components, namely transcription factors (TFs) and cis-regulatory elements. Thus far, identifying transcription factors has been simpler due to their conserved protein-coding domains, whereas discovering active cis-regulatory elements within the genome has been more challenging.

When transcriptomics came of age, methods relying on coexpression matrices were used to infer regulatory networks (Langfelder and Horvath, 2008; Huynh-Thu et al., 2010). However, as these approaches do not identify direct cis-regulatory interactions, methods were developed to detect TF binding. ChIP-seq (chromatin immunoprecipitation followed by sequencing) and CUT&TAG (cleavage under targets and tagmentation) can be used to profile the binding of individual TFs in the genome, but these techniques are restricted to TFs for which antibodies are available (Johnson et al., 2007; Kaya-Okur et al., 2019). Therefore, chromatin signatures, in combination with TF motifs, have been used to infer cis-regulatory regions on a larger scale. Enhancer-associated histone modifications, such as acetylated lysine 27 of histone H3 (H3K27ac) and lysine 4 on histone 3 (H3K4me1), can also be identified by using ChIP-seq (Shlyueva et al., 2014), while open chromatin can be identified by using DNase-seq (DNase I hypersensitive site sequencing) (Boyle et al., 2008), FAIRE-seq (formaldehyde-associated isolation of regulatory elements followed by sequencing) (Giresi et al., 2007), MNase-seq (micrococcal nuclease sequencing) (Yuan et al., 2005), and ATAC-seq (Assay for transposase-accessible chromatin using sequencing) (Buenrostro et al., 2013). In the fast-evolving era of single-cell sequencing, scRNA-seq and scATAC-seq (Buenrostro et al., 2015) have been combined to generate multiome data (Badia-i-Mompel et al., 2023; Baysoy et al., 2023). To infer GRNs from these datasets, a plethora of algorithms have recently been developed, including ANANSE (Xu et al., 2021), SCENIC+ (Bravo González-Blas et al., 2023), GRANIE and GRANPA (Kamal et al., 2023), and PANDO (Fleck et al., 2023). Nonetheless, all the methods rely on the same underlying biological logic: TFs are assigned to CREs that are assigned to target genes, and these regulatory interactions are used to construct networks.

As regenerative and non-regenerative species mostly share the same underlying molecular toolkit of genes, differences in regenerative abilities are also thought to arise from differences in the cis-regulatory code (Srivastava, 2021). So far, enhancers active during tissue-level regeneration have been identified in *Drosophila* imaginal discs (Harris et al., 2016), the African killifish (Wang, W., et al., 2020), and most extensively in zebrafish organs (Kang et al., 2016; Goldman et al., 2017; Goldman and Poss, 2020; Thompson et al., 2020; Cao et al. 2022; Sun et al. 2022; Cigliola et al., 2023). In urodeles, snRNA-seq have been used to understand

telencephalon regeneration after injury (Lust et al., 2022). Furthermore, acoels and *Hydra* have begun to shed light on the cis-regulatory code in whole-body regeneration (Gehrke et al., 2019; Murad et al., 2021). Against this background, the demand for understanding regulatory mechanisms in planarian regeneration is ever-increasing.

1.2 TALE factors in metazoans

Understanding a gene regulatory network requires an understanding of the genes within that network. Here, the evolution of TALE factors and the regulatory activities of Pbx, Prep, and Meis are reviewed as background for the third and fourth chapters.

1.2.1 TALE factors are a subclass of homeobox genes

TALE (Three Amino Acids Loop Extension) factors constitute a subclass of homeobox genes alongside the other subclasses ANTP, PRD, LIM, POU, HNF, SINE, CUT, PROS, ZF, and CERS (Holland et al., 2007). Homeobox genes share the homeodomain of around 60 residues, consisting of three α -helices that bind to DNA (Bürglin and Affolter, 2016). Many homeobox genes are involved in regulating embryonic development and cell fate specification, the most famous being the eponymous class of Hox factors (Bürglin and Affolter, 2016). TALE factors are further subdivided into five subclasses: PBC or PBX, MEINOX or MEIS and PREP, TGIF, IRO (Iroquois), and MKX (Mohawk) (Mukherjee and Bürglin, 2007).

The *PBX* genes (Pre-B-cell leukemia homeobox) were discovered as translocation partners with E2a in human pre-B acute lymphocytic leukaemia (Kamps et al., 1990; Nourse et al., 1990). At the same time, the homolog of *PBX* in flies, *extradenticle* (*exd*), was named after the extra denticles observed in the abdominal region of larvae in mutation screens (Wieschaus et al., 1984; Peifer and Wieschaus, 1990). Mice, humans, and chicks have four *Pbx* genes (*Pbx1-4*), while zebrafish has six (*pbx1a*, *pbx1b*, *pbx2*, *pbx3a*, *pbx3b*, *pbx4* or *lazarus*) (Mukherjee and Bürglin, 2007). Misleadingly, *pbx4* or *lazarus* in zebrafish is an orthologue of mammalian *Pbx3*, and no orthologue of *Pbx4* appears to exist in the zebrafish genome (Longobardi et al., 2014). *Exd* is the single orthologue in *D. melanogaster*, while *C. elegans* has three PBC genes: the conserved *ceb-20*, and the divergent *ceb-40* and *ceb-60* (Mukherjee and Bürglin, 2007).

The MEINOX (a contraction of MEIS and KNOX) clade includes Meis and Prep/Pknox. *Meis* (Myeloid Ecotropic viral Integration Site) genes were discovered by viral insertion mutagenesis in mouse (Moskow et al., 1995; Steelman et al., 1997), while *Prep* (PBX Regulatory Protein) or *Pknox* (PBX/Knox1-related) genes were discovered when screening human cDNA libraries (Chen et al., 1997; Berthelsen et al., 1998). In amniote model species, humans and mice have two *Prep* genes (*Prep1* and *Prep2*) and three *Meis* genes (*Meis1-3*), while chick has two *Prep* genes (*Prep1* and *Prep2*) but only two *Meis* genes (*Meis1* and *Meis2*) following loss of *Meis3* (Mukherjee and Bürglin, 2007; Sánchez-Guardado et al., 2011). In zebrafish, *prep* is represented by three copies (*prep1a*, *prep1b*, *prep2*) and *meis* by five (*meis1a*, *meis1b*, *meis2a*, *meis2b*, *meis3*) (Mukherjee and Bürglin, 2007). In *D. melanogaster*, the single MEINOX member *homothorax* (*hth*) was named after the mutational phenotype: homogenisation in the morphology of thoracic denticle bands (Jürgens et al., 1984; Rieckhof et al., 1997). In *C. elegans*, the homologue of *Meis* is *unc-62*, which derives its name from the uncoordinated (Unc) phenotype observed in the first forward genetic screens (Brenner, 1974; Van Auken et al., 2002). In addition, *C. elegans* has a divergent orthologue of *Prep*, *psa-3*, which has lost its homeodomain but retained the conserved MH-A and MH-B domains (Mukherjee and Bürglin, 2007). Originally, the presence of a single MEINOX homologue in *D. melanogaster* gave rise to the simplistic interpretation that *Prep* genes had arisen in the deuterostome lineage, but this idea was revised when both *Meis* and *Prep* were discovered in other insects and the cnidarian *N. vectensis*, tracing the divergence between *Prep* and *Meis* further back to the base of the metazoan tree (Mukherjee and Bürglin, 2007).

TGIF (Transforming growth factor- β -induced factor or 5'-TG-3' interacting factor) factors were described as a new subclass of TALE factors binding to the promoter of CRBP II (Bertolino et al., 1995). Human

and mouse have four *TGIF* genes (*TGIF1*, *TGIF2*, *TGIFLX*, *TGIFLY*), while chick and zebrafish both have two (*tgif1* and *tgif2*) (Hu et al., 2011). Therefore, the sex chromosome-linked *TGIFLX* and *TGIFLY* are eutherian innovations, and it appears that at least *TGIFLX* evolved from autosomal *TGIF2* through retrotransposition (Blanco-Arias et al. 2002). In their phylogenetic analysis, Mukherjee and B r glin (2007) found that the presence of *Tgif2* was restricted to the vertebrate clade, whereas the presence of a single *tgif* gene in the *N. vectensis* genome suggest that the *Tgif* clade diverged from an ancestral MEINOX gene at or before the base of the metazoan phylogeny. In *Drosophila*, the etymology of the name of the *Tgif* homologue *achintya* (*achi*) derives from the Sanskrit word meaning “that which is beyond thought and contemplation”, while *vismay* is Hindi for “surprise” (Ayyar, 2003; Hyman et al., 2003). However, *achi* and *vis* in *D. melanogaster* are the outcome of a recent gene duplication event, as the closely related *D. pseudoobscura* only has a single gene copy (Mukherjee and B r glin, 2007).

The IRO class was defined by three genes in the Iroquois complex of *D. melanogaster*, *araucan* (*ara*), *caupolican* (*caup*), and *mirror* (*mirr*) (Mukherjee and B r glin, 2007). In *D. melanogaster*, the first mutant of this complex had suppressed lateral bristles in the dorsal mesothorax but kept its bristles in the central region, thus giving rise to a bristle pattern resembling the hairstyle of Iroquois American Indians (Dambly-Chaudi re and Leyns, 1992; Leyns et al., 1996). Consequently, *araucan* and *caupolican* are named after the Araucanian people and their chief Caupolican who led the resistance against the Spanish invaders of Chile. In contrast, *mirror* is expressed in the dorsal half of the eye and was named after this mirror-image distribution (McNeill et al., 1997). Human, mouse and chick have six *Irx* genes (*Irx1-6*), while zebrafish has 11 copies (*irx1a*, *irx1b*, *irx2a*, *irx3a*, *irx3b*, *irx4a*, *irx4b*, *irx5a*, *irx5b*, *irx6a*, *irx7*) following whole-genome duplication (Mukherjee and B r glin, 2007; Dildrop and R ther, 2004; Feijo et al., 2004). Interestingly, in both vertebrates and *D. melanogaster* *Irx* genes are clustered in the genome, but this likely reflects independent tandem gene duplications rather than a homologous cluster structure (Mukherjee and B r glin, 2007; Kerner et al., 2009). *C. elegans* has a single IRO gene *irx-1*, and *Irx* genes are present in the genomes of *N. vectensis* and *H. vulgaris*, also indicating that the *Irx* clade diverged at or before the base of the metazoan tree (Mukherjee and B r glin 2007).

Lastly, the MKX (Mohawk homeobox) was named after one of the Iroquois nations, the Mohawk, and was discovered in mouse when querying for *Irx* genes in the mouse genome (Anderson et al., 2006; Mukherjee and B r glin, 2007). Indeed, the *Mkx* subclade belongs to the same clade as the *Irx* genes, and appears to be robust against gene duplications as Mukherjee and B r glin (2007) found a single copy in most deuterostome and protostome species in their phylogenetic analysis. Underwhelmingly, the *Mkx* homologue in *D. melanogaster* was overlooked by developmental geneticists, and is known as *CG11617* and is poorly characterised unlike the genes of the other TALE subclasses (Mukherjee and B r glin, 2007).

All in all, TALE genes predate metazoans and their evolutionary origin can be traced possibly all the way to the last common eukaryote ancestor (Derelle et al., 2007; de Mendoza et al., 2013). Importantly, although the *Hox* genes are likely the best known homeobox genes, they have arisen in the metazoan clade and are not the most basal lineage in the homeobox gene phylogeny (Merabet and Galliot, 2015; B r glin and Affolter, 2016). In contrast, TALE factors are present in animals, fungi (*M-ATYP* and *CUP*), plants (*KNOX* and *BEL*), and unicellular eukaryotes, rendering them some of the most ancient transcription factors in eukaryotes (Bharathan et al., 1997; B r glin et al., 1997; Derelle et al., 2007; B r glin and Affolter, 2016; Joo et al., 2018). In the homeobox gene phylogeny, the bifurcation into TALE and non-TALE lineages spans the entire eukaryotic domain of life, suggesting that the divergence into TALE and non-TALE branches predates eukaryotes, and that the last common ancestor of eukaryotes possessed at least a single TALE and a non-TALE homeobox gene (Bharathan et al., 1997; B r glin, 1997; Derelle et al., 2007). To conclude, TALE factors are an ancestral class of homeobox genes that are deeply embedded in the framework of metazoan multicellularity.

1.2.2 TALE factors possess conserved functional domains

To execute their conserved functions in metazoans, TALE factors rely on a set of conserved functional domains in the protein sequence. As the name Three Amino Acids Loop Extension implies, TALE factors contain an atypical homeodomain (HD) containing a three-amino-acids extension between the first and second helix, further supporting their monophyletic status (Bürglin, 1997; Mukherjee and Bürglin, 2007). These three amino acids are mostly proline-tyrosine-proline (PYP), but alanine-tyrosine-proline (AYP) is also found in the TGIF class (Mukherjee and Bürglin, 2007). This extension forms a flexible linker and hydrophobic pocket that influences DNA binding (Passner et al., 1999; Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Moens and Selleri, 2006; Mukherjee and Bürglin, 2007; Saadaoui et al., 2011; Bürglin and Affolter, 2016; Dard et al., 2018, 2019). This hydrophobic pocket also interacts with the Hox factors upon DNA binding (see Hox cofactors) (Passner et al. 1999). In addition to the atypical HD, the TALE factors contain other subclass-specific domains that confer subclass-specific functionality (Figure 1.3).

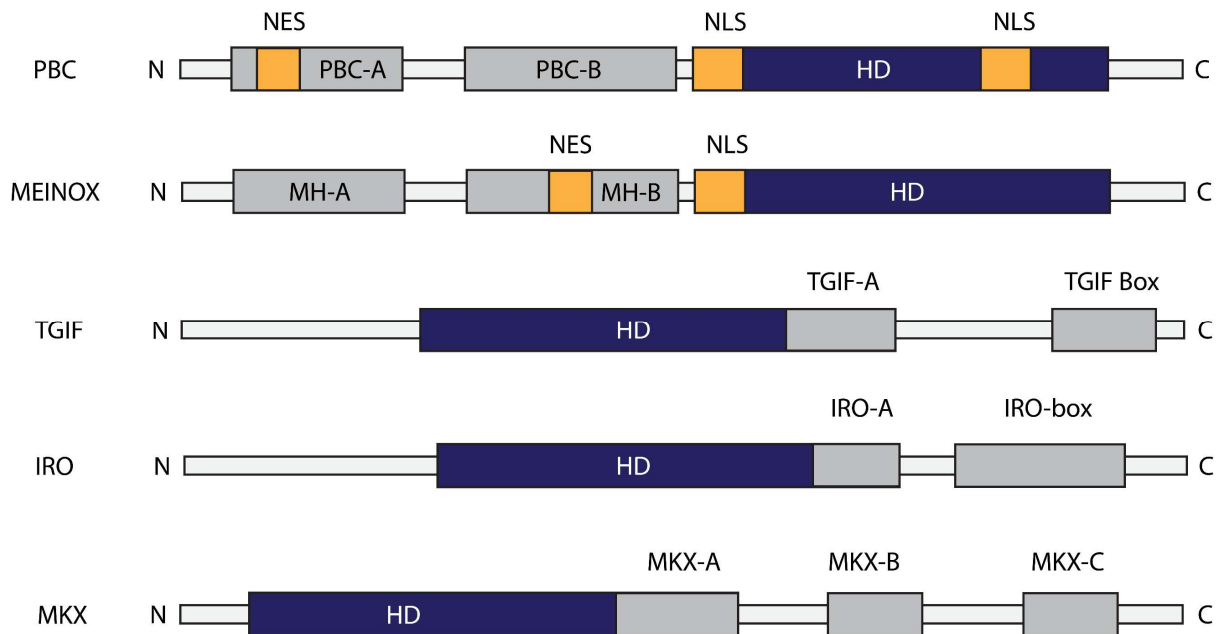


Figure 1.3 Functional domains of TALE proteins. The homeodomain is marked in navy-blue, functional domains of each subclass are marked in dark grey, and nuclear localization signals (NLS) and nuclear export signals (NES) are marked in orange (Reichlmeir et al. 2021).

Towards the N terminus, the PBC proteins contain the conserved domains PBC-A and PBC-B that are necessary for heterodimerisation with MEIS or PREP (see 1.2.3 MEINOX and PBC proteins form dimers) (Figure 3.1) (Longobardi et al., 2014). In addition, the PBC proteins contain nuclear localisation signals (NLS) and nuclear export signals (NES) that regulate subcellular distribution (see 1.2.4 Nuclear localisation). The weaker NLS1 (sequence RRKRR) is located in the N-terminal end of the HD, the stronger NLS2 (sequence KRIRYKKNI) is positioned in the third helix of the HD, a NES-like sequence lies in the PBC-A domain, and the phosphorylation state of serine residues (S187, S193, S202, S209, and S218) in the PBC-B domain also control nuclear localisation (Abu-Shaar et al., 1999; Berthelsen et al., 1999; Saleh et al., 2000; Saleh et al., 2000b; Kilstrup-Nielsen et al., 2003). In the pre-omics era, alternatively spliced isoforms for Pbx1 (Pbx1a and Pbx1b) and Pbx3 (Pbx3a to d) differing in both N-terminal (Pbx3) and C-terminal (Pbx1 and Pbx3) ends were identified and experimentally validated (Monica et al., 1991; Di Rocco et al., 1997; Milech et al., 2001). These isoforms display distinct expression patterns during mammalian development,

proteins post-translationally, as dimerisation is essential for PBX proteins to enter the nucleus for transcriptional activation.

1.2.4 Nuclear localisation controls TALE activity

As discovered in fly embryos, the activities of Exd and Hth are regulated through their dynamic subcellular localisation (Mann and Abu-Shaar, 1996; Aspland and White, 1997). Hth is both necessary and sufficient to induce cytoplasmic to nuclear translocation of Exd, while Exd is necessary for the nuclear localisation of Hth (Rieckhoff et al., 1997; Abu-Shaar and Mann, 1998; Kurant et al., 1998; Jaw et al., 2000). Similarly, the nuclear localisation of Pbx1 correlates with the expression of Meis1 and Meis2 in the developing limb buds in mouse and chicken (González-Crespo et al., 1998; Capdevilla et al., 1999; Mercader et al., 1999). Also in frog and zebrafish embryos, nuclear localisation of Meis proteins was functionally demonstrated to depend on Pbx (Vlachakis et al., 2001; Maeda et al., 2002). In the cytoplasm, PBX proteins interact with non-muscle myosin heavy chain II B that prevents them from entering the nucleus (Huang et al., 2003). Also PREP2 requires PBX1 for its nuclear localisation, and interaction with actin and microtubule cytoskeletons facilitates retention of PREP2 in the cytoplasm (Fognani et al., 2002; Haller et al., 2004).

The balance between nuclear import and export governs nuclear localisation of the Exd:Hth dimer, mediated by NLS and NES sequences. The NLS sequence in Hth is dispensable, while Exd has two NLSs in the HD (Abu-Shaar et al., 1999; Berthelsen et al., 1999; Ryoo et al., 1999; Saleh et al., 2000b). Therefore, MEINOX proteins enter the nucleus with the aid of the NLS in PBC proteins. Since repression of the nuclear export pathway accumulates Exd/Pbx in the nucleus, the NESs were proposed to be in either PBC-A (Berthelsen et al., 1999) or PBC-B (Abu-Shaar et al., 1999). The lack of a proper NES sequence in the PBC-A domain led to the idea that PBC-A promotes nuclear export indirectly by folding onto and masking the NLSs in the HD (Saleh et al., 2000). However, Kilstrup-Nielsen et al. (2003) found two adjacent NESs of the leucine-rich type in the PBC-A domain that mediate direct contact with the CRM1 export receptor. Furthermore, PKA was found to phosphorylate serine residues in the PBC-B domain *in vitro*, and PKA activation blocks nuclear export of PBX *in vivo* (Kilstrup-Nielsen et al., 2003). In another study, disruptive point mutations indicated the presence of a NES in the PBC-B domain (Stevens and Mann, 2007). Taken together, both the PBC-A and PBC-B domains appear to regulate nuclear export of the PBC:MEINOX dimer.

1.2.5 TALE proteins are Hox cofactors

The “Hox specificity paradox” refers to the seemingly contradictory observation that Hox factors regulate specific developmental programs but bind highly similar DNA sequences (TAAT) *in vitro* (Merabet and Mann, 2016). The existence of additional binding partners, so called Hox cofactors, were offered as a partial explanation to this paradox. After pioneering studies on fly development, the TALE factors rose to fame and became known as these essential Hox cofactors. In *D. melanogaster*, both *exd* and *hth* mutants resemble Hox mutants, but they do not affect Hox gene expression, suggesting that they modulate HOX protein binding (Peifer and Wieschaus, 1990; Rauskolb et al., 1993; Rauskolb and Wieschaus, 1994; Rieckhof et al., 1997; Kurant et al., 1998; Pai et al., 1998). Similarly, mutants of *pbx*, *prep*, and *meis* were discovered to resemble Hox mutants and were reasoned to act as cofactors of Hox proteins in the vertebrate hindbrain and *C. elegans* (Pöpperl et al., 1995, 2000; Ferretti et al., 2000; Van Aucken et al., 2002; Samad et al., 2004). The TALE-HOX partnership appears to be conserved even beyond bilaterians, as TALE factors in the cnidarian *N. vectensis* are also able to interact with HOX proteins (Ferrier et al., 2014; Hudry et al., 2014; Merabet and Galliot, 2015). However, this partnership seems to have arisen in metazoans, as TALE factors from the unicellular *Acanthamoeba* are unable to interact with HOX proteins (Hudry et al., 2014).

In the wake of the enthusiasm sparked by the first observed mutants, an extensive research program was initiated to understand the structural biology of the PBC:HOX dimer and cooperative DNA binding dynamics (Chan et al., 1994; van Dijk and Murre, 1994; Chang et al., 1995; Johnson et al., 1995; Knoepfler

and Kamps, 1995; Lu et al., 1995; Neuteboom et al., 1995; Phelan et al., 1995; Chan and Mann, 1996; Chang et al., 1996; Lu and Kamps, 1996; Peltenburg and Murre, 1996; Shen et al., 1996; Lu and Kamps, 1997; Neuteboom and Murre, 1997; Phelan and Featherstone 1997; Shen et al., 1997a; Piper et al., 1999; Passner et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Joshi et al., 2007; Slattery et al., 2011; Dard et al., 2018). Indeed, formation of the PBC:HOX complex was found to increase and modulate DNA binding specificity of Hox factors (Chang et al., 1996; Shen et al., 1996; Lu and Kamps, 1997; Neuteboom and Murre, 1997; Joshi et al., 2007; Slattery et al., 2011). PBC proteins interact with the hexapeptide motif (YYPWMK) situated N-terminal to the homeodomain of HOX proteins (Lu and Kamps, 1996; Peltenburg and Murre, 1996; Shen et al., 1996; Shen et al., 1997; Passner et al., 1999; Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Saadaoui et al., 2011). Crystallography of complexes in *D. melanogaster* and mammals revealed that the hexapeptide forms strong contacts within the hydrophobic pocket created by the homeodomain in PBC proteins (Passner et al., 1999; Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Joshi et al., 2007). In the PBC:HOX complex, each homeodomain binds one half of an octameric DNA sequence, and the third helix of each homeodomain inserts into the main groove of the DNA double helix (Passner et al., 1999; Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Joshi et al., 2007). However, PBC:HOX dimers can form even in the absence of the hexapeptide, suggesting that the interaction between PBC and HOX is more versatile than initially thought (Galant et al., 2002; Hudry et al., 2012; Foos et al., 2015). In fact, later studies have revealed that binding specificity also depends on the extended and unstructured regions in Hox proteins (Joshi et al., 2007). In the current model, binding specificity is provided by extended regions of the Hox/Exd complex that insert into the minor groove and recognise a sequence-dependent DNA structure (Joshi et al., 2007).

The first experiments indicated that Exd seems to interact with and increase the binding specificity of Hox proteins from Labial to AbdA, but not the posterior AbdB (Chan et al., 1994; van Dijk and Murre, 1994; Johnson et al., 1995; Chan and Mann, 1996). However, a SELEX-seq screening indicates that Exd can dimerise with all eight Hox proteins in *D. melanogaster* (Slattery et al., 2011). Similarly, the first *in vitro* studies suggested that vertebrate Pbx proteins can dimerise with the ten first Hox paralog proteins, but not with posterior-group Hox11-Hox13 proteins (Chang et al., 1995; Knoepfler and Kamps, 1995; Lu et al., 1995; Phelan et al., 1995; Chang et al., 1996; Neuteboom and Murre, 1997; Shen et al., 1996; Shen et al., 1997a). However, bimolecular fluorescence complementation (BiFC) in human cells revealed that HOXA11, HOXD12, and HOXB13 can interact with PBX1, and that HOX proteins from the posterior groups can form ternary HOX:PBX1:MEIS1 complexes (Dard et al., 2018). Lastly, genome-wide ChIP-seq has demonstrated the co-occupancy of PBC and Hox proteins (paralog groups 1 to 7) also *in vivo* (Donaldson et al., 2012; De Kumar et al., 2017; Bridoux et al., 2020; Feng et al., 2022).

With respect to the MEIS:HOX dimer, MEIS1 was first discovered to dimerise with posterior HOX proteins (HOX9 to HOX13), but later experiments also revealed interactions with anterior Hox proteins (Shen et al., 1997b; Williams et al., 2005). In these studies, MEIS proteins were shown to interact with HOX proteins through C-terminal sequences (18 residues in MEIS1 and 93 residues in MEIS2) (Williams et al., 2005). ChIP-exo and CAP-selex experiments have further validated interaction between MEIS and posterior HOX proteins (Jolma et al., 2015). In a genomic context, ChIP-seq in *Drosophila* and mouse embryos along with hematopoietic cells have indicated the co-occupancy of MEIS and both anterior and posterior HOX proteins (Choo et al., 2011; Fan et al., 2012; Huang et al., 2012; Penkov et al., 2013; Delgado et al., 2021).

As PBC proteins dimerise with MEINOX and HOX proteins through different interaction surfaces, it allows for the formation of trimeric MEINOX:PBC:HOX complexes. Indeed, the first trimeric complex to be observed was the Prep1:Pbx1:Hoxb1 complex *in vitro* (Berthelsen et al., 1998a). In *Drosophila* embryos, the trimeric Hth:Lab:Exd complex is required for the activity of an autoregulatory enhancer of *lab* (Ryoo et

al., 1999; Ebner et al., 2005). Likewise, Pbx:Meis:Hox and Pbx:Prep:Hox trimers have been found to assemble on enhancers and drive Hox gene expression in the vertebrate hindbrain (Jacobs et al., 1999; Ferretti et al., 2000; Vlachakis et al., 2000; Ferretti et al., 2005). Adjacent Pbx/Hox and Prep/Meis binding sites facilitate trimerisation, but the trimeric complex can form even without Meis binding to DNA (Berthelsen et al., 1998a; Ryoo et al., 1999; Shanmugam et al., 1999; Ferretti et al., 2000; Ferretti et al., 2005). Nonetheless, the MEIS C-terminal is required for the formation of trimeric complexes (Shanmugam et al., 1999). Lastly, ChIP-seq studies have also found tripartite co-occupancy of Pbx, Meis, and Hox proteins in the genome (Amin et al., 2015; Bridoux et al., 2020).

1.2.6 TALE factors also interact with non-HOX components

Although TALE factors first gained a reputation as cofactors to *Hox* genes, later research has unveiled numerous Hox-independent functions. A combination of *in vitro* studies, protein-protein interaction screens and ChIP-seq experiments suggest that Pbx proteins interact with multiple other homeobox proteins, other transcription factors such as basic helix-loop-helix factors (bHLH), and miscellaneous other proteins (Table 1.1). Among non-Hox genes, the best characterised Pbx interactors are Pdx1 and myogenic bHLH factors that are involved in the differentiation of the pancreas and muscle cells, respectively (Table 1.1). Likewise, the non-Hox interactome of Meis proteins has been discovered to include notable interactions between Pax and Dlx proteins in neuronal development (Table 1.2). The interactome of Prep is not as well characterised, but includes a set of other transcription factors (Table 1.3). Beyond metazoans, the TALE factor MATalpha2 has been discovered to interact with the other homeobox gene MATa1 (Li et al., 1995; Bürglin and Affolter, 2016).

Table 1.1 Non-Hox interactions with Pbx.

Homeobox proteins	
Pdx1	(Peers et al., 1995; Goudet et al., 1999; Liu et al., 2001; Kim et al., 2002)
Emx2	(Capellini et al., 2010)
En	(Kobayashi et al., 2003)
Pax6	(Mikkola et al., 2001)
FoxC1	(Berry et al., 2005)
Tlx3	(Rhee et al., 2004)
POU2F1	(Rave-Harel et al., 2004)
bHLH proteins	
TCF3 or E2A	(Knoepfler et al., 1999)
MyoD	(Knoepfler et al., 1999; Maves et al., 2007)
Myogenin	(Knoepfler et al., 1999)
Myf6	(Knoepfler et al., 1999)
Myf5	(Knoepfler et al., 1999)
HAND2	(Losa et al., 2023)
Other transcription factors	
ZNF462	(Laurent et al., 2007; Laurent et al., 2009a, b)
Klf4	(Bjerke et al., 2011)
TRalpha	(Wang et al., 2001)
Glucocorticoid receptors	(Subramaniam et al., 2003; Ravasi et al., 2010)
Eralpha	(Magnani et al., 2011)
Smad2,3,4	(Bailey et al., 2004)
Other proteins	
NMHCb	(Huang et al., 2003)
Zipper	(Huang et al., 2003)

PBXIP1	(Abramovich et al., 2000)
ZNHIT3	(Ravasi et al., 2010)
ZDHHHC7	(Ravasi et al., 2010)

Table 1.2 Non-Hox interactions with Meis.

Homeobox proteins	
Otx2	(Agoston and Schulte, 2009)
Pax3, Pax6, Pax7	(Agoston et al., 2012, 2014)
Dlx2, Dlx3	(Agoston et al., 2014; Jolma et al., 2015)
En	(Kobayashi et al., 2003)
Pdx1	(Swift et al., 1998; Liu et al., 2001)
Tlx1, Tlx2	(Milech et al., 2010; Ravasi et al., 2010)
Emx2	(Ravasi et al., 2010)
Evx1	(Jolma et al., 2015)
Drgx	(Jolma et al., 2015)
Onecut2	(Jolma et al., 2015)
bHLH proteins	
TCF3 or E2A	(Knoepfler et al., 1999)
MyoD	(Knoepfler et al., 1999; Maves et al., 2007)
Myogenin	(Knoepfler et al., 1999)
Myf6	(Knoepfler et al., 1999)
Myf5	(Knoepfler et al., 1999)
Max	(Jolma et al., 2015)
Other transcription factors	
Ey	(Bessa et al. 2002)
Tsh	(Bessa et al. 2002)
Klf4	(Bjerke et al. 2011)
Ikaros	(Aubrey et al. 2022)
Lobe	(Singh et al 2011)
Androgen receptor	(Cui et al. 2014)
Eomes	(Jolma et al. 2015)
Elf1	(Jolma et al. 2015)
Other proteins	
PKA	(Huang et al. 2005)

Table 1.3 Non-Hox interactions with Prep.

Transcription factors	
Pax6	(Mikkola et al., 2001)
POU2F1	(Rave-Harel et al., 2004)
Tlx1	(Ravasi et al., 2010)
Rfx3	(Ravasi et al., 2010)
Smad 2,3,4	(Bailey et al., 2004)
NF-Y	(Ladam et al., 2018; Stanney 3 rd et al., 2020)
Other proteins	
MYBBP1a	(Díaz et al., 2007; Oriente et al., 2008)
4EHP	(Villaescusa et al., 2009)
ZBTB3	(Ravasi et al., 2010)

1.2.7 TALE proteins are modulators of the epigenomic landscape

Pioneer factors are transcription factors that have the unique property to engage with closed chromatin, initiate chromatin opening, enable binding of additional factors, and drive cell fate specification and re-specification. The versatility of TALE factors has raised the question of whether they also might act as pioneers (Grebbe et al., 2017). Currently, the pioneering status of TALE proteins is supported by preoccupation at cis-regulatory regions, correlation with changes in histone modifications, and direct interactions with histone-modifying enzymes. However, the putative pioneering role of TALE factors is based on a handful of diverse case studies, including some in pathological contexts, which makes it difficult to assess the prevalence and importance of pioneering.

At least in a few studied cases, TALE factors occupy binding sites before others. At the promoter of *MYOG*, the binding of PBX1 precedes the binding of the master regulator MYOD driving muscle differentiation, and bound PBX1 is observed before other factors in undifferentiated neuronal progenitor cells (Berkes et al., 2004; Maves et al., 2007; Grebbe et al., 2016). In branchial arches, Meis1/2 binds to its targets before *Hoxa2*, while PBX1 occupancy precedes and enables the binding of the estrogen receptor in breast cancer tissue (Magnani et al., 2011; Amin et al., 2015). In zebrafish embryos, Pbx, Prep, and Meis bind to the promoter of *boxb1a* before other factors, and the maternally deposited Pbx and Prep form a trimeric regulatory complex with NF-Y prior to the binding of tissue-specific factors (Choe et al., 2014; Ladam et al., 2018). Altogether, the dynamics of transcription factor binding suggests that TALE factors can bind to compacted chromatin and promote the binding of other factors.

Furthermore, in zebrafish blastulas, Pbx and Prep/Meis promote the increase in active (H3K27ac) and decrease in repressive (H3K27me3) histone modifications and recruit RNA polymerase II to trigger weak initial transcription of *boxb1a* before another Hox protein also binds to stabilise transcription levels in a cross-regulatory loop (Choe et al., 2014). At the same promoter, Meis binds to Pbx and displaces HDACs (histone deacetylase) in exchange for CBP (a histone acetyltransferase), which drives the change in histone modifications (Choe et al., 2009). The loss of Prep and Pbx is also associated with the reduction in H3K27ac levels genome-wide (Ladam et al., 2018). Lastly, a mechanistic case study of the *Dcx* promoter in neuronal progenitors revealed that Meis interacts with prebound Pbx1 and recruits PARP1 that modifies and releases histone H1, recruits RNA polymerase II, and facilitates *Dcx* transcription (Hau et al., 2017).

Studies in cell cultures suggest that Pbx proteins interact with a transcriptional corepressor complex including HDACs, SIN3B, and N-CoR/SMRT (Saleh et al., 2000a). The N-terminal domain in Pbx1 has been observed to mediate the interaction with CBP, while the C-terminal is associated with the corepressors N-CoR/SMRT (Asahara et al., 1999). In mesenchymal cells, Pbx1 associates with HDACs and seems to keep chromatin inactive at osteoblast-related genes, as loss of Pbx1 increases the recruitment of histone acetylases (p300) (Gordon et al., 2010; Gordon et al., 2011). In leukaemogenic samples, Meis1 binding sites are associated with increased levels of H3K4me1 and CBP/p300 binding (Huang et al., 2012).

To summarise, the properties of TALE factors and circumstantial evidence suggests pioneering activity, yet it cannot be ruled out that TALE-marked sites are already primed by another pioneer. Hence, it has been suggested that TALE factors serve as “second-rank” pioneers that synchronise and facilitate the activity of histone-modifying enzymes and tissue-specific transcription factors in establishing lineage-specific transcriptional programs (Schulte and Geerts 2019; Bobola and Sagerström, 2024).

1.2.8 The DNA binding sites of TALE factors are well characterised

As transcription factors, an important part of the functionality of TALE factors is their binding to DNA sequences. *In vitro* selection of target sequences by protein complexes, along with the analysis of a number of regulatory elements within promoters of target genes, established the consensus binding sequence for PBC:HOX complexes as 5'-TGATNNATNN-3' (termed as the “OCTA” motif) (Chan et al., 1994; Lu et

al. 1995; Pöpperl et al., 1995; Chan and Mann, 1996; Chang et al., 1996; Shen et al., 1996; Grieder et al., 1997; Maconochie et al., 1997; Shen et al., 1997; Phelan and Featherstone, 1997; Lu and Kamps, 1996; Passner et al., 1999; Piper et al., 1999; Ryoo and Mann, 1999; Tümpel et al., 2007). The consensus site consists of directly adjacent PBX and HOX half-sites (Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003). The binding site for monomeric PBX was also identified to be a subvariant of this, namely TGATTGAT (LeBrun and Cleary, 1994). Similarly, monomeric Hth/Meis/Prep was discovered to prefer TGACAG (the “HEXA” motif) (Chang et al., 1997; Shen et al., 1997; Berthelsen et al., 1998b; Jacobs et al., 1999; Ferretti et al., 2000; Ferretti et al., 2005; Tümpel et al., 2007), while the preferred binding site for the Pbx:Meis/Prep heterodimer was identified as TGATTGACAG (the “DECA” motif) (Chang et al., 1997; Knoepfler et al., 1997; Knoepfler and Kamps, 1997).

Afterward, the first ChIP-seq experiments on embryonic mouse trunks were able to reproduce the HEXA, OCTA, and DECA motifs (Penkov et al., 2013). The sequencing results revealed that Meis and Prep select different DNA-binding sequences in the genome, as Meis prefers the HEXA and OCTA motifs, while Prep prefers the DECA motif (Penkov et al., 2013). Additionally, Prep binding sites were enriched in promoters and housekeeping genes, while Meis binds to distal enhancers of developmental genes (Penkov et al., 2013). Additional ChIP-seq experiments on other mouse tissues also found that Meis binds to the HEXA motif, Hox-Pbx to the OCTA motif, and Prep to the DECA motif (Amin et al., 2015; Marcos et al., 2015; Laurent et al., 2015; De Kumar et al., 2017). Furthermore, Prep and Pbx co-occupy the DECA motif in the genome of zebrafish blastulas (Ladam et al., 2018). Overall, an evolutionarily conserved set of motifs has been linked to TALE factors, but the results are still somewhat ambiguous concerning the overlap of motifs preferred by which monomeric TALE factors and which by TALE-TALE and TALE-HOX complexes.

1.2.9 TALE factors regulate the cell cycle and cell physiology

Predating metazoan multicellularity, TALE factors have been associated with several aspects of cell cycle control. PBX1 was discovered as a proto-oncogene in human leukaemia, and the role of PBX in tumorigenesis has expanded into its own subdiscipline (Kamps et al., 1990; Nourse et al., 1990; Liu et al., 2022). Carcinogenesis has been explained by the aberrant activity of Pbx proteins in cell proliferation and apoptosis, epithelial-mesenchymal transition (EMT), invasion and metastasis, and cancer stemness (Liu et al., 2022). Similarly, Meis proteins were recognised as oncogenes in leukaemias (Nakamura, 1996a, b; Smith et al., 1997; Lawrence et al., 1999). In contrast, Prep1 acts as a tumour suppressor as demonstrated by increased DNA damage and tumorigenesis in cells and hypomorphic mouse strains with a reduced level of Prep expression (Longobardi et al., 2010; Iotti et al., 2011, 2012). In tumorigenesis, Meis1 and Prep1 compete for Pbx1 and counteract each other (Dardaei et al., 2014, 2015). Furthermore, single nucleotide substitutions, gene deletions, and gene amplifications of TALE genes are implicated in many human tumours (Blasi et al., 2017).

Expectedly, TALE proteins have been found to control both directly and indirectly several components of the cell cycle machinery also during embryonic development. Meis1/2 controls the expression of *cyclin D1* (*Ccnd1*) and *c-Myc* (*Myc*), while *Ccnd1* controls the expression of *Meis2* and forms an autoregulatory loop that drives proliferation of retinal progenitors (Bessa et al., 2008; Heine et al., 2008; Bienvenu et al., 2010). In a similar manner, Hth acts indirectly via an intermediary layer of nuclear receptors to repress *Stg* (*string*, homolog of *Cdc25*) that is required for cell cycle progression in the fly retina (Bessa et al., 2002; Lopes and Casares, 2010, 2015; Neto et al., 2017). In contrast, Meis is required for the activation of Cdk inhibitors p15, p16, and p21 to slow down the cell cycle in cardiomyocytes (Mahmoud et al., 2013).

Furthermore, the genomic distribution of Prep binding in the mouse embryonic trunk suggested that Prep1 is mostly involved in basal cell functions rather than development (Penkov et al., 2013; Laurent et al., 2015). In the mouse and zebrafish embryos, Prep1-deficiency has been linked to increased levels of apoptosis, suggesting that Prep is involved in the suppression of apoptosis during embryonic development (Deflorian

et al., 2004; Micali et al., 2009; Fernandez-Diaz et al., 2010). Prep proteins have been found to interact with the nuclear lamina and regulate the stiffness of the nucleus, while ablation of Prep1 raises the expression of components in the oxidative phosphorylation cascade (Kanzleiter et al., 2014; Palmigiano et al., 2018; Purushothaman et al., 2022). All in all, TALE factors are involved in both basal cell functions maintaining proper cell physiology and the coordinated proliferation of cells.

1.2.10 TALE factors play diverse roles in metazoan development

TALE factors are true sculptors of the embryo, as Pbx, Prep, and Meis have been implicated in the development of almost every mammalian organ and the maintenance of adult stem cells (Selleri et al., 2019). Although a lot of our knowledge of TALE factors is based on mouse development and associated biomedical research, experiments in *D. melanogaster* along with some other organisms have begun to shed light on possible evolutionarily shared developmental functions of TALE proteins, including axial patterning, the outgrowth of both vertebrate and invertebrate appendages, and the development of the visual system.

1.2.10.1 TALE proteins mould the mammalian embryo and maintain adult stem cells

During mouse development, *Pbx* genes display overlapping expression domains and functions, both also gene-specific developmental roles (Selleri et al., 2019). *Pbx1* is expressed in almost all tissues and organ systems during development, and loss of *Pbx1* leads to embryonic death with hypoplasia or malformations in virtually all organs (Schnabel et al., 2001; Selleri et al., 2001). *Pbx2* is also expressed in multiple tissues, but deletion of *Pbx2* does not affect viability (Selleri et al., 2004). Expression of *Pbx3* is not as widespread but can be detected in the nervous system, the forelimb but not the hindlimb, mesenchyme, and ovaries, and *Pbx3*(-/-) mice die within one day after birth due to central respiratory failure (Rhee et al., 2004; Di Giacomo et al., 2006). In contrast, the expression of *Pbx4* is restricted to the testes, and *Pbx4* appears to not have as a systemic developmental function as its paralogues (Wagner et al., 2001).

In both mouse and chick, *Meis1* and *Meis2* are expressed in the primitive streak and lateral plate mesoderm during gastrulation, and then in the neural tube, eye, inner ear, migrating neural crest cells, and the cardiovascular system at the late somite stages (Coy and Borycki, 2010; Cecconi et al., 1997; Sánchez-Guardado et al., 2011). Systemic knockout and functional inactivation of *Meis1* and *Meis2* in mouse embryos are embryonic lethal and mainly display defects in the eyes, haematopoiesis, and the cardiovascular system (Hisa et al., 2004; Azcoitia et al., 2005; Stankunas et al., 2008; Machon et al., 2015).

In the mouse embryo, the Prep1 protein is found in the developing first branchial arch, the hepatic primordium, the midgut, the lung bud, neuronal tissues, and weakly in the heart (Ferretti et al., 1999). Deletion of *Prep1* causes embryonic lethality, as the embryos do not gastrulate and the epiblast deteriorates, while hypomorphic mouse strains develop until E17.5 and have overall organ hypoplasia, haematopoietic defects, and vascular and eye anomalies (Ferretti et al., 2006; Fernandez-Diaz et al., 2010; Rowan et al., 2010). The mammalian paralogue *Prep2* is highly expressed in the heart, brain, skeletal muscles and the ovary, and to a lesser degree in the placenta, lungs, pancreas, prostate, and small intestine, but is not known whether *Prep2* has paralogue-specific functions in mammalian development (Haller et al., 2002; Fognani et al., 2002).

Besides embryonic functions, TALE proteins are involved in adult stem cell biology as a part of haematopoiesis (DiMartino et al., 2001; Azcoitia et al., 2005; Di Rosa et al., 2007; Pillay et al., 2010; Cvejic et al., 2011; Modica et al., 2014). Haematopoiesis is driven by haematopoietic stem cells (HSCs) that give rise to the common myeloid progenitors (CMPs) and granulocyte-macrophage-lymphocyte progenitors (GMLPs) (Jagannathan-Bogdan and Zon, 2013). In turn, GMLPs differentiate into B and T cells, while CMPs give rise to erythrocytes, thrombocytes and monocytes. In adults, Pbx and Meis are required for maintaining the pool of HSCs (Ficara et al., 2008; Kocabas et al., 2012; Ficara et al., 2013; Ariki

et al., 2014), while Prep1 appears to be dispensable for the maintenance of adult HSCs (Yoshioka et al., 2015). Furthermore, Pbx1 regulates B-cell commitment (Sanyal et al., 2007), and Prep1 is required for both B and T cell differentiation (Penkov et al., 2005, 2008; Iotti et al., 2012; Yoshioka et al., 2015). All in all, Pbx, Meis, and Prep are widely involved in organogenesis and cell fate specification during both embryonic development and adult cell turnover in mammals.

1.2.10.2 TALE proteins are involved in axial patterning

Evidence from different metazoans raise the possibility for an evolutionarily conserved role for TALE factors in axial patterning. The primary or anterior-posterior axis of bilaterians is patterned by posterior Wnt/beta-catenin signalling, while the secondary or dorsal-ventral axis is patterned by BMP signaling (Petersen and Reddien, 2009; Niehrs, 2010; Loh et al., 2016; Petersen, 2023; Holzem et al., 2024). Indeed, investigations into the positional information system of planarians have greatly contributed to our understanding of the evolution of the mechanisms and signalling pathways that pattern the bilaterian body axes (see 1.3 Patterning in planarian regeneration). As Hox factors pattern the bilaterian AP axis, this raises the question how *Hox* genes are linked to the patterning activity of the Wnt signalling pathway. In the cnidarian *N. vectensis*, a study suggests that knockdown of *Hox* genes affects oral-aboral patterning and the expression of the Wnt signalling pathway, but the mechanistic link to Wnt signalling remains unclear (DuBuc et al., 2018; Rentzsch and Holstein, 2018; Holstein, 2022). Furthermore, as TALE factors are evolutionarily conserved cofactors of Hox factors, this raises the possibility of an integrated Wnt/Hox/TALE regulatory module in patterning the bilaterian body plan.

Vertebrate tail development and regeneration have begun to uncover an interrelationship between Wnt, Hox, and TALE factors. As in *D. melanogaster*, mammalian mutants of *Pbx* and *Meis* display homeotic transformations and mispatterned axial skeletons (Capellini et al., 2008; Berenguer et al., 2020; López-Delgado et al., 2021). During somitogenesis, neuromesodermal progenitors (NMPs) in the tailbud first migrate posteriorly to the progenitor zone, and then change direction and migrate bilaterally, enter the maturation zone, and differentiate into mesodermal progenitor cells that form the unsegmented pre-somitic mesoderm (Mariani et al., 2021). In this context, TALE proteins regulate the migration and differentiation of NMPs by opening chromatin and permitting the Wnt effector LEF1 to bind to its target genes and drive the differentiation program into paraxial mesoderm (Mariani et al., 2021). In the current model, PBX:PREP1 dimers first bind loosely to chromatin, and then a PBX:MEIS2:HOX1 ternary complex increases chromatin accessibility to facilitate the activation of Wnt-responsive genes (Mariani et al., 2021). Furthermore, Wnt signalling along with Pbx and Meis are required for tail regeneration in *Xenopus tropicalis*, but the detailed mechanistic interrelationship between these remains unknown (Lin and Slack, 2008; Kakebeen et al., 2020).

The vertebrate hindbrain is another well-studied model system that provides further insight into the regulatory interactions between Wnt, Hox, and TALE factors. During neurulation, the ectodermal neural plate invaginates and seals up to give rise to the neural tube. At the anterior end, the neural tube expands and forms three primary vesicles: the forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon). The hindbrain becomes compartmentalised into seven rhombomeres (r1 to r7) that organise the neuronal populations and adjacent craniofacial tissues. The hierarchical GRN governing hindbrain segmentation involves a primary AP signalling module, a secondary segment subdivision module, and a tertiary intrasegmental AP patterning module (Krumlauf and Wilkinson, 2021). During the AP signalling module, cooperative interactions between the Fgf, Wnt, and retinoic acid signalling pathways initiates segmentation along the AP axis. In this phase, mesodermal Wnt3a activates *Meis* and *Pbx*, which in turn induce the expression of *Fgf3/8* and the biosynthetic enzyme *Raldh2* (Waskiewicz et al., 2002; Dibner et al., 2004; Elkouby et al., 2010; Gutkovich et al., 2010; Vitobello et al., 2011; Schulte and Frank, 2014). *Raldh2* produces RA in the mesoderm from which it diffuses to form a

posterior-to-anterior gradient required for segmenting the hindbrain into rhombomeres (Schulte and Frank, 2014). In the segment subdivision module, a rhombomere-specific TALE-mediated Hox code provides segmental identity to each rhombomere (Schulte and Frank, 2014). Here, Hox binding activity is modulated by dimeric and trimeric complexes involving Pbx, Meis, Prep, and Hox, as disruption of their activities perturbs rhombomere segment identity formation (Pöpperl et al., 1995; Salzberg et al., 1999; Pöpperl et al., 2000; Dibner et al., 2001; Vlachakis et al., 2001; Waskiewicz et al., 2001, 2002; Choe et al. 2002; Maeda et al. 2002; Deflorian et al., 2004; Vaccari et al. 2010). Taken together, TALE factors have been shown to act as cofactors to Hox genes in Wnt-regulated contexts, but the exact mechanisms of interaction and the broader evolutionary significance of the interplay between Wnt, Hox, and TALE factors is still poorly understood.

1.2.10.3 TALE factors grow out vertebrate and invertebrate appendages

TALE factors have been well characterised in the development of both vertebrate and invertebrate appendages. The tetrapod limb develops from limb buds derived from the posterior lateral mesoderm, neighbouring somites, and the overlying ectoderm. Limb bud growth is fuelled by the proliferative mesenchyme in the progress zone, the anterior-posterior axis is patterned by the zone of polarizing activity (ZPA), and the limb apex is covered by the apical ectodermal ridge (AER). Pbx proteins are necessary for both proximal and distal limb development (González-Crespo et al., 1998; Selleri et al., 2001; Capellini et al., 2006; Capellini et al., 2011; Losa et al., 2023). Pbx expression in the mesenchyme is necessary for hindlimb bud development, and Pbx proteins execute some of their function through cooperative binding with Hand2 (Losa et al., 2023). Later in the developing bud, Pbx proteins are also necessary for *Sbb* expression in the ZPA (Capellini et al. 2006). Similarly, *Meis* genes are expressed in the lateral plate mesoderm, regulate limb initiation, and are then involved in patterning the growing limb along the anterior-posterior and proximal-distal axes (Capdevila et al., 1999; Mercader et al., 1999; Mercader et al., 2000; Delgado et al., 2009; Delgado et al., 2020; Delgado et al., 2021). Additionally, the patterning activity of Meis has been shown to be recapitulated in regenerating axolotl limbs (Mercader et al., 2005).

In insects, appendages can be subdivided into two categories. Dorsal appendages include wings and halteres, while ventral appendages include antennae, feeding appendages, legs and genitalia. From proximal to distal, the leg consists of coxa, trochanter, femur, tibia, tarsus and pretarsus. The leg develops from the leg disc that telescopes out from the centre to form the leg. Similar to vertebrates, *Exd* and *Hth* are involved in proximodistal development of the insect leg (González-Crespo et al., 1998; Kojima, 2004). In the leg disc of *D. melanogaster*, *dll* is expressed in the center, while *hth* is expressed in the peripheral ring. *Exd* is expressed throughout the leg disc but is localised to the nucleus in the *hth*-expressing domain that develops into the proximal part of the leg (Kojima, 2004). The proximodistal patterning mechanism of *D. melanogaster* has been found to be conserved in other arthropods (Kojima, 2004; Bruce and Patel, 2020). Furthermore, cephalopods have evolved limbs independently, yet *hth* and *exd* are involved in the proximal regionalisation of limbs (Tarazona et al., 2019). Altogether, Pbx and Meis have been found to be involved in appendage outgrowth and patterning in many metazoan lineages.

1.2.10.4 Meis has a conserved role in the development of the visual system

Moreover, the role of Meis is conserved in the gene regulatory networks of eye development in both flies and vertebrates. In flies, the eye imaginal disc invaginates from the ectoderm, while the optic vesicle in vertebrates protrudes from the neural tube (Schulte and Frank, 2014). Famously, *pax6* is necessary and sufficient for eye development in both phyla (Halder et al., 1995). The eye imaginal disc is transversed by the morphogenetic furrow that proceeds from posterior to anterior. In the anterior, Hth maintains retinal progenitors in the undifferentiated and proliferating state, while Hth is downregulated in the posterior and the postmitotic neuronal differentiation trajectory is initiated (Pai et al., 1998; Bessa et al., 2002; Lopes and Casares, 2010, 2015). In the vertebrate retina, neurons and glia arise from multipotent retinal progenitor

cells (Schulte and Frank, 2014). Differentiation occurs in periodic waves, which initiate in the central retina and advance towards the margin (Schulte and Frank, 2014). Loss of *Meis* disturbs vertebrate eye development (Zhang et al., 2002; Hisa et al., 2004; Marcos et al., 2015), and similar to flies, *Meis* maintains the pool of undifferentiated progenitors (Bessa et al., 2008; Heine et al., 2008; Dupacova et al., 2021). In both the retina and lens ectoderm, *Meis* proteins are direct regulators of the master regulator Pax6 (Zhang et al., 2002; Marcos et al., 2015; Antosova et al., 2016).

1.2.11 TALE factors have diversified in spiralian

Interestingly, although TALE factors are an evolutionarily conserved class of homeobox genes and are deeply embedded in metazoan development as outlined above, phylogenetic studies have revealed that TALE factors have diversified in spiralian and given rise to lineage-specific genes (Paps et al., 2015; Morino et al., 2017; Barton-Owen et al., 2018; Phuangphong et al., 2021; Morino et al., 2022; Morino and Yoshikawa, 2023). Spiralian or lophotrochozoans comprise 14 out of 36 metazoan phyla, including molluscs, annelids, rotifers, and the platyhelminths. Morphological evolution has been proposed to be mainly driven by changes in the expression of a conserved set of toolkit genes rather than by sequence evolution in protein-coding genes (Davidson and Erwin, 2006; Carroll, 2008; Carroll, 2013), but recent research has revived the importance of asymmetric gene duplications as a force of morphological evolution (Holland et al., 2017). Indeed, the diversification of TALE genes in spiralian constitutes one of the largest known expansion of homeobox genes yet described (Paps et al., 2015; Morino et al., 2017; Barton-Owen et al., 2018; Phuangphong et al., 2021; Morino et al., 2022).

A pioneering phylogenetic analysis based on eight spiralian species discovered orphan TALE genes not belonging to any of the conserved TALE classes and classified them into seven spiralian-specific subclades (*TALE I* to *VII*) (Paps et al., 2015). In another analysis, Barton-Owen et al. (2018) also uncovered orphan spiralian TALE genes and proposed a classification scheme based on 19 TALE clades (*TALE I* to *XIX*). In contrast, Morino et al. (2017) reconstructed a monophyletic spiralian-specific clade of divergent TALE genes and named it *SPILE* (SPIralian-specific taLE). Later analyses of mollusc sequences have refined the phylogeny and subdivided the *SPILE* clade into four subclades, *SPILE-A/C*, *SPILE-B*, *SPILE-D* and *SPILE-E*, based on five defining *SPILE* genes in the model limpet species *Nipponacmea fuscoviridis* that have been used for functional studies (Morino et al., 2017; Morino et al., 2022; Morino et al., 2023). Interestingly, the *SPILE-B* clade is distinguished by a duplication of the homeodomain, while the conserved PADRE domain N-terminal to the homeodomain is present in A/C, B, and D clades (Paps et al., 2015; Morino et al., 2017; Barton-Owen et al., 2018; Phuangphong et al., 2021; Morino et al., 2022). Furthermore, *SPILE* genes appear to have been evolving most rapidly in the mollusc and annelid lineages, while less divergence has occurred in platyhelminths and rotifers (Morino et al., 2017; Barton-Owen et al., 2018). Genes in the *SPILE-D* and *SPILE-E* clades duplicate rarely in molluscs, while genes in the *SPILE-A/C* clade have duplicated multiple times in the molluscan lineage (Morino et al., 2022). Interestingly, *SPILE* genes appear to have been lost in the cephalopods altogether, which Morino et al. (2022) surmise is connected to the loss of the spiralian cleavage pattern.

Early analysis of transcriptomic data indicated that the expression of many spiralian-specific genes peaks either during early development or at the trochophore larval stage (Paps et al., 2015). As the name implies, many spiralian undergo spiralian cleavage, in which the zygote first undergoes two rounds of cell division along the animal-vegetal axis, giving rise to four blastomeres A, B, C, and D, termed macromeres (Lambert, 2010; Henry, 2015; Martín-Durán and Marlétaz, 2020). The macromeres then divide and generate the “first quartet” of micromeres positioned diagonally on top of the macromeres (Figure 1.4). Subsequently, the macromeres divide and produce the second and third quartets, while the micromeres divide within their quartet groups (Figure 1.4). In early embryos of molluscs and annelids, the expression of many *SPILE* genes is specific to a subset of blastomeres (Figure 1.4) (Morino et al., 2017; Phuangphong et al., 2021). This led

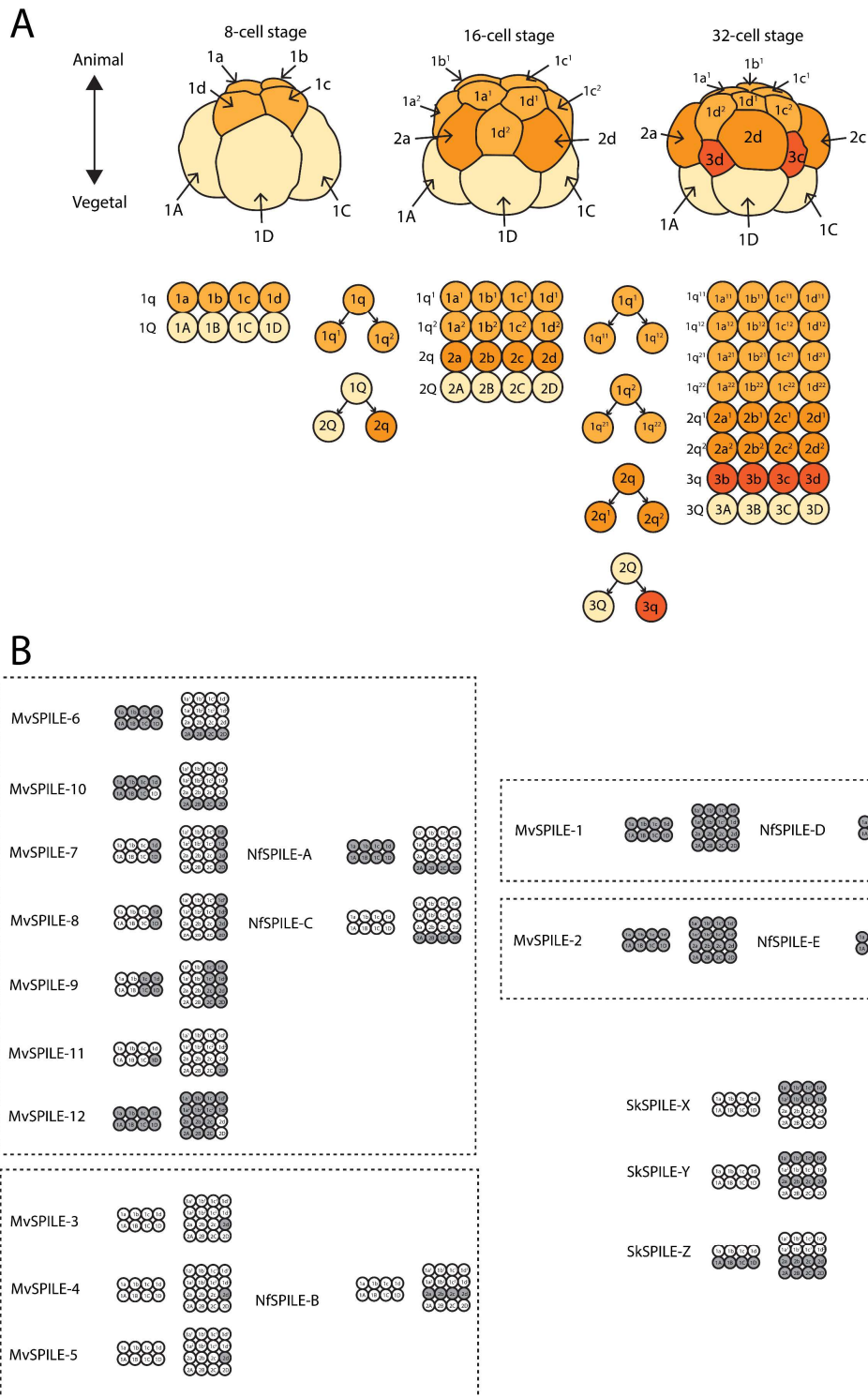


Figure 1.4 Spiralian cleavage and SPiLE gene expression in spiralian embryos. **(A)** Schematic illustration of spiralian cleavage. At the 8-cell stage, the embryo contains 4 macromeres (1Q) and four micromeres (1q) that constitute the first quartet. Macromeres divide and give rise to the second quartet (2q), and then the third quartet (3q). **(B)** Gene expression patterns for SPiLE genes in 8-cell stage and 16-cell stage embryos of *Mytilisepta virgata* (Mv), *Nipponacmea fuscoviridis* (Nf), and *Spirobranchus kraussii* (Morino et al. 2017, Phuangphong et al. 2021). Cells expressing the SPiLE gene are marked in dark grey. SPiLE genes of the same clade are grouped. The diagrams correspond to the schematic cleavage pattern in (A).

to the idea of blastomere-specifying functions for spiralian-specific genes, and under- and overexpression experiments have indicated that at least *SPILE* genes in *N. fuscoviridis* cross-regulate each other's expression patterns as well as the expression patterns of other transcription factors (Morino et al., 2017; Morino et al., 2023). However, the functional role of *SPILE* genes beyond molluscs, including flatworms and planarians, remains to be investigated.

1.2.12 Summary of TALE factors

To conclude, TALE factors are an ancient subfamily of homeobox genes that are deeply incorporated into the meshwork of metazoan life, yet are also underlying evolutionary innovations. Against this background, TALE factors also play their part in patterning the regenerating planarian.

1.3 Patterning in planarian regeneration

Gene regulatory networks require an understanding of the underlying biological system. Here, the dynamics of a repatterning planarian is reviewed to understand the broader context in which Pbx, Prep, and Meis function to drive anterior and posterior regeneration, the subject of the fourth chapter.

1.3.1 Pattern formation in the context of regeneration

How do we make heads and tails of planarian regeneration? Stem cells and their progeny provide the building blocks for regeneration, but these blocks need to be laid in an organised spatiotemporal manner to precisely regenerate the missing tissue. Whole-body regeneration in planarians may be the best-understood model system for studies on regeneration polarity and pattern formation as a part of regeneration. Commonly, regeneration polarity refers to the initial specification of the identity of structures to be regenerated, whereas regenerative pattern formation encompasses patterning events from wounding to a fully regenerated structure. Pattern formation is central to the field of developmental biology; however, regeneration brings the concept to new dimensions (Wolpert et al., 1969; French et al., 1976; Green and Sharpe, 2015). As opposed to the relatively deterministic nature of embryonic development, regeneration begins from a stochastic and variable point, as different types of wounds result in varying proportions in the remaining tissue fragments (Čapek and Müller, 2019; Reddien, 2022). Mechanistically, the process necessitates several critical steps: the quantitative sensing of tissue loss; the re-establishment of the positional information system in accordance with the new body proportions; the coordinated regrowth of the lost tissue; and the integration of the reforming structure with the pre-existing tissue (Čapek and Müller, 2019; Reddien, 2022).

As demonstrated by classical grafting experiments, tissues are able to interact and induce developmental fates through cell-to-cell communication mediated by signalling molecules (Browne, 1909; Spemann and Mangold, 1924). Morphogen gradients and reaction-diffusion systems represent the two main theoretical models of pattern formation, both positing that signaling molecules elicit cellular responses in a concentration-dependent manner (Wolpert, 1969; Turing, 1952; Green and Sharpe, 2015). The concept of a morphogen gradient is intuitive; a morphogen is secreted at the source, transported from the source, and establishes a concentration gradient that instructs cell fate specification (Wolpert et al., 1969; Kicheva and Sharpe, 2023). However, the minimal model of a concentration gradient does not explain how the patterning system remains robust to alterations in the size of the patterning field, i.e. the property of size-invariance (Čapek and Müller, 2019). Thus, the gradient model was improved into a “source and sink” or “synthesis-diffusion-degradation” model, wherein a signaling source and sink ensure that a size-invariant gradient is formed (Čapek and Müller, 2019). Later theoretical advancements have also explained size-invariance by additional mechanisms, including a mobile modulator that modifies degradation and diffusion rates, opposing gradients, and feedback interactions (Čapek and Müller, 2019). The morphogen gradient relies on a prepattern, whereas the reaction-diffusion model puts forth that a diffusible local autoactivator and long-range inhibitor can generate stable concentration patterns from a homogeneous initial point in a self-organising manner (Turing, 1952; Gierer and Meinhardt, 1972). Although sometimes presented as

mutually exclusive, these models are complementary and may act upstream, downstream, and in parallel with each other to create biological patterns (Green and Sharpe, 2015).

Morphogen diffusion provides an elegant theoretical framework for biological patterning, but it does not address the biophysical properties of signalling molecules. How do morphogens operate within the intricate meshwork of tissues? How are morphogen gradients generated in the extracellular medium, how are the morphogens transported, and how is the gradient interpreted, particularly at low concentrations that obscure the interpretation of concentration differences? (Stapornwongkul and Vincent, 2021; Kicheva and Sharpe, 2023; Liberali and Schier, 2024). Given the nature of the extracellular compartment, passive and non-directional diffusion may be too simplistic a mechanism for morphogen transport; yet experimental evidence supports diffusion in many instances (Stapornwongkul and Vincent, 2021; Liberali and Schier, 2024).

To this end, other modes of morphogen transportation have been proposed. In the relay mechanism, morphogens do not diffuse over long distances, but rather activate their own expression in adjacent cells and launch a chain reaction that drives signalling forward. Moreover, extracellular matrix components such as heparan sulfate proteoglycans (HSPGs) and cellular protrusions may also facilitate morphogen transport. In planar transcytosis, a cascade of cell-to-cell exocytoses and endocytoses would steer ligands along the plane of the epithelium. As some morphogens are lipidated and display reduced solubility in the aqueous extracellular medium, alternative transport mechanisms involving lipid-binding chaperones, exosomes, lipoprotein particles, and micelles have been suggested (Stapornwongkul and Vincent, 2021; Kicheva and Sharpe, 2023).

1.3.2 *Entwicklungsmechanik* of planarians informs on pattern formation

Regenerative pattern formation in planarians is studied following different types of amputation. *S. mediterranea* can regenerate heads and tails following transverse amputations anywhere along the AP axis, but the rate of regeneration decreases more posteriorly (Evans et al., 2011). Along the ML axis, thin and thick fragments of *S. mediterranea* regenerate lateral structures after parasagittal amputations, indicating that the positional information of the midline is reformed. Likewise, the worms also heal from a variety of oblique and triangular cuts. Nonetheless, some long-known regeneration abnormalities in planarians inform on the nature of positional mechanisms. Occasionally, thin cross-pieces resulting from transverse cuts regrow heads instead of tails at posterior-facing wounds; these “Janus heads” or heteromorphs attracted the interest of T. H. Morgan when he formulated his views on the mechanisms of regeneration polarity (Morgan, 1898, 1900). Furthermore, partial longitudinal splitting in the head or tail duplicates the body axis and gives rise to two-headed or two-tailed worms if the tissue stumps are not allowed to merge after wounding (Morgan, 1900). In a less well-known case, if the partial split from the posterior is extended forwards to the head, ectopic heads may regenerate at the base of the split (Morgan, 1900). In a similar T-cut, ectopic heads have been observed to form at the anterior-facing cut surfaces (Brøndsted, 1955).

In the first attempt to explain planarian regeneration in mechanistic terms, Morgan theorised the existence of some “structure or composition” in the pre-existing tissue that guides neoblast behavior in the regenerating blastema (Morgan, 1904). For Morgan, the existence of “head stuff” and “tail stuff” did not necessarily imply a gradient but nevertheless some physicochemical mechanism acting differentially at anterior-facing and posterior-facing wounds. Child (1911) became a strong proponent of the idea of a gradient and quantitative differences along the planarian body axis, which he believed were metabolic in nature. In the century that followed, a series of grafting experiments revealed that the cross-transplantation of tissues from different axial positions may cause ectopic outgrowths as a result of intercalary regeneration, restoring the missing positional identity between the donor and recipient tissues (Santos, 1929, 1931; Okada and Sugino, 1934; Schewtschenko, 1936a,b; Okada and Sugino, 1937; Miller, 1938; Sugino, 1940, 1941; Saló and Baguña, 1985; Kobayashi et al., 1999; Kato et al., 2001; Saito et al., 2003; Witchley et al., 2013; Oderberg

et al., 2017). Among the grafts, the anterior pole possesses unique organising properties, as it can induce the growth of a complete ectopic head when transplanted to lateral flank tissues (Oderberg et al., 2017). To sum up, the robust regenerative scenarios in planarians, combined with grafting experiments, suggest the existence of a constitutive positional information system along the body axes of these animals. In the last two decades, the molecular underpinnings of this information system have been investigated.

1.3.3 A molecular coordinate system provides positional information in planarians

In planarians, pattern formation is required both during homeostatic cell turnover and injury-induced restorative regeneration. The positional information system provides a molecular coordinate system along the anteroposterior (AP), dorsoventral (DV), and mediolateral (ML) axes, and is rescaled following injury to accommodate new body proportions. In the planarian literature, position control genes (PCGs) have been defined as genes that are constitutively and regionally expressed along any of the body axes, display patterning phenotypes following knockdowns, or are predicted to encode components of the same signalling pathways included in the set of PCGs (Witchley et al., 2013; Scimone et al., 2016; Reddien et al., 2018; Reddien et al., 2022). In this definition, PCGs include components of signalling pathways (Wnt, BMP, FGFR-like, Slit, Netrin) and certain transcription factors, but exclude signalling molecules orchestrating regeneration that are not expressed in an axially restricted manner in the adult worm (e.g. Hedgehog). Although the same core signalling pathways of the positional information system are shared between homeostasis and injury-induced regeneration, additional factors are involved in reconfiguring positional information into a new tissue context. The positional information system instructs neoblast fate specification (Wurtzel et al., 2017), but how the signalling cues are translated into neoblast behaviour is still incompletely understood.

Along the AP axis, extensive knockdown experiments of Wnt signalling components and inhibitors have revealed that loss of Wnt signalling activity leads to posterior heads, while increased Wnt activity generates ectopic tails at anterior-facing wounds (Petersen and Reddien, 2008; Gurley et al., 2008; Iglesias et al., 2008; Adell et al., 2009; Petersen and Reddien, 2009b; Gurley et al., 2010; Almuedo-Castillo et al., 2011; Iglesias et al., 2011; Petersen and Reddien, 2011; Reuter et al., 2015; Owen et al., 2015; Sureda-Gómez et al., 2015; Scimone et al., 2016; Lander and Petersen 2016; Sureda-Gómez et al., 2016; Stückemann et al., 2017; Tewari et al., 2019; Pascual-Carreras et al., 2021; Pascual-Carreras et al., 2023). In situ hybridisation (ISH) has revealed that certain Wnt ligands are expressed in a posterior-to-anterior gradient, and immunostainings and Western blotting from tissue fragments have indicated the existence of a gradient of Wnt signalling activity (Sureda-Gómez et al., 2015; Scimone et al., 2016). Intriguingly, differences in regenerative abilities of planarians species have been attributed to evolutionary modifications in Wnt signalling dynamics. In the partially regeneration-deficient species such as *Phagocata kawatsui*, *Dendrocoelum lacteum* and *Proctotyla fluvialilis*, inhibition of beta-catenin is able to restore anterior regeneration (Liu et al., 2013; Sikes et al., 2013; Umesono et al., 2013). It has been hypothesised that the requirement for Wnt signalling in the reproductive system may constitute an evolutionary trade-off between robust regenerative abilities and effective sexual reproduction (Vila-Farré et al., 2023).

Furthermore, FGFR-like (FGFRL) genes are expressed in the head and prepharyngeal regions and are involved in the patterning of the head and the trunk (Cebrià et al., 2002; Scimone et al., 2016). Even though components of the Hh and Erk signalling pathways are not expressed in a graded manner in adult worms, they control AP polarity during regeneration. Reduced Hh signalling impairs posterior regeneration, while increased signalling impairs anterior regeneration (Rink et al., 2009; Yazawa et al., 2009). At least in *D. japonica*, Erk activity is biased to the anterior during regeneration (Umesono et al., 2013; Agata et al., 2014).

The planarian DV axis is patterned by BMP signalling (Molina et al., 2007; Orii and Watanabe, 2007; Reddien et al., 2007; Gaviño and Reddien, 2011; Molina et al., 2011). Increased BMP signalling dorsalises planarians, while reduced signalling ventralises the body plan during both homeostasis and injury-induced

regeneration. A cross-inhibitory module between laterally expressed *wnt5* and medially expressed *slit* patterns the ML axis, as silencing *slit* expands lateral identities, while inhibition of *wnt5* extends the midline laterally (Cebrià et al., 2007; Adell et al., 2009; Gurley et al., 2010; Almuedo-Castillo et al., 2011; Oderberg et al., 2017; Atabay et al., 2018). However, it is unclear whether *wnt5* and *slit* regulate each other directly or indirectly.

The majority of the signalling molecules of the molecular coordinate system are expressed in muscle cells. The body wall musculature (BWM) of planarians has three layers: the outer layer of circular fibres along the ML axis, the middle layer of diagonal and thin longitudinal fibres along the AP axis, and an inner layer of thick longitudinal fibres along the AP axis (Cebrià et al., 2016; Scimone et al., 2017). Additionally, dorsoventral muscle fibres extend along the DV axis (Cebrià et al., 2016; Witchley et al., 2017). Most of the signalling molecules (Wnt, BMP, and FGFRL signalling components) are expressed in the body wall musculature, with some PCGs, such as *wntP-2* and *wnt11-2*, expressed throughout several muscle cell types, whereas some genes, such as *wnt11-1* (circular fibres) and *slit* (DV fibres), are restricted to certain subtypes of muscle cells (Witchley et al., 2013; Lander and Petersen 2016; Scimone et al., 2016; Scimone et al., 2017, 2018; Cote et al., 2019). Functional muscle fibres are necessary for proper regeneration, as the loss of longitudinal fibres eliminates the formation of both anterior (*notum*+) and posterior (*wnt1*+) poles, while the loss of circular fibres bifurcates the AP axis (Scimone et al., 2017). Additionally, disruption of DDX24 (DEAD box RNA helicase) and Phred-1 (a divergent neurexin-1 homolog), which regulate proper muscle biology, also leads to patterning defects in regenerating worms (Sarkar et al., 2022; Adler and Sánchez Alvarado, 2017).

How the patterning systems of different body axes interact is poorly understood, or whether any of the body axes exert a primary directive role. Recently, it has been discovered that BMP signalling regulates the localisation of the *wnt1* expression domain at the posterior pole during homeostasis, while inhibition of *wnt1* shifts *bmp4* expression patterns (Clark and Petersen, 2023). Additionally, *bmp4* is involved in ML patterning by activating *slit* and repressing *wnt5* expression (Clark and Petersen, 2023). Therefore, intricate cross-regulations appear to contribute to the robustness of the positional information system.

In addition to signalling molecules, the bioelectrical state of planarians also influences regeneration outcomes. The bioelectrical state of tissues arises from ion concentration differences across cell membranes, and it can influence cell behaviour during both development and regeneration (Harris et al., 2021). In planarians, perturbations to the function of gap junctions and ion channels, with subsequent alterations in electrical potentials, may cause misspecification of regeneration polarity (Nogi and Levin, 2005; Beane et al., 2011; Zhang et al., 2011; Durant et al., 2016; Durant et al., 2019). However, the coordination between bioelectrical and molecular signalling in planarians in carrying out proper regeneration is still poorly understood.

1.3.4 Signalling pathways in planarian regeneration

To further understand the signalling pathways involved in planarian regeneration, we will first reduce the problem by examining the components of the different signalling pathways and the knockdown phenotypes generated over two decades of research in molecular genetics.

1.3.4.1 Wnt signalling

Wnt signalling proceeds from the production and secretion of Wnt ligands from the sending cell to intercellular transport and signal transduction in the receiving cell. In the endoplasmic reticulum of the sending cell, membrane-bound Porcupine adds lipid moieties to the Wnt ligands, which are then delivered by the transport protein EVI/WLS to the vesicular transport machinery and Golgi apparatus (Holzem et al., 2024). From the Golgi apparatus, Wnt ligands are secreted to the extracellular compartment, wherein several mechanisms have been proposed to aid in the transport of the lipophilic Wnt ligands, including

vesicular transport on exosomes, attaching lipoproteins, HSPGs, and cell protrusions (Holzem et al., 2024). During morphogen transport, the activity of Wnt ligands can also be inhibited by a number of antagonists, including Notum, WIF1, and DKK (Holzem et al., 2024).

In the receiving cell, the Wnt signal is transduced through either the canonical beta-catenin-dependent pathway, the canonical Wnt/STOP pathway, the non-canonical planar cell polarity pathway, or the calcium-dependent pathway (Holzem et al., 2024). In the beta-catenin-dependent pathway, the Wnt ligand binds to and brings together the receptors Frizzled and LRP5/6, and assemble the signalosome complex (Dsh, Axin, CD1alpha, GSK3). This signalosome complex breaks up the degradation complex, which degrades beta-catenin in the absence of the activating Wnt ligand, allowing the co-repressor Groucho to enter the nucleus and repress the activation of Wnt-responsive genes. When not degraded, beta-catenin then enters the nucleus and turns the LEF/TCF factors into activators of gene expression. In the PCP pathway, Wnt ligands bind to the receptor pair Frizzled-Ror, activating Rho GTPases, which in turn activate mainly kinesins and cytoskeletal binding proteins that drive changes in cell shape and movement. In the calcium-dependent pathway, Wnt binds to Ryk, activates a phospholipase that frees calcium ions into the cytoplasm, where they act as secondary messengers to modulate the activity of various enzymes, transcription factors, and translation factors.

In planarians, the various subcomponents of the Wnt pathway contribute to proper signalling activity. The genome of *S. mediterranea* contains a single *wntless*, nine *wnt* ligands, three *sFRP* inhibitors, a single *notum*, nine *frizzled* receptors, three *hfp* receptors, two *beta-catenin* genes, a two *APC* genes, and two *axin* genes (Gurley et al., 2008; Iglesias et al., 2008; Petersen and Reddien, 2008; Adell et al., 2009; Gurley et al., 2010; Iglesias et al., 2011; Witchley et al., 2013; Scimone et al., 2016; Stückemann et al., 2017; Pascual-Carreras et al., 2021). Of the Wnt ligands, *wnt-1*, *wnt11-1*, *wnt11-2*, and *wnt11-5/wntP-2* are expressed in a posterior-to-anterior gradient in intact adult worms (Gurley et al., 2008; Petersen and Reddien, 2008; Adell et al., 2009; Gurley et al., 2010; Sureda-Gómez et al., 2015). Among these, *wnt1* is expressed in the posterior pole, *wnt11-1* and *wnt11-2* are expressed from the posterior pole to the pharynx, and *wnt11-5/wntP-2* is expressed from posterior pole to the pre-pharyngeal region. In contrast, Wnt inhibitors are expressed in the anterior: *notum* in the anterior pole, *sFRP-1* in the far anterior, and *sFRP-2* in the anterior half (Gurley et al., 2010; Petersen and Reddien, 2008; Petersen and Reddien, 2011). Although the *beta-catenin-1* transcript is present ubiquitously along the AP axis, the nuclear localisation of the beta-catenin-1 protein displays a posterior-to-anterior gradient in the intact adults (Gurley et al., 2008; Iglesias et al., 2008; Petersen and Reddien, 2008; Sureda-Gómez et al., 2016; Stückemann et al., 2017). In contrast, *beta-catenin-2* is involved in cell adhesion rather than in Wnt signalling (Chai et al., 2010).

Among the Wnt signalling components, knockdown of the intracellular molecules *beta-catenin-1* and *APC-1* results in the strongest phenotypes. Silencing of *beta-catenin-1* reverses polarity during regeneration and gives rise to two-headed worms, while knockdown of *APC-1* has the opposite effect and results in two-tailed worms. Interestingly, knockdown of *beta-catenin-1* is able to disrupt polarity during homeostasis and give rise to “radial-like hypercephalized” worms, whereas inhibition of *APC-1* does not eliminate the anterior brain or result in “hypercaudal” worms (Gurley et al., 2008).

To understand the main driver of *beta-catenin-1* signalling, the posteriorly expressed Wnt ligands can be sorted into a posterior “Wnt code” with respect to the severity in regeneration outcomes after knockdown. Knockdown of *wnt1* results in two-headed or tailless planarians (Petersen and Reddien, 2009b; Gurley et al., 2010; Sureda-Gómez et al., 2015), *wnt11-2(RNAi)* worms display the tailless phenotype with no duplicated pharynx (Sureda-Gómez et al., 2015), in *wnt11-1(RNAi)* worms the tails are shorter and pharynx sometimes duplicates, and in *wnt11-5/wntP-2(RNAi)* worms the tail is normal but the pharynx duplicates. Thus, *wnt1* appears to be the primary ligand that drives beta-catenin signalling during regeneration, whereas *wnt11-2* and *wntP-2* appear to have secondary roles later on in the regeneration process. Given phenotype

data and expression patterns, the receptors *frz**d1*, *frz**d4-1*, and *lhp**5/6* are transducing the posterior Wnt signals (Pascual-Carreras et al. 2021).

1.3.4.2 Hedgehog signalling

In the Hedgehog signalling pathway, the Hh ligand is lipidated (cholesterol and palmitate) and bound to the membrane before being extracted from the membrane by DISP1 and shielded by SCUBE2, which ensures the transport of the Hh ligand to the receiving cell (Zhang and Beachy, 2023). There, the Hh ligand binds to the receptor Patched (Ptc), which suppresses the transmembrane protein Smoothened (Zhang and Beachy, 2023). In the absence of the Hh signal, Ptc inhibits Smoothened (Smo) activity and the Gli transcription factors are sequestered by Kif7 and Sufu and phosphorylated by PKA and CK1, which leads to proteasomal degradation, yielding the repressor form (GLIR). When Smo is no longer inhibited, the Gli proteins are protected from proteolytic processing and enter the nucleus in their active form (GLIA) (Zhang and Beachy, 2023). Interestingly, the signalling dynamics in vertebrates and flies display differences in the requirement of the SuFu protein and the primary cilium (Lee et al., 2016).

The genome of *S. mediterranea* contains a single *hh* ligand, a single *ptc* receptor, a single *smo* receptor, a single *sufu*, and three *gli* genes, of which only *gli-1* is involved in Hh signalling (Rink et al., 2009). Neither cilia nor the signal transduction components Cos2, Kif2, Kif7, Fused, and Iguana are required for functional Hh signalling in *S. mediterranea* (Rink et al., 2009). Although Hh signalling genes are not expressed in a graded fashion in intact worms, Hh signalling is nonetheless required for proper regeneration. Increased Hh signalling levels (knockdown of *ptc* or *sufu*) increase mitotic activity and result in a headless phenotype, while decreased Hh signalling (knockdown of *hh*, *smo*, and *gli-1*) decreases cell proliferation levels and leads to a tailless phenotype (Rink et al., 2009; Yazawa et al., 2009). Alterations in Hh signalling affects Wnt signalling during regeneration but not in intact adults, indicating that the patterning effect of Hh signalling is regeneration-specific (Rink et al., 2009; Yazawa et al., 2009).

1.3.4.3 Growth factors: FGF, ERK and JNK signalling

The role of growth factors encompassing the FGFR, ERK, and JNK signalling pathways is the most enigmatic in planarian regeneration. FGFs bind to their tyrosinase kinase receptors (FGFRs) and activate downstream signalling pathways including MAPK/ERK, PI3K/AKT, PLC γ , and STAT (Xie et al., 2020). However, FGFRs are atypical FGFRs that lack the intracellular tyrosine kinase domain, and the downstream transduction signalling mechanism is poorly understood, even though the FGFR family is conserved throughout metazoans (Bertrand et al., 2009; Trueb, 2011). As the extracellular part resembles FGFRs, the FGFR family may act as decoy receptors that inhibit FGF signalling (Trueb, 2011). In turn, the MAPK/ERK signal transduction cascade includes a chain reaction of phosphorylation: MAP3K, MAP2K/MEK, and MAPK/ERK. The MAPK family of kinases includes three subfamilies: ERKs, JNKs, and p38s (Lavoie et al., 2020). ERK induces the expression of immediate early genes including members JUN, FOS, MYC, and EGR families, and mostly drive a cell proliferation program (Lavoie et al., 2020). Meanwhile, the JNK signalling pathway mainly activates the AP-1 transcription factor and has been implicated in both cell proliferation and cell death (Weston and Davies, 2007).

In *S. mediterranea*, the FGFR family includes the defining member *nou-darake* (*ndk*, “brains everywhere” in Japanese) and five *noudarake-like* genes, *ndl-1* to *ndl-5* (Cebrià et al., 2002; Scimone et al., 2016). All FGFR genes are expressed in the anterior, and loss of *ndk* expands brain tissues towards the posterior both during homeostasis and in regenerating worms without perturbing trunk identity (Cebrià et al., 2002; Scimone et al., 2016). As *ndk* can bind FGF in the *Xenopus* embryos, this supports the decoy receptor model in which *ndk* binds to an diffusing FGF ligand and restricts its signalling activity to the anterior (Adell et al., 2010; Agata et al., 2014). However, the presence of a FGF ligand has not been established, and only recently has a FGF ligand been identified in *D. japonica* (Auwal et al., 2020). However, although *fgf* is expressed in the anterior and is upregulated following amputation, knockdown of *fgf* does not result in regeneration defects

as the current model would predict (Auwal et al., 2020). Therefore, further research into the FGF-FGFR-FGFRL problem is required in planarians.

The ERK pathway is rapidly activated at wounds and required for wound-induced gene expression, later becoming restricted to the anterior (see 1.3.5.1 Early generic wound response). The JNK signalling components *mapk7/hem*, *jnk*, and two transcription factors *juml-1* and *juml-2* have been described in the genome of *S. mediterranea*, and Jnk signalling is required for coordinating mitosis, apoptosis as well as posterior regeneration (Wagner et al., 2012; Wenemoser et al., 2012; Almuedo-Castillo et al., 2014; Tejada-Romero et al., 2015).

1.3.4.4 TGF- β superfamily

The TGF- β superfamily includes subfamilies TGF- β , BMP, activin and inhibin, and Lefty (Kitisin et al., 2007). The TGF- β signals bind to type I and type II receptors that are serine-threonine kinases (Kitisin et al., 2007). The type II receptors include BMPRII, ActRIIA, ActRIIB, and TGF β RII. The type I receptors include seven members, ALK1 to ALK7. The TGF- β ligands binds to type II receptors, leading to the phosphorylation and activation of type I receptors. Type I receptors phosphorylate receptor-activated Smads of the signal transduction cascade. TGF- β and activin activates Smad2 and 3, while BMPs activates Smad1, 5, and 8. The receptor-activated Smads, along with the common Smad4, form the R-Smad-Smad4 complex that enters the nucleus and triggers gene expression. Noggin and chordin are extracellular inhibitors of BMP signalling, while follistatin (fst) can inhibit both BMPs and activin.

In *S. mediterranea*, two activins *activin-1* and *activin-2*, a single receptor *actR*, and a single *smad2/3* have been described (Gaviño et al., 2013; Roberts-Galbraith et al., 2013; Cloutier et al., 2021). As for the BMP signalling pathways, the genes *bmp1*, *bmp4*, *admp-1a*, *admp-1b*, *smad1*, *smad4*, *noggin-1*, *noggin-2*, and eight *noggin-like* (*nlg1* to *nlg8*) have been identified (Molina et al., 2007; Reddien et al., 2007; Gaviño and Reddien, 2011; Molina et al., 2011). *Bmp4* is expressed in a dorsal-to-ventral gradient centred at the dorsal midline, while *admp-1* (i.e. *admp-1a/b*) is expressed laterally and ventrally (Reddien et al., 2007; Molina et al., 2007; Gaviño and Reddien, 2011; Molina et al., 2011). The *bmp/admp* regulatory circuit patterns the DV axis both during regeneration and homeostasis (Molina et al., 2007; Orii and Watanabe, 2007; Reddien et al., 2007; Gaviño and Reddien, 2011; Molina et al., 2011). During homeostasis, the most severe phenotypes develop an ectopic CNS on top of the pre-existing CNS (Cebrià et al., 2007; Molina et al., 2007; Reddien et al., 2007; Gaviño and Reddien, 2011; Molina et al., 2011). However, it is unclear whether *admp-1* regulates *bmp4* directly or indirectly, but *admp-1* expression is negatively regulated by Bmp signaling during homeostasis and growth (Gaviño and Reddien, 2011).

1.3.4.5 Hox genes

The genome of *S. mediterranea* harbours thirteen Hox genes, the anterior (*box1*, *box3a*, *box3b*, *box4a*, and *box4b*), the central (*lox4a*, *lox5a*, and *lox5b*), and the posterior (*post-1*, *post-2a*, *post-2b*, *post-2c*, *post-2d*). The first planarian Hox genes were first cloned already 30 years ago (García-Fernández et al., 1991; Balavoine and Telford, 1995), but knockdown of them failed to give visible phenotypes for a long time, despite the fact that five Hox genes, *box3b*, *box4b*, *lox5a*, *post2c*, *post2d*, display axially restricted expression in adult worms (Currie et al., 2016; Scimone et al. 2016). However, recent experiments have implicated a role for *post-2d* in posterior regeneration (Tewari et al., 2019). Instead, it appears that some of the planarian Hox genes (*box1*, *box3a*, *box3b*, *lox5b*, and *post2b*) are required for asexual reproduction by spacing cryptic fission planes along the AP axis (Arnold et al., 2019; Arnold et al., 2021).

1.3.4.6 Hippo/Yorkie pathway and kinases

The Hippo/Yorkie pathway is a conserved signalling pathway in development that regulates organ size and cell fate (Manning et al., 2020). Cell adhesion and cell polarity influence the activity of the core kinase cassette, which includes Hippo (mammalian MST1 and MST2), salvador (SAV1), Mats (MOBK1A and

MOBKL1B) and Warts (LATS1 and LATS2). The kinase cassette normally inhibits the downstream transcription co-activator Yki (YAP and TAZ) and the partner transcription factor Sd (TEAD1/2/3/4), but upon activation, Yki/YAP/TAZ enters the nucleus (Manning et al., 2020).

In planarians, Yki limits stem cell proliferation and has been implicated in axial regeneration and the rescaling of the positional information system after injury (Lin and Pearson, 2014, 2017). Recently, it was discovered that the Hippo signalling components Merlin and Warts are required for the function of *pak1* (p21-activated kinase), which inhibits beta-catenin-dependent Wnt signalling along the AP axis and acts synergistically with beta-catenin-independent Wnt signalling along the ML axis (Doddihall et al., 2024). The STRIPAK complex is a multimolecular structure comprising kinases and phosphatases and is involved in signalling pathways such as Hippo, MAPK/ERK, cytoskeleton modelling (Shi et al., 2016). In planarians, STRIPAK complex components restrict *wnt1*+ cells to the tail tip and are thus involved in axial scaling (Schad and Petersen, 2020). Furthermore, the non-receptor tyrosine kinase *src-1* is a global repressor of anterior identities in parallel or downstream of Wnt/FGFR signalling pathways (Bonar et al., 2022).

1.3.5 Regeneration cascade

In planarians, injury-induced regeneration can be divided into three phases: the early phase (0-18 hpa), which involves the expression of generic wound response genes; the second phase (18-48 hpa), during which positional information is re-established; and the late phase (48 hpa onwards), which involves fate-specific cell differentiation and morphogenesis. In the early phase, muscles contract and seal the wound, a thin epidermal film covers the wound, and the basal membrane reforms. In the second phase, asymmetric *notum* expression at anterior-facing wounds inhibits Wnt activity and rescales the positional information system. In the late phase, the blastemal outgrowth is fuelled by neoblast progeny and orchestrated by distal blastemal poles that orchestrate morphogenesis. Over the past two decades, advancements in developmental genetics and transcriptomics have helped to unravel the genes regulating different phases of regeneration. Recently, single-cell transcriptomics of regenerating worms have mapped out transient regeneration-activated cell states, while a three-dimensional spatiotemporal transcriptomic atlas has further informed on the genes involved in the regeneration process (Benham-Pyle et al., 2021; Cui et al., 2023b).

1.3.5.1 Early generic wound response

The early generic wound response involves the expression of around 200 wound response genes, a general mitotic peak in neoblasts throughout the worm, and a local apoptotic response in differentiated cells at the wound site (Baguña, 1976; Saló and Baguña, 1984; Pellettieri et al., 2010; Wenemoser and Reddien, 2010; Wenemoser et al., 2012; Wurtzel et al., 2015). The early phase is initiated by all types of injuries, including small incisions, and is stem cell-independent, while the missing tissue response occurs following tissue loss. If tissue is lost, the missing tissue response is initiated, which includes the extended expression of the wound response genes beyond 24 hpa, a second mitotic peak in neoblasts at the wound site, and a systemic apoptotic response in differentiated tissues (Pellettieri et al., 2010; Wenemoser and Reddien, 2010; Wenemoser et al., 2012). However, it seems that disruption of the positional information system primarily drives regeneration and the missing tissue response mainly accelerates it, as planarians can regenerate slowly even in the absence of the three hallmarks of the MTR, thanks to basal turnover rates in the neoblasts (Tewari et al., 2018). Incisions in patterning-deficient backgrounds (*beta-catenin-1(RNAi)*, *APC-1(RNAi)*, and *patched(RNAi)*) cause ectopic outgrowths, further suggesting that an early generic wound response occurring in a disrupted positional system launches the missing tissue response and restorative regeneration (Owlarn et al., 2017).

The gene expression program includes the upregulation of around 200 genes (Wenemoser and Reddien, 2012; Wurtzel et al., 2015). The very first genes include “immediate early genes” such as *jun-1*, *fos-1*, *egl-1*, *egr-2*, *egr-3*, and *egr-4*, whose expression does not depend on new protein synthesis. The program also includes broadly expressed stress-related genes (*HSP90*, *HSP70*, *HSP40*, *egr-2*), a neoblast subprogram

involving histone methyltransferases, cell cycle factors, chromatin remodelling factors, and transcription factors (*H2B*, *topbp1*, *rrm2b*, *runt-1*, *strf*, *runt-2*), and an epidermis program (*jun-1*, *TRAF-1*, *ston*, *hadrian*), and the muscle subprogram (patterning genes, i.e. members of Wnt, BMP, and TGF-beta pathways), whose expression is translation-dependent (Wenemoser et al., 2012; Wurtzel et al., 2015).

Early wound-induced gene expression and the missing tissue response are dependent on Erk signalling (Tasaki et al., 2011a; Owlarn et al., 2017, Fan et al., 2023). Erk activates the missing tissue response through *fst* induction (Owlarn et al., 2017). After injury, ERK signalling activity propagates at a speed of 1 mm/h from the wound site along longitudinal muscle fibres to areas distant from the wound (Fan et al., 2023). Interestingly, wound-induced gene expression occurs both at distal and proximal sites, and the distal response is dependent on Erk signalling (Fan et al., 2023). At distal sites, ERK signalling is required for the mitotic response in neoblasts, but the functional link is unclear (Fan et al., 2023). Moreover, the ERK signal might hypothetically elicit a reaction in distal tissues that informs on body proportions to the proximal wound site, but no backwards-propagating signal has yet been characterised (Fan et al., 2023). If ERK signalling is inhibited during the onset of regeneration, regeneration-deficient dormant tissue fragments are formed, but if the generic wound response is triggered again in dormant tails, functional regeneration occurs (Owlarn et al., 2017; Fan et al., 2023). As regeneration proceeds, the pERK signal is shifted from differentiated cells to neoblast-derived blastemal cells at 24 hpa (Tasaki et al., 2011a). Additionally, JNK signalling is active in the postblastemal region and is required to properly induce the expression of early wound-induced genes, the control of the two mitotic peaks, and both of the apoptotic responses (Tasaki et al., 2011b; Almuedo-Castillo et al., 2014). Although JNK signalling affects the early generic phase, it is only strictly required for the onset of the missing tissue response (Almuedo-Castillo et al., 2014).

Interestingly, a subprogram of wound-induced gene expression includes the expression of the patterning genes *fst*, *notum* and *wnt1* in pre-existing muscle fibres at all wound sites (6 hpa to 18 hpa) (Petersen and Reddien, 2009; Gurley et al., 2010; Gaviño et al., 2013; Scimone et al., 2017). Induction of early expression of *wnt1* (0-24 hpa) is independent of *beta-catenin-1* or *APC-1*, while wound-induced expression of *notum* depends on *beta-catenin-1* (Petersen and Reddien, 2009; Rink et al., 2009). Furthermore, as indicated earlier, wound-induced *fst* is required for the missing tissue response, but in a separate regulatory gate, *fst* suppresses *wnt1* at anterior-facing wounds by inhibiting *activin-1* or *activin-2* that likely transduce their signal through a *actR-1* and *smad2/3* circuit (Gaviño et al., 2013; Roberts-Galbraith et al., 2013; Tewari et al., 2018). Additionally, wounding induces the expression of *foxD* (6 hpa) at the ventral midline muscle cells if the midline is perturbed (Scimone et al., 2014a). Interestingly, Hh signalling inhibits wound-induced *foxD* expression, while wound-induced *foxD* expression is independent of Wnt signalling levels (Scimone et al., 2014a).

1.3.5.3 Rescaling of positional information

The start of the second phase (18 hpa to 48 hpa) is marked by re-establishment of the positional information system, as observed by the rescaling of critical PCGs to correspond to the body proportions of the tissue fragment. Resetting of positional information is required for regeneration, as the loss of muscle fibers and PCG expression disrupts the second phase (Scimone et al., 2017). Positional rescaling is also required for remodelling tissues in the pre-existing fragment (Umesono et al., 2013; Lander and Petersen, 2016). During the second phase (18 hpa), neoblasts migrate and accumulate in the postblastemal region of the wound site, where they divide, differentiate, and neoblast progeny feed the outgrowth of the blastema (Dubois, 1949; Saló and Baguña, 1984; Newmark and Sánchez Alvarado, 2000; Eisenhoffer et al., 2008; Wenemoser and Reddien, 2010).

The onset of axial re-establishment is marked by the asymmetric expression of *notum* in muscle cells at anterior wound sites, which is the first gene to be differentially expressed at anterior- vs posterior-facing wounds (Petersen and Reddien, 2011; Wurtzel et al., 2015). Anterior expression of *notum* reduces the level

of Wnt activity at the anterior, reestablishing a posterior-to-anterior gradient corresponding to the regenerating fragment. As regeneration proceeds, the expression of *notum* becomes restricted to the stem cell-derived anterior pole, which acts as a signalling center, patterning the outgrowing head (Petersen and Reddien, 2011; Witchley et al., 2013; Vásquez-Doorman and Petersen, 2014). Although early *notum* expression is dependent on *beta-catenin-1*, a head is able to form in *notum;beta-catenin-1(RNAi)* double knockdowns, indicating that low beta-catenin signalling levels are sufficient to drive anterior regeneration. Conversely, at posterior wound sites, *wnt1* expression becomes progressively restricted to the stem cell-derived posterior pole that drives posterior specification (Petersen and Reddien, 2009; Gurley et al., 2010; Hayashi et al., 2011). However, although differential Wnt activity provides an elegant model for anterior vs. posterior patterning, beta-catenin-1 is in fact observed to be nuclearised in both anterior and posterior blastemas from 12-24 hpa onwards (Sureda-Gómez et al., 2016). Nuclearisation is dependent on *wnt1* in the posterior blastema but not in the anterior (Sureda-Gómez et al., 2016).

However, what signal controls *notum* expression in anterior vs. posterior wounds? Recent efforts have begun to find upstream regulators of *notum*. Knockdown of *activin-2* eliminates asymmetric expression of *notum* and is required to prevent *notum* from being activated at posterior-facing wounds (Cloutier et al., 2021). *Activin-2* is expressed in muscle cells, and disruption of the longitudinal muscle fibers, in combination with knockdown of *activin-2*, rescues asymmetric expression of *notum*, suggesting that *activin-2* somehow modifies the synthesis of longitudinal muscle fibers and their inherent polarity competence, which is used during axis re-establishment (Cloutier et al., 2021). In addition, the transcription factor *alx-3* appears to be required for *notum* expression in the anterior (Akheralie et al., 2023).

During the second phase, ERK signalling activity becomes restricted to the anterior blastema, at least in the species *D. japonica* (Umesono et al., 2013; Agata et al., 2014). Knockdown of *beta-catenin* increased the ERK levels in both anterior and posterior blastemas, indicating that *beta-catenin* has the capacity to downregulate ERK signalling levels throughout the worm (Umesono et al., 2013). Based on these data, Agata et al. (2014) propose that the default trajectory of neoblasts is towards the anterior fate and brain neurons, while Wnt signalling suppresses this in the posterior end by inhibiting ERK activity levels. Furthermore, the anteriorly expressed *ndk* also has the ability to suppress ERK signalling (Umesono et al., 2013). Knockdown of *ndk* expands the ERK signalling boundary from the anterior blastema towards the pharynx and displays a posterior expansion of brain tissues towards the pharynx. Based on these findings, Umesono et al. (2013) hypothesise that an anterior FGF signal activates ERK signalling. Since *ndk* does not have the intracellular signalling domain, it would act as a decoy receptor, trapping FGF ligands and ensuring that FGF signalling activity does not extend beyond the brain towards the posterior. However, this model has its limitations; an *fgf* ligand has been characterised in *D. japonica*, but it does not seem to have as a critical role as posited by the model (Auwal et al., 2020).

Another line of evidence for anterior fate bias in neoblasts comes from the observation that early brain development appears to be uncoupled from Wnt signalling (Evans et al., 2011; Iglesias et al., 2011). Following increased Wnt signalling levels, small brain primordia form in the regenerated ectopic tail of two-tailed worms, but these do not mature into a functional brain (Evans et al., 2011; Iglesias et al., 2011). In wildtype worms, the brain primordia form at the distal tip of the blastema, while in two-tailed worms the primordia form at the boundary between the blastema and the postblastema (Cebrià et al., 2002; Iglesias et al., 2011). Therefore, neoblast may differentiate towards an anterior fate in the postblastemal region independently of Wnt signalling, but a low Wnt signalling environment further allows subsequent differentiation at the anterior end. As the brain primordia appear close to the pharynx in homeostatic two-tailed worms, Iglesias et al. (2011) also speculate that the pharynx might act as an additional signalling center, motivated by the fact that several signalling factors are expressed in the esophageal region.

In the regenerating tail, Hh signalling promotes the expression of *wnt1* and drives the posterior specification program, while increased levels of Hh signalling do not affect the expression of *notum* (Rink et al., 2009; Yazawa et al., 2009; Petersen and Reddien, 2011). At 18 hpa, *bb* is expressed in ventral nerve cords and *wnt1* in muscle cells, suggesting that the Hh signal might be transported via neural axons to the target cells (Yazawa et al., 2009). Differences in regenerative rate along the AP axis is dependent on Hh signalling, suggesting that Hh signalling temporarily inhibits anterior regeneration (Evans et al., 2011). Furthermore, JNK signalling ensures that the anterior and posterior poles arise from neoblasts (Almuedo-Castillo et al., 2014; Tejada-Romero et al., 2015). Specifically, neoblast-expressed *juml-1* is required for specifying the posterior pole lineage (Tejada-Romero et al., 2015).

In addition to the dynamic signalling pathways, a range of transcription factors, including the TALE factors, begin to specify the blastemal poles at the end of the second phase. The nuclear receptor *nr4-a* is expressed in muscle fibres and is necessary for the positioning of both the anterior and posterior poles to the distal tips of the outgrowing blastemas (Li et al., 2019). In the anterior blastema, *notum*, *fst*, and *foxD* are first expressed in the muscle fibres, while *zic-1* is expressed in neoblasts. Towards the end of the second phase, the expression of all these factors becomes restricted to a subset of differentiating neoblast progeny that give rise to the anterior pole (Gaviño et al., 2013; Scimone et al., 2014a; Vázquez-Doorman and Petersen, 2014; Vogg et al., 2014). In the second phase blastema (24 hpa), the number of anterior pole progenitors (*zic-1*+) is dependent on low levels of Wnt signalling and the activity of *pbx* and *fst* (Vázquez-Doorman and Petersen, 2014). Furthermore, the factors *prep*, *pbx*, *islet-1* and *pitx* are expressed diffusely in both anterior and posterior blastemas, and are involved in the specification of the anterior (*prep*, *pbx*) and posterior (*pbx*, *islet*, *pitx*) poles (Felix and Aboobaker, 2010; Hayashi et al., 2011; Blassberg et al., 2013; Chen et al., 2013; Currie and Pearson, 2013; März et al., 2013).

1.3.5.4 Anterior specification and morphogenesis

In the third phase (48 hpa onwards), the poles assemble at the tips of blastemas and act as signalling centers, patterning the blastema along with signalling cues from the pre-existing fragments. As transplanted anterior poles can induce the ectopic outgrowth of heads, this indicates that they provide signalling cues along all the three body axes (Oderberg et al., 2017). Anterior pole progenitors differentiate while migrating towards the tip of the blastema, where they coalesce into a cluster (Scimone et al., 2014a; Vogg et al., 2014). The muscle fibers of the pre-existing fragment are required for the formation and organisation of the anterior pole, as the loss of longitudinal fibers eliminates pole formation, while disruption in the circular fibres splits the pole into two and gives rise to bifurcated worms (Scimone et al., 2017). Moreover, signals from the pre-existing fragment direct the positioning of the pole at the pre-existing midline and DV boundary (Oderberg et al., 2017). Initially (48 hpa), the pole progenitors are specified from neoblasts at the ML median plane but broadly along the DV axis, before progressively accumulating at the DV median plane (Oderberg et al., 2017). The anterior pole is then required to rescale the expression boundaries of PCGs (Vázquez-Doorman and Petersen, 2014).

In third-stage blastemas, the expression of the markers *notum*, *fst*, *foxD*, and *zic-1* is restricted to the anterior pole (Koinuma et al., 2003; Roberts-Galbraith and Newmark, 2013; Scimone et al., 2014a; Vázquez-Doorman and Petersen, 2014; Vogg et al., 2014). Other genes expressed in the anterior pole include *sFRP-1*, *ndk*, *ndl-4*, and *prep*, but not *sFRP-2* or the midline and DV boundary markers *slit* and *admp-1* (Scimone et al., 2014a). Knockdown experiments have revealed that the formation of the anterior pole is dependent on *notum*, *fst*, *foxD*, *zic-1*, *prep*, and *pbx* (Felix and Aboobaker, 2010; Blassberg et al., 2013; Chen et al., 2013; Scimone et al., 2014a; Vázquez-Doorman and Petersen, 2014; Vogg et al., 2014). Among these factors, homeostatic knockdowns suggest that *foxD* and *zic-1* act upstream of *notum* and *fst* in the pole specification cascade (Vogg et al., 2014). Even though *notum* expression is eliminated after double-knockdown of *zic-1* and *beta-catenin-1*, heads form at both ends, suggesting that *zic-1* and *notum* are dispensable in a low Wnt

environment (Vásquez-Doorman and Petersen, 2014). Although *pitx* and *islet-1* are expressed at both anterior and posterior poles, they are required only for the formation of the posterior pole (Currie and Pearson, 2013; März et al., 2013). Nonetheless, *pitx* and *islet-1* along with *foxD*, *pbx*, and *prep* are required for extending the midline (*slit*+ cells) from the anterior pole to the postblastema in the pre-existing fragment (Currie and Pearson, 2013; März et al., 2013; Scimone et al., 2014a). As only around 50 transcripts are differentially regulated between the anterior pole and the flanking tissues, it appears that the specification and function of the pole can be reduced to a set of critical transcription factors mentioned above (Li et al., 2019).

1.3.5.5 Posterior pole and specification

Likewise, the posterior pole forms at the tip of the posterior blastema and expresses the markers *wnt1*, *islet-1*, and *pitx* (Petersen and Reddien, 2009; Hayashi et al., 2011; Currie and Pearson, 2013; März et al., 2013). The formation of the posterior pole is also dependent on the activity of these, and *pitx* is further required for the formation of the midline between the pole and the postblastema in pre-existing fragment as in the anterior blastema (Petersen and Reddien, 2009; Hayashi et al., 2011; Currie and Pearson, 2013; März et al., 2013). In the posterior, transcriptomic profiling has identified around 200 pole-specific transcripts compared to adjacent tissues, suggesting that the signalling ligands and transcription factors regulating the specification of the posterior pole have also been mostly found (Li et al., 2019).

According to expression patterns and phenotypic data following knockdowns, a hierarchical “Wnt code” model has been proposed. Early injury-induced *wnt1* expression in pre-existing muscle cells at the end of the second phase induces the late-stage expression of *wnt1* in a subsets of neoblasts and initiates the posterior specification program. As the receptors *frz4-1* and *lrp5/6* display the strongest phenotypes, this could indicate that these receptors mediate the signal of *wnt1* to specify the posterior pole. In addition to *wnt1*, the Wnt ligands *wnt11-2*, *wnt11-1*, and *wntP-2* are expressed in the posterior blastema. As the knockdown of these ligands display weaker phenotypes than that of *wnt1*, it is likely that these act downstream of the *wnt1-frz4-1-lrp5/6* circuit to specify posterior identities. *Wnt11-2* is also required for the specification of the posterior pole, while silencing *wnt11-1* gives rise to shorter tails but does not eliminate the formation of the posterior pole, and knockdown of *wntP-2* does not affect tail regeneration but is involved in patterning the trunk (Adell et al., 2009; Gurley et al., 2010; Sureda-Gómez et al., 2015). The knockdown of *frz4-1* also produces tailless animals, suggesting that *wnt11-2* could transduce its signal through this receptor (Pascual-Carreras et al., 2021). In the third stage, the expression of *wntP-2* is first induced in the posterior blastema, but the expression boundary later extends forward to form a posterior-to-anterior gradient. Taken together, early-injury induced *wnt1* from differentiated cells could induce its own expression in pole-forming neoblasts or neoblast progeny through *frz4-1-lrp5/6*. In the pole-forming cells, *wnt1* also induces the expression of the second rank module containing *wnt11-2* and *frz4-1*, which helps to stabilise the specification of the posterior pole. In addition, *wnt1* from posterior pole cells could induce the expression of *wnt11-2* and *frz4-1* in adjacent posterior cells that drive tail regeneration around the posterior pole cells. In turn, *wnt11-2* secreted from these adjacent cells would induce the expression of the third module involving *wnt11-1* that ensures the correct morphology of the tail, and subsequently *wntP-2* to ensure the regeneration of central regions.

Gene expression patterns along the AP axis and sequencing after knockdown of *beta-catenin-1* and *wnt1* have begun to identify the targetome of Wnt signalling (Owen et al., 2015; Reuter et al., 2015; Lander and Petersen, 2016; Scimone et al., 2016; Stückemann et al., 2017; Tewari et al., 2019; Pascual-Carreras et al., 2023). Together, these studies have observed a core downstream posterior specification program including the posterior Wnt components (*wnt11-1*, *wnt11-2*, *wntP-2*, *frz4-1*, *frz4-2*, *axinB*, *wntless*), posteriorly expressed *Hox* genes (*box4b*, *lox5a*, *post-2c*, *post-2d*), and the transcription factor *sp5* and the zinc finger *tsb*. Additionally, it appears that these core downstream effector are evolutionarily conserved, as they are downregulated after

knockdown of *beta-catenin-1* in acoels and zebrafish (Reuter et al., 2015; Tewari et al., 2019; Pascual-Carreras et al., 2023). In turn, the downstream effector *sp5* functions with *wntP-2* to position the trunk region along the AP axis (Tewari et al., 2019).

1.3.5.6 Trunk specification

In the trunk or the worm's central region, a regulatory module including the FGFR family member *ndl-3*, the Wnt ligand *wntP-2*, and the kinase *ptk7* have been shown to establish axial proportions, as knockdown of these genes results in worms with a second ectopic pharynx and mouth (Sureda-Gómez et al., 2015; Lander and Petersen, 2016; Scimone et al., 2016). Additionally, the loss of *wntP-2* expands the *ndl-3* expression domain towards the posterior (Lander and Petersen, 2016; Scimone et al., 2016), indicating cross-regulatory interactions. However, the direct and exact signalling mechanisms of the genes patterning the trunk are not known.

1.3.6 Planarian eye regeneration is self-organising

In addition to tissue loss and the regeneration cascade, the planarian eyes have provided further insight into the process of regeneration and pattern formation. The planarian eyes are discrete and easily accessible, making them a good experimental model to study planarian cell dynamics during regeneration. Two models for neoblast behaviour in regeneration have been proposed: targeted and non-targeted mode of regeneration. In the targeted mode of regeneration, neoblasts specifically give rise to progenitors of lost organs (Thi-Kim Vu et al., 2015; Bohr et al., 2021). In contrast, eyes regenerate in a non-targeted mode of regeneration, in which passive cell turnover replaces the cells over time (LoCascio et al., 2017). Neoblasts give rise to eye progenitors in a broad area extending from eyes to near the pharynx (LoCascio et al., 2017, Atabay et al., 2018). The specialised eye neoblasts are interspersed among other neoblasts. Upon eye resection, the number of eye progenitors is not increased, and continuous cell turnover regenerates the eyes passively over time (LoCascio et al., 2017; Reddien et al., 2021). Furthermore, it has been shown that both the eye-external positional information system and the self-organisation of eye progenitors contribute to eye regeneration (Atabay et al., 2018; Hill and Petersen, 2018). If the PCG expression domains are shifted following RNAi, the eye progenitors migrate to their correct position in the PCG expression domain if no eyes are present (external positional information). However, if eyes are present, the eye progenitors incorporate into the pre-existing eye regardless of PCG gene expression (self-organisation). Taken together, planarian eyes have demonstrated that, at least in some instances, planarian regeneration is not only dependent on the positional information system, but also the self-regulatory behaviour of planarian cells.

1.3.7 Summary of planarian regeneration

Whole-body regeneration in planarians relies on a molecular coordinate system of signalling pathways. This system is dynamically rescaled after injury. At wound sites, blastemas with signalling centres form that orchestrate downstream regeneration. The TALE factors *pbx* and *prep* are required for anterior pole specification, while *pbx* is additionally required for forming the posterior pole.

1.4 Aims of thesis

Before this thesis, little was known about gene regulatory networks in planarian whole-body regeneration. Therefore, the aim of this thesis is to understand regulatory mechanisms of the components of planarian regeneration: adult pluripotent stem cells and the dynamic cascade of mechanisms that pattern the regenerating planarian. The regulatory mechanisms of planarian stem cells are holistically studied as a system-wide gene regulatory network, while this thesis takes a reductionist approach to patterning mechanisms, focusing specifically on the family of TALE homeobox genes.

Chapter 1 provides a brief overview of planarians as a model system and their stem cells, along with a more comprehensive review on TALE factors and pattern formation in the context of a regenerating planarian.

Chapter 2 investigates gene regulatory networks in planarian stem cells. To this end, it provides a reannotation of the planarian genome, compiles a list of transcription factors, identifies enhancer-like regions active in stem cells, and integrates this information to construct gene regulatory networks modelling planarian stem cell dynamics. In this study, our research questions are:

- What are the transcription factors in the genome of *S. mediterranea*?
- Which enhancers are active in planarian stem cells?
- Which transcription factors are binding to the enhancers?
- What is the gene regulatory network of planarian stem cells?
- What does the gene regulatory network reveal about stem cell maintenance and lineage differentiation programs?

Our hypothesis was that enhancers are enriched in the vicinity of neoblast-specific genes, and a putative master regulator of pluripotency is enriched in these enhancers. However, our findings did not uncover a clearly enriched pluripotency factor in the network, although we found multiple enhancer-like regions in association with uncharacterised transcription factors, providing resources for further investigation into the mechanisms of stem cell maintenance and lineage differentiation in planarian stem cells.

Chapter 3 studies the phylogeny of planarian TALE homeobox genes, given that the transcription factor screening in the previous chapter had uncovered multiple uncharacterised TALE factors in the genome of *S. mediterranea*, and that TALE factors have diversified in spiralian. In this chapter, our research questions are:

- How have the TALE factors of *S. mediterranea* evolved?
- What are the homologues of TALE factors in the genome of *S. mediterranea*?
- Are there lineage-specific TALE factors in *S. mediterranea*?
- What are the gene expression dynamics of TALE factors in *S. mediterranea*?
- What are the functional roles of TALE factors in *S. mediterranea*?

Our hypothesis was that both conserved and lineage-specific TALE factors are present in the genome of *S. mediterranea*. Indeed, we found factors in conserved families (*irx-1*, *irx-2*, *irx-3*, *irx-4*, *prep*, *pbx-1*, *pbx-2*, *meis-1*, *meis-2*, *meis-like*) as well as factors in lineage-specific clades (*SPILE-Z1*, *SPILE-Z2*, *SPILE-E1*, *SPILE-E2*, *PALÉ*). Expression analysis and preliminary RNAi screening suggest roles in both early development and organogenesis.

In the fourth chapter, we asked whether other TALE factors are required for axial regeneration alongside the previously characterised *pbx-1* and *prep*. Furthermore, we sought to understand the targetome of TALE factors in axial regeneration and to infer TALE-independent regulatory interactions using chromatin accessibility in regenerating blastemas. Our research questions are:

- Are other TALE factors besides *pbx-1* and *prep* involved in axial regeneration?
- Which genes do *prep*, *pbx-1*, and *meis-2* regulate in axial regeneration?
- Which direct regulatory interactions between patterning genes are most likely, given chromatin accessibility data and previous knowledge?

We found that the Meis family member *meis-2* is involved in posterior regeneration. Integration of transcriptomic and accessibility data, along with some experimental evidence, suggests that *prep* and *pbx-1* are direct regulators of anterior pole markers and FGFR genes in the anterior blastema, while *pbx-1* and *meis-2* are direct regulators of *wnt1* and other genes of posterior specification regulatory cassette.

Chapter 2

Identification of putative enhancer-like elements predicts regulatory networks active in planarian adult stem cells

Jakke Neiro, Divya Sridhar, Anish Dattani, Aziz Aboobaker

This chapter has been published as a first author original research article in eLIFE (<https://doi.org/10.7554/eLife.79675>). Jakke Neiro and Aziz Aboobaker are corresponding authors. It was published online on the 23rd of August, and is reproduced here.

Divya Sridhar and Anish Dattani generated the H3K27ac ChIP-seq and ATAC-seq data. Dattani compiled the publicly available RNA-seq data used for the expression-driven reannotation of the planarian genome and provided preliminary guidance on the analysis. Jakke Neiro conducted the computational analysis for all sections, wrote the manuscript, and revised it following reviews.

Jakke Neiro

Contribution Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing

Divya Sridhar

Contribution Investigation, Methodology, Writing - review and editing

Anish Dattani

Contribution Formal analysis, Investigation, Visualization, Methodology, Writing - review and editing

Aziz Aboobaker

Contribution Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Funding

Medical Research Council (MR/T028165/1)

Biotechnology and Biological Sciences Research Council (BB/J014427/1)

Biotechnology and Biological Sciences Research Council (BB/M011224/1)

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Chapter 2. Identification of putative enhancer-like elements predicts regulatory networks active in planarian adult stem cells

Jakke Neiro, Divya Sridhar, Anish Dattani, Aziz Aboobaker

Planarians have become an established model system to study regeneration and stem cells, but the regulatory elements in the genome remain almost entirely undescribed. Here, by integrating epigenetic and expression data we use multiple sources of evidence to predict enhancer elements active in the adult stem cell populations that drive regeneration. We have used ChIP-seq data to identify genomic regions with histone modifications consistent with enhancer activity, and ATAC-seq data to identify accessible chromatin. Overlapping these signals allowed for the identification of a set of high-confidence candidate enhancers predicted to be active in planarian adult stem cells. These enhancers are enriched for predicted transcription factor (TF) binding sites for TFs and TF families expressed in planarian adult stem cells. Footprinting analyses provided further evidence that these potential TF binding sites are likely to be occupied in adult stem cells. We integrated these analyses to build testable hypotheses for the regulatory function of TFs in stem cells, both with respect to how pluripotency might be regulated, and to how lineage differentiation programs are controlled. We found that our predicted GRNs were independently supported by existing TF RNAi/RNA-seq datasets, providing further evidence that our work predicts active enhancers that regulate adult stem cells and regenerative mechanisms.

2.1 Introduction

The molecular and evolutionary mechanisms of regeneration remain underexplained compared to animal development. This can be attributed to the greater technical difficulty historically associated with studying the molecular mechanisms of adult biology compared with development. However, an ever-broadening repertoire of model organisms for regeneration, progress in understanding the variety of cellular and molecular mechanisms used across taxa, and advances in experimental tools are serving to close this gap. One area of slower progress has been our knowledge of the regulatory logic of regenerative mechanisms, with relatively few studies in highly regenerative models providing precise insight into the epigenetic regulation of regeneration. While transcription factors (TFs) have been assigned functions, and in some cases a list of likely targets through transcriptome analysis, regulatory logic has been studied in only a few cases (Pascual-Carreras et al., 2020; Wang et al., 2020). This is particularly true in the highly regenerative planarian model system, where only a few studies credibly address epigenetic and regulatory mechanisms (Duncan et al., 2015; Dattani et al., 2018; Mihaylova et al., 2018; Dattani et al., 2019; Pascual-Carreras et al., 2020).

Enhancers are distal-acting elements that regulate transcription initiation when TFs attach to TF binding sites, containing conserved sequences. In a regenerative context, active enhancers have been identified at the tissue level in zebrafish, the African killifish *Nothobranchius furzeri*, and the fruit fly *Drosophila melanogaster*, and at the whole-body level in the acoel worm *Hofstenia miamia* and the cnidarian *Hydra vulgaris* (Harris et al., 2016; Kang et al., 2016; Goldman et al., 2017; Gehrke et al., 2019; Murad et al., 2019; Yang and Kang, 2019; Wang et al., 2020). Despite their importance as a model for understanding regeneration, there are only a few putative enhancers identified or implicated in planarians (Pascual-Carreras et al., 2020), partly due to the lack of transgenic approaches allowing for direct functional testing of potential regulatory

elements. Hence, other analytical methods must be used to discover enhancer regions de novo across the genome. A set of putative planarian TFs have been identified based on available transcriptomic data, but genome-wide TF interaction with potential enhancers has not been elucidated (Swapna et al., 2018; Pascual-Carreras et al., 2020). Ultimately, TF interactions with enhancers can be used to construct gene regulatory networks (GRNs) (Duren et al., 2017; Lowe et al., 2019; Miraldi et al., 2019; Duren et al., 2020; Janssens et al., 2022).

In planarians, regeneration is based on a population of somatic pluripotent stem cells called neoblasts that can differentiate into all cell types of the adult body plan and are the only cycling cells in the adult worm (Aboobaker, 2011; Rink, 2013; Zhu and Pearson, 2016). Nonetheless, how neoblasts dynamically retain pluripotency, specify fate, and differentiate remains unknown from a regulatory network perspective, while a number of different models at the cell biological level have been hypothesized (Adler and Sánchez Alvarado, 2015; Swapna et al., 2018; Raz et al., 2021). Fluorescence-activated cell sorting (FACS) separates cells into three compartments: S/G2/M cell-cycle stage neoblasts (known as X1), G1 stage neoblasts and postmitotic progeny (known as X2), and differentiated cells (known as Xins, as the insensitive to radiation treatment) (Hayashi et al., 2006; Labbé et al., 2012; Onal et al., 2012; Romero et al., 2012; Solana et al., 2012). Analysis of the X1 cells has been used to catalog genes expressed in neoblasts, and these studies have revealed that neoblasts are heterogeneous and can be subdivided into multiple classes based on their gene expression profiles (van Wolfswinkel et al., 2014; Fincher et al., 2018; Plass et al., 2018). Some classes of fate-specified neoblasts are termed specialized neoblasts, which include precursors to eyes, protonephridia, epidermis, intestine, pharynx, neurons, and muscle (van Wolfswinkel et al., 2014; Raz et al., 2021). Fate-specific transcription factors (FSTFs) are expressed in S/G2/M phase and are thought to direct fate specification into different cell types (Raz et al., 2021). Evidence from current experimental data supports a model where specialized neoblasts can divide asymmetrically giving rise to one fate-specified, postmitotic daughter and a proliferative neoblast that may still specify different fates (Raz et al., 2021). Genome-wide identification of active enhancers in the X1 compartment would shed light on the GRNs regulating the dynamic behavior of neoblasts in planarian regeneration (Labbé et al., 2012; Onal et al., 2012; Solana et al., 2012). Currently, no such predictions of enhancers or GRNs of TFs exist.

Identifying enhancers is challenging and especially so in non-model organisms, but various genome-wide high-throughput sequencing techniques have revealed signatures indicative of enhancers (Shlyueva et al., 2014; Tomoyasu and Halfon, 2020). Chromatin accessibility has proven to be a universal attribute of active enhancers and other regulatory regions in eukaryotes (Thomas et al., 2011; West et al., 2014; Zhu et al., 2015; Daugherty et al., 2017; Klemm et al., 2019). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) has become the standard method of mapping open chromatin regions in various taxa (Buenrostro et al., 2013; Li et al., 2019; Yan et al., 2020). However, chromatin accessibility is not unique and specific to enhancers, and thus other complementary methods are used to discriminate enhancers from other open chromatin regions, particularly at the point of initial identification.

Enhancers are known to be flanked by nucleosomes with specific histone modifications (Shlyueva et al., 2014). In both mammals and *D. melanogaster*, acetylated lysine 27 of histone H3 (H3K27ac) is known to mark active enhancers together with mono-methylation of lysine 4 on histone H3 (H3K4me1), whereas H3K4me1 alone marks predetermined or poised enhancers (Heintzman et al., 2007; Creighton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011; Bonn et al., 2012; Arnold et al., 2013; Calo and Wysocka, 2013). However, poised enhancers are not always activated, most enhancers are activated without a prior poised state, and some poised enhancers are later actively repressed, indicating that the poised enhancer state is not necessarily indicative of a pre-activation state (Bonn et al., 2012; Rada-Iglesias et al., 2011; Koenecke et al., 2017). Nonetheless, the bivalent epigenetic signature of H3K27ac and H3K4me1 seems to be a conserved indicator of active enhancers in metazoans and has been successfully used for enhancer detection

in non-model organisms (Gaiti et al., 2017a, Jänes et al., 2018). In planarian neoblasts, the same histone modifications mark the active, suppressed, and bivalent state of promoters as in vertebrates (Dattani et al., 2018). Taken together, these data suggest that epigenetic marks may have a conserved association with regulatory elements across bilaterians (Schwaiger et al., 2014; Sebé-Pedrós et al., 2016; Gaiti et al., 2017b, Dattani et al., 2018).

Here, we significantly improve upon the annotation of the planarian genome (Grohme et al., 2018), take a computational approach to identify all TFs in this annotation, identify putative enhancers in the planarian genome supported by multiple lines of evidence, and then construct hypothetical GRNs active in neoblasts. We find that multiple enhancers have evidence of FSTF-mediated regulation, supporting the view that fate specification occurs in the S/G2/M phases. The FSTFs of different cell types seem to cross-regulate each other, revealing the potential for a dynamic GRN governing neoblast fate specification. We identified enhancers linked to several unstudied TFs implicating them in potentially central roles in neoblast TF GRNs. Enhancers linked to well-known planarian positional genes suggest regulatory mechanisms for some of the known links between these genes implicated by phenotypic studies. Finally, we show that our GRNs predict TF regulatory interactions implicated by available data measuring gene expression changes after TF RNAi. Overall, this work provides a foundation for future work on the regulatory logic of planarian stem cell biology and identifies many candidate TFs with predicted roles in regulating adult stem cells.

2.2 Results

2.2.1 Annotation with a full range of transcriptome samples identifies more than 3000 new protein coding genes in the *Schmidtea mediterranea* genome

We refined and extended the current annotation of the *S. mediterranea* genome (SMESG.1 genome and SMESG high-confidence annotation at Planmine; Brandl et al., 2016). We performed a genome-directed annotation based on the genome sequence and 183 independent RNA-seq datasets, including data both from whole worms and cell compartments (Figure 2.1A and B; this new annotation has also been used in García-Castro et al., 2021, but is described in more detail here). By including a diverse and large set of RNA-seq data, we sought to characterize transcripts undetected in individual studies and annotation attempts (Hoff and Stanke, 2013; Hoff and Stanke, 2019). Furthermore, we calculated proportional expression values for each cell population defined by FACS using approaches established previously (Figure 2.1A; Dattani et al., 2018).

In total, our expression-driven annotation process identified 91,068 transcripts at 28,097 genomic loci (Figure 2.1B). The annotation process validated all gene models in the SMESG high-confidence annotation, as all loci, transcripts, and exons were also found in our new annotation. In total, 50,213 transcripts were identified as putative new isoforms of previously identified loci (Figure 2.1C). Furthermore, 7412 new loci with 10,636 transcripts were found (Figure 2.1B, C). The protein coding potential of these new transcripts was assessed by defining putative open-reading frames (ORFs) and scanning for protein structures (Figure 2.1B). In total, 3121 new loci with 4752 transcripts were predicted to be coding, while 4291 loci with 5884 transcripts were predicted to be non-coding (Figure 2.1B, C).

The newly described isoforms across the genome were slightly shorter than known transcripts (median length 1618 bp vs. 1656 bp), while the new coding and non-coding transcripts were much shorter (median length 583 and 388 bp, respectively) (Figure 2.1D). However, the mean transcripts per kilobase million (TPM) value measured across the RNA-seq samples for new isoforms did not differ much from the levels of previously known transcripts, while the mean TPM values for new coding and non-coding transcripts were slightly higher (Figure 2.1, S2.A). This suggests that the main advantage of our approach was to discover shorter transcripts (and encoded proteins) not found by previous annotation approaches (Figure 2.1D, Figure 2.1, S2.1). We sorted all annotated genes to FACS enrichment groups (Figure 2.1H), using

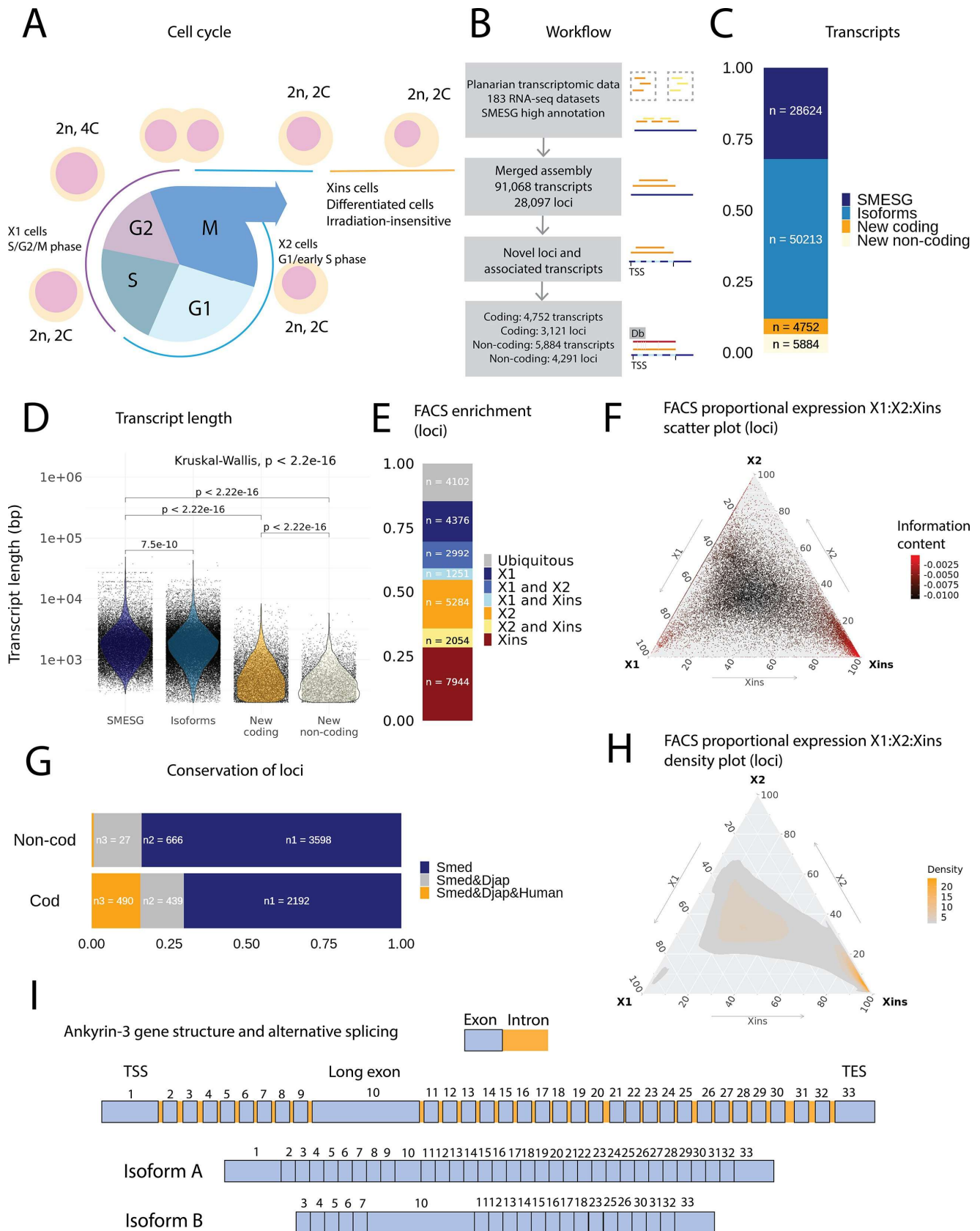


Figure 2.1 An expression-driven genome annotation of *Schmidtea mediterranea*.

(A) Diagram of the planarian neoblast cell cycle illustrating how the fluorescence-activated cell sorting (FACS) cell compartments X1, X2, and Xins relate to different cell-cycle phases. **(B)** Overview of methodology for annotating the *S. mediterranea* genome and identifying novel coding and non-coding transcripts. In total, 183 RNA-seq datasets were aligned to the SMESG.1 genome (Planmine). Novel transcripts were assembled by using the SMESG-high-confidence annotation as a reference (Planmine). **(C)** The proportion of transcript identities in the new annotation. ‘SMESG’ are transcripts present in the known SMESG high-confidence annotation, ‘Isoforms’ are transcripts that are isoforms or splicing variants of known transcripts in the SMESG high-confidence annotation, and ‘New coding’ and ‘New non-coding’ are previously unknown transcripts deemed to have coding and non-coding potential, respectively. **(D)** The transcript lengths of known transcripts (SMESG, $n = 28624$), isoforms ($n = 50213$), new coding transcripts ($n = 4752$), and new non-coding transcripts ($n=5884$). Kruskal-Wallis test is used to compare all groups and Wilcoxon rank-sum test is used for pairwise comparisons. **(E)** FACS categorization of 28,003 total loci into enrichment groups (Dattani et al., 2018). The enrichment groups are X1 (X1 proportional expression $\geq 50\%$), X2 (X2 proportional expression $\geq 50\%$), Xins (Xins proportional expression $\geq 50\%$), X1 and X2 (X1 + X2 proportional expression $\geq 75\%$, neither enriched in X1 nor X2), X1 and Xins (X1 + Xins proportional expression $\geq 75\%$, neither enriched in X1 nor Xins), X2 and Xins (X2 + Xins proportional expression $\geq 75\%$, neither enriched in X2 nor Xins), and ubiquitous (loci not categorized into the enrichment groups above and with roughly equal proportion in all three groups). **(F)** Ternary plot of proportional expression values in the X1, X2, and Xins cell compartments. Each dot represents one single locus. The information content metric represents the enrichment of a locus to one of the three compartments, meaning that gray loci are unspecific while red loci are specific to a single compartment (see ‘Materials and methods’ for details). **(G)** Conservation of new coding and non-coding loci to loci in *Dugesia japonica* and humans. ‘Smed’ (n_1) are loci exclusive to *S. mediterranea*, ‘Smed&Djap’ (n_2) are loci with a homolog in *D. japonica* but not in humans, and ‘Smed&Djap&Human’ (n_3) are loci with homologs both in *D. japonica* and humans. **(H)** Ternary plot of the density distribution of proportional expression values in the X1, X2, and Xins cell compartments. The plot gives a visual condensation of ternary plot F. The density distribution is estimated by two-dimensional kernel density estimation. **(I)** Schematic of the gene structure of ankyrin-3 and two example isoforms that are alternatively spliced (SMEST010564001.1.p1 and MSTRG.3975.46.p1). The detailed genomic track of all the isoforms is given in Figure 2.1, S2.1F. TSS, transcriptional start site; TES, transcript end site.

previously described methods (Dattani et al., 2018). The number of genes with enriched expression was highest in the Xins compartment and lowest in X1 cells (Figure 2.1E). As expected by overall lineage relationships, the shared enrichment between X1 and Xins is less common than between X1 and X2 and between X2 and Xins cells, congruent with the fact that X1 cells are all neoblasts, X2 is an amalgam of G1 neoblasts and differentiating postmitotic progeny, and Xins cells are their collective differentiation products (Figure 2.1E). By inspecting the distribution of proportional expression values, we also see that the distribution is shifted towards the Xins compartment (Figure 2.1F, H), thus overall expression of genes is enriched in differentiated cell types.

The majority of the new coding loci were unique to *S. mediterranea* (70 %), but 30% had a homolog in the *Dugesia japonica* genome (E-value cutoff 10^{-10} , Figure 2.1G; An et al., 2018). Only 27 non-coding loci (0.6 %) had a potential homolog in the human genome, but 666 loci (16 %) had a homolog in the *D. japonica* genome (Figure 2.1G). The Gene Ontology (GO) analysis of new coding loci revealed that the new genes were enriched for a function in RNA biosynthesis, metabolic processes, dephosphorylation, deacetylation,

transmembrane transport, and mitochondrial functions (Figure 2.1, S2.1C–E). These data suggest that many of our newly annotated genes are conserved amongst the group of planarians used for regeneration and stem cell research.

Some genes in the new annotation displayed high levels of alternative splicing, including homologs of lectin, ankyrins, and dystonin (Figure 2.1I, Figure 2.1, S2.1B, Table 2.1). In mammals, ankyrin-3 (ankyrin-G) is a structural protein localized to the axon initial segment (AIS) and the nodes of Ranvier, and alternative splicing is known to underlie its functional diversity (Figure 2.1I, Figure 2.1, S2.1F; Hopitzan et al., 2005; Lopez et al., 2017; Nelson and Jenkins, 2017). A planarian homolog of ankyrin-3 has 112 isoforms, including isoforms with one long exon, supporting the finding that giant ankyrin-based cytoskeleton of the AIS may have been present in the last common ancestor of bilaterians (Figure 2.1I, Figure S2.1F; Jegla et al., 2016).

Table 2.1 Genes with more than 100 transcript isoforms.

Number of isoforms	Gene	Description
571	TMEM25	Transmembrane protein 25
291	CLEC18B	C-type lectin domain family 18 member B
206	ANK2	Ankyrin 2
143	GABPB2	GA binding protein transcription factor subunit
138	PACRG	Parkin coregulated
112	ANK3	Ankyrin 3
106	DST	Dystonin
104	ANK3	Ankyrin 3

Taken together, these initial example analyses of our new annotation, particularly the discovery of many hundreds of new loci and thousands of putative alternative isoforms, suggest that it will have an important utility for the research community studying all aspects of planarian biology.

2.2.2 Comprehensive annotation of planarian transcription factors highlights a diversity of unknown zinc fingers

We screened for TFs in the new annotation using the same approach as Swapna et al., 2018 (Figure 2.2A). We validated the TF potential through a systematic protein domain annotation, assessed for homology to known TFs, and manually reviewed the list of TFs to assign them to planarian TFs present in the literature and databases (Figure 2.2A). Altogether, we found predicted 551 TFs in *S. mediterranea*, of which we found 248 to be described in the planarian literature as a named molecular sequence (i.e., the sequence was assigned a proper TF name; for details, see ‘Materials and methods’). The naming of planarian TFs in the literature was mostly consistent, but some inconsistencies were found (Supplementary file 1). We classified the TFs into four structural categories: basic domains, zinc domains, helix-turn-helix domains, and other domains (Stegmaier et al., 2004). Most basic domains had been described by Cowles et al., 2013, but we still identified new homologs of *Atf*, *Batf*, *Creb*, *Htf*, *Mad*, *Matf*, *MyoD*, *Npas*, and *Pdp* family members, each with broad established roles in metazoan biology (Figure 2.2B, Supplementary file 1).

In contrast, we identified multiple uncharacterized zinc finger domain TFs (ZNFs), many of which have not received much consideration yet in planarian regeneration research. While some of these unstudied

```

graph LR
    A[New annotation] --> B[982 TFs  
Pfam, SUPERFAMILY]
    B --> C[551 TFs  
Literature reference]
    C --> D[263 TFs  
166 motifs]
    subgraph Tools
        direction LR
        I[Interproscan]
        L[BLAST Literature review]
        J[Jaspar]
    end
    I --- B
    L --- C
    J --- D
  
```

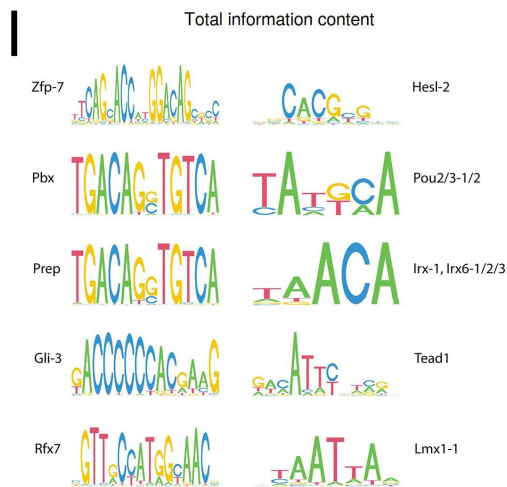
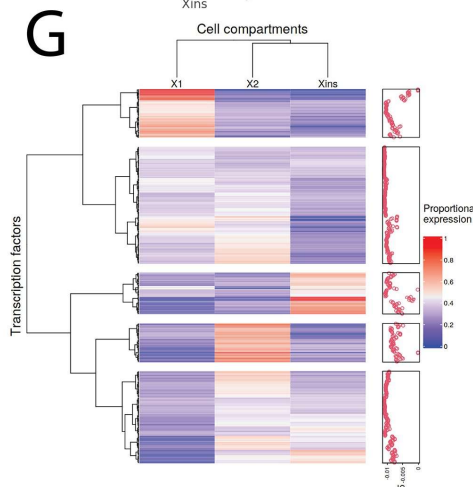
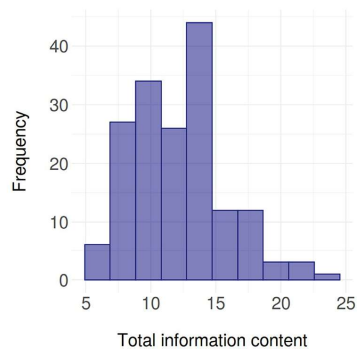
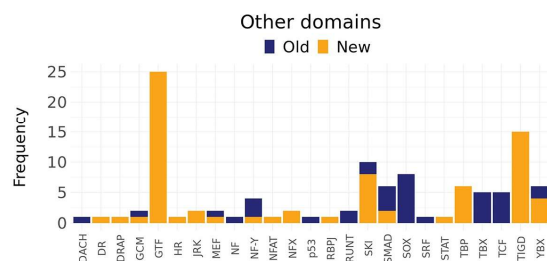
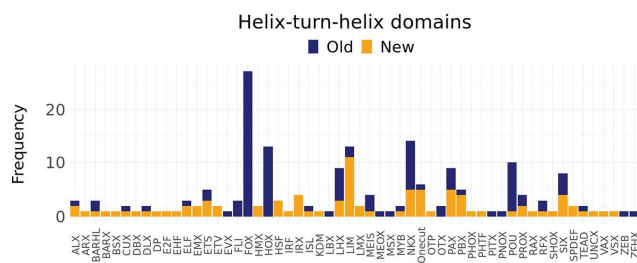
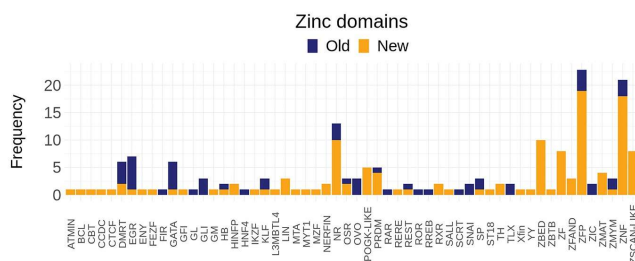
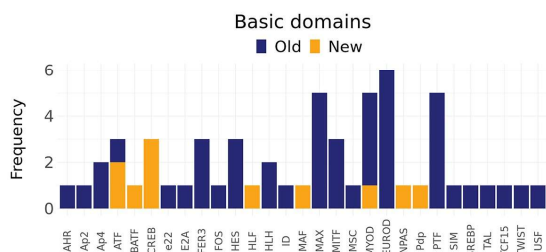


Figure 2.2 Annotation of planarian transcription factors.

(A) Diagram of the transcription factor (TF) annotation process. The coding transcripts identified in the new genome annotation were screened for conserved TF structures by using Interproscan and the Pfam and SUPERFAMILY databases. Homologs were then searched for the TFs with BLAST, and the planarian literature was reviewed to assign screened TFs to previously described. The manually reviewed TF proteins were then used to predict motifs. **(B–E)** Identified TF families with basic domains **(B)**, **(C–E)**. ‘Old’ refers to TFs present in the literature, while ‘New’ are TFs without a reference in the literature or new additional homologs. **(F)** Ternary plot of proportional expression values in the X1, X2, and Xins cell compartments. Each dot represents one single TF. The information content metric represents the enrichment of a locus to one of the three compartments (see 'Materials and methods' for details). **(G)** Heatmap and hierarchical clustering of proportional fluorescence-activated cell sorting (FACS) expression profiles for all TFs. **(H)** Histogram of the total information content of the identified motifs. **(I)** Example sequence logos for some of the best characterized motifs. The motifs of pbx (MA0782.1) and prep (MA0783.1) have almost identical PWMs, but they are listed as separate motif entries in JASPAR.

ZNFs could be assigned to well-known ZNF families such as GATA, KLF, EGR, and PRDM, many could not be assigned to well-described families. The nomenclature of ZNFs was based on the naming of human proteins to which they have the highest identity, and hence many appear in the ZNF and ZNP categories (Figure 2.2C). Interestingly, we also find ZNFs related to SCAN-domain containing zinc fingers (ZSCAN) and pogo transposable elements with KRAB domain (POGK) previously only described in vertebrates (Emerson and Thomas, 2011; Gao and Qian, 2020a). However, these planarian ZNFs do not contain the SCAN or KRAB domains, and the similarity arises from the DNA-binding domains. While members of the pogo transposable element superfamily are found throughout the metazoans, the KRAB subfamily is specific to vertebrates (Gao et al., 2020b). For this reason, we have provisionally named these ZNFs Zscan-like and POGK-like (Figure 2.2C).

For helix-turn-helix domains, new homologs were found for several described TF families, including *Nkx* and *Six*, and a few TFs belonging to families not previously described in planarians were also discovered, such as *Lmx* and *Sbox* (Figure 2.2D). Some TFs were also newly annotated for other domain families (Figure 2.2E).

We proceeded to allocate proportional expression values to the TFs with respect to the X1, X2, and Xins cell compartments (Dattani et al., 2018). The distribution of the proportional expression values was even, although a slight skew towards the X1 compartment was evident (Figure 2.2F, G). Most TFs were assigned to the X1 and Xins compartments and the least to the X2 compartment (Figure 2.2F, G). In other words, few TFs were specific to the X2 compartment, while most TFs were specifically expressed in stem cells and their nascent progeny (X1 and X2) or in progeny and differentiated cells (X2 and Xins). This would be expected by overall lineage relationships. There was no compartment-specific enrichment for different TF domains (χ^2 test, $p=0.2$).

We then moved to assigning predicted target binding motifs to the annotated set of planarian TFs. Most studies in non-model organisms have tended to use motifs directly from only a single-model organism or solely rely on de novo motif enrichment without reference to the TFs actually present in the studied organism (Gaiti et al., 2017a, Gehrke et al., 2019; Murad et al., 2019). However, we used the same approach as Siebert et al., 2019 and searched the JASPAR database for TFs with the highest similarity to predict motifs for planarian TFs (Figure 2.2A). In total, we found 166 motifs that were assigned to 263 TFs with a normal distribution of motif information value (Figure 2.2H, Supplementary file 1). The most informative

motifs were found for *zfp-7*, *pbx*, *prep*, *gli-3*, and *rfx-7*, while the least informative motifs were for *hesl-2*, *pon2/3*, *irx*, *tead1*, and *lmx1* (Figure 2.2I).

2.2.3 Histone modifications and chromosome accessibility mark enhancer-like regions

To identify potential enhancer regions, we analyzed previously generated ChIP-seq data with respect to the enhancer-associated histone modification H3K4me1 in X1 cells (Mihaylova et al., 2018). Furthermore, we sequenced the epigenome of X1 cells with respect to the histone modification H3K27ac to identify genomic regions indicative of an active enhancer state. H3K27ac was enriched at promoter regions, suggesting that this epigenomic feature for this histone modification is conserved in planarians (Figure 2.3, S2.2A, B; Gaiti et al., 2017b). We identified 37,345 H3K27ac peaks and 13,868 H3K4me1 peaks that were generally less than 200 bp wide (Figure 2.3A, B). At H3K4me1 peaks, the H3K27ac signal was strongest at the peak center at almost all peaks, while at H3K27ac peaks the H3K4me1 displayed a bimodal peak at the peak center of most peaks (Figure 2.3C, D). This pattern of H3K4me1 flanking H3K27ac peaks at enhancers has been previously described in mammals, and our data suggest that this may be evolutionarily conserved within metazoans (Gorkin et al., 2012; Spicuglia and Vanhille, 2012; Pundhir et al., 2016).

As our ChIP-seq data followed well-established enhancer-like patterns, we used the ChIP peaks to select putative enhancer-like regions. We calculated a mean peak value at all peaks with respect to the H3K27ac and H3K4me1 signal (Figure 2.3E) and selected all H3K27ac peaks that were at most 500 bp from H3K4me1 peaks and determined these 5529 peaks to be an initial set of putative active enhancer-like regions in cycling adult stem cells. The H3K27ac and H3K4me1 signals correlated at these enhancer-like regions more than at the other peaks in the genome (Figure 2.3F).

Furthermore, we calculated the mean peak value with respect to the change in H3K4me1 signal upon RNAi-mediated *Smed-lpt* knockdown (corresponding to the N-terminus of mammalian *ml3/4* or *kmt2c/d*, see Mihaylova et al., 2018). In mammals, *MLL3* and *MLL4* are two paralogous methyltransferases of the COMPASS family (SET1/MLL) that regulate enhancer activity by mono-methylating H3K4 (Wang et al., 2021). In addition, the *ml3/4* methyltransferase complex associates with the histone acetyltransferase *p300/CBP* that mediates H3K27 acetylation at enhancers and thus gives rise to the active enhancer landscape (Wang et al., 2021). The enhancer-like regions were clearly more responsive to the knockdown of *Smed-lpt* as compared to random points within the genome, and the response was most evident at the center of putative enhancer-like regions (Figure 2.3G, H and K). Thus, the H3K4me1 reduction after *Smed-lpt* knockdown provides further functional support for the identification of active enhancers in planarian stem cells as these are targets for active histone methylase activity associated with enhancers.

We optimized and performed ATAC-seq on X1 cells to measure high-resolution chromatin accessibility in conjunction with the histone modifications (Buenrostro et al., 2013) as a potential further source of evidence. Enhancer-like regions identified by ChIP-seq analysis had a more accessible chromatin configuration than random points in the genome, and the peaks of open regions were positioned at the center of the predicted enhancer-like regions implicated by ChIP-seq data (Figure 2.3I, J and L). Thus, ATAC-seq provides independent evidence for these regions being putative active enhancers in planarian stem cells.

2.2.4 Gene regulatory networks involving fate-specific transcription factors in neoblasts

Having defined enhancer-like regions (Supplementary file 2), integration with both TF and potential target gene expression data allowed us to begin constructing preliminary GRNs to demonstrate the utility of our neoblast enhancer predictions. First, we assigned enhancer-like regions to their closest gene, assuming a high probability that these will be the putative target genes of enhancer-like regions (Supplementary file 2).

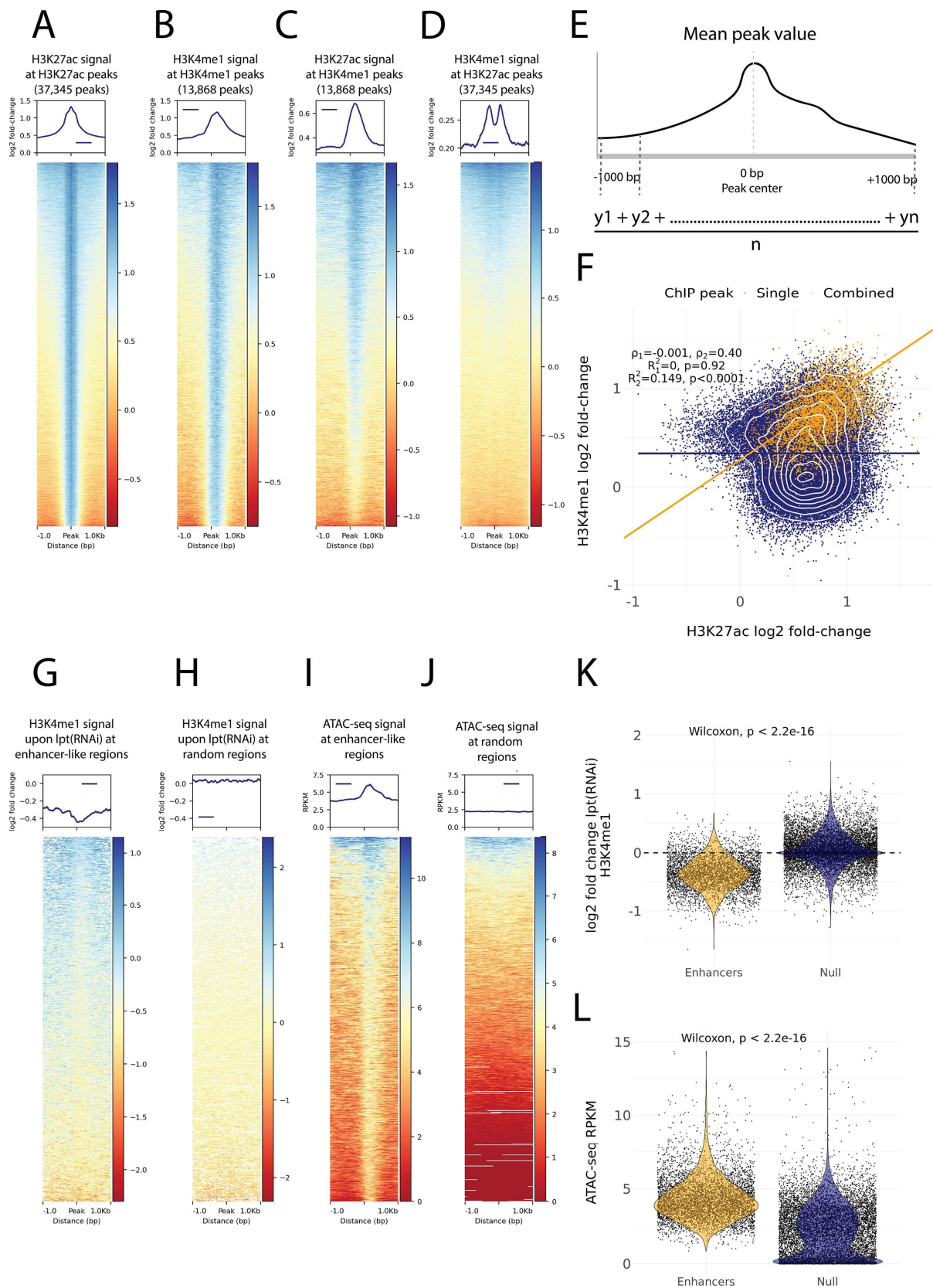


Figure. 2.3 Histone modifications and chromatin accessibility of enhancer-like regions.

(A) H3K27ac ChIP-seq signal around independently called H3K27ac peaks. **(B)** H3K4me1 ChIP-seq signal around independently called H3K4me1 peaks. **(C)** H3K27ac signal around the same H3K4me1 peaks as in (B). The signal centers around the center region. **(D)** The H3K4me1 signal around the same H3K27ac peaks as in (A). The signal displays a bimodal shape around the center region. The signal of the y-axis and the heatmap in (A–D) is given as the log₂ fold-change relative to the input and the peak width is 2000 bp. **(E)** A visual diagram of the calculation of the mean peak value. Mean peak value is defined as the mean of signal values taken at an interval of 10 bp from 1000 bp upstream to 1000 bp downstream of the peak center. **(F)** Scatter plot of H3K27ac and H3K4me1 log₂ fold-change mean peak values at the ChIP-seq peaks. H3K4me1 peaks that are at most 500 bp from a H3K27ac peak are defined as combined peaks and putative enhancer-like regions (model 2, orange), while the non-overlapping H3K4me1 and H3K27ac peaks are referred to as single peaks (model 1, navyblue). **(G)** The change in the H3K4me1 signal upon lpt (mll3/4) RNAi at the combined peaks or putative enhancer-like regions. The change is strongest at the peak center. The log₂ fold-change is calculated as the RNAi value relative to the wildtype value. **(H)** The H3K4me1 lpt(RNAi) signal at 10,000 random regions in the genome. **(I)** The ATAC-seq signal at the combined peaks or putative enhancer-like regions. The signal is given as reads per kilo base per million mapped reads (RPKM) **(J)** The ATAC-seq signal at 10,000 random regions in the genome. **(K)** Violin plot of H3K4me1 lpt(RNAi) signal at putative enhancer-like regions (n = 5529) and random regions (n = 10,000) in the genome. The H3K4me1 signal is reduced at enhancers. Wilcoxon rank-sum test is used for pairwise comparisons. **(L)** Violin plot of the ATAC-seq signal at putative enhancer-like regions (n = 5529) and random regions (n = 10,000) in the genome. The ATAC-seq signal is higher at enhancers. Wilcoxon rank-sum test is used for pairwise comparisons.

The distance from the enhancer-like regions to the target gene transcription start site (TSS) varied from being proximal to the promoter region to being as far as 89 kb away (Figure 2.4A). We found that TFs themselves were significantly enriched within the set of all predicted target genes (Figure 2.4B). Furthermore, these TFs were in turn enriched for the 43 FSTFs previously shown to be expressed in S/G2/M neoblasts (Figure 2.4B; Raz et al., 2021), again suggesting that our predicted set of enhancers are real. In addition to TF activity, the predicted target genes linked to enhancer-like regions were enriched for RNA metabolic and RNA biosynthetic processes and other biosynthetic processes, as well as transcription and regulation of gene expression (Figure 2.4C). These data suggest that our enhancer predictions include real enhancers involved in regulating key aspects of neoblast biology.

To establish more persuasive and direct regulatory links between TFs and enhancer-like regions, we used ATAC-seq footprinting to detect potentially bound motifs in the planarian stem cell genome (Bentsen et al., 2020; Figure 2.4D, Supplementary file 3). ATAC-seq footprints are short inaccessible or less accessible regions within an otherwise accessible region, indicative of DNA binding by a TF or another DNA-binding protein (Bentsen et al., 2020; Figure 2.4D). As for the raw ATAC-seq signal, footprint scores were higher in enhancer-like regions than in random regions of the genome (Figure 2.3, S2.2C, D). Overall footprinting analysis found 22,489 putatively bound TF motifs in the enhancer-like regions but no bound motifs in random regions, providing further support that these regions are potential enhancers. The TFs with the most binding sites predicted by footprint analysis in the genome overall were *irx-1/irx6-1/2/3*, *egr4-1/2*, *smad1/9*, *sp5*, and *nf-ya1/2*, while the TFs with the least bound motifs were *mef2-1*, *mef2-2*, *phox*, *pou6-2*, and *lmx1-2* (Figure 2.4E Supplementary file 3 and Supplementary file 4). Overall, these data provide a resource of studying GRNs and specific regulatory interactions in planarians neoblasts. We found no obvious

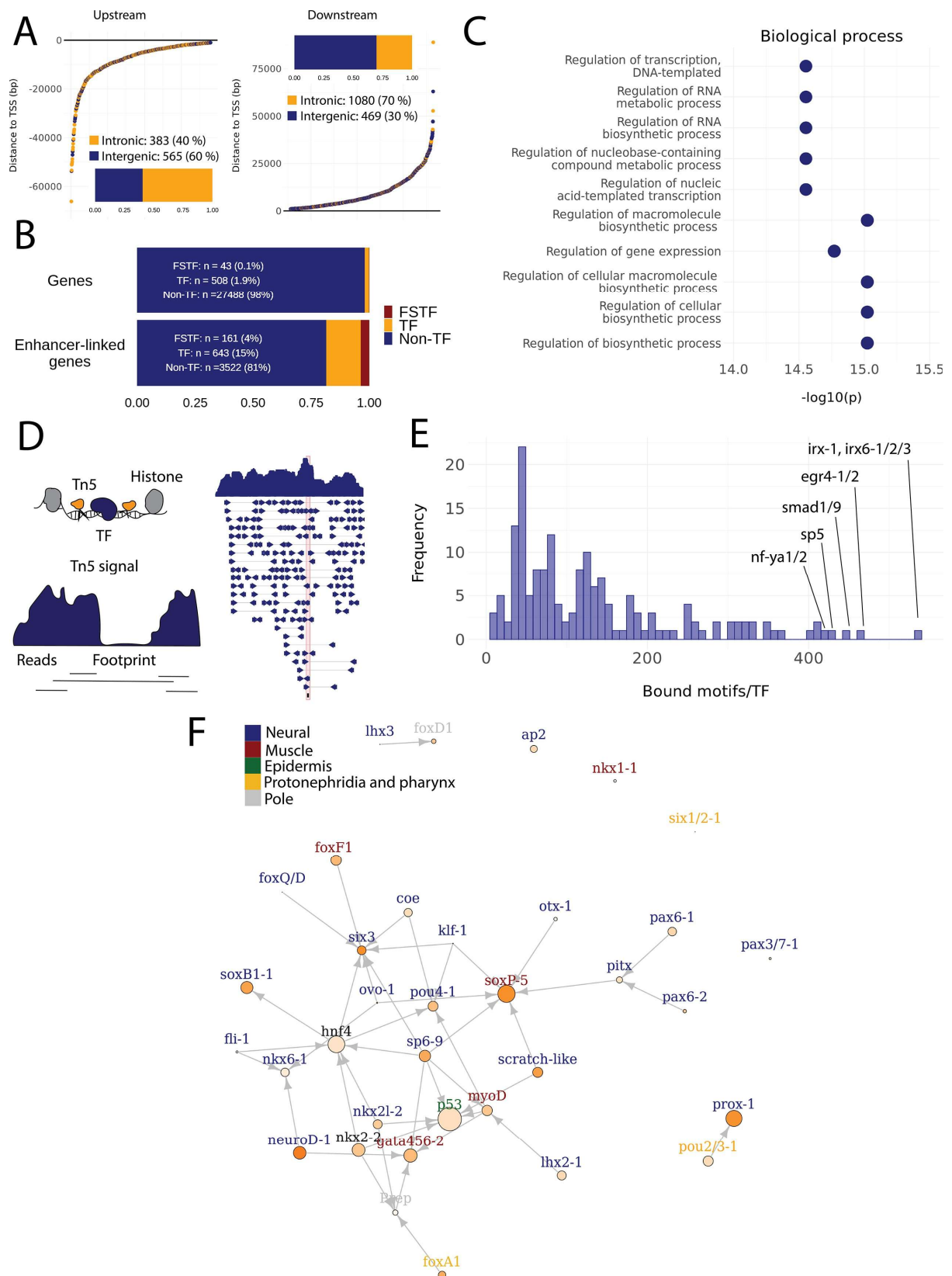


Figure 2.4 Predicted gene regulatory networks involving fate-specific transcription factors in neoblasts.

(A) Distance of enhancer-like regions to the transcription start site (TSS) of nearest genes. The nearest genes are assumed to be the putative target genes of the enhancer-like regions. **(B)** The proportion of fate-specific transcription factors (FSTFs), transcription factors (TFs), and non-TF genes of all planarian genes and of the target genes linked to enhancer-like regions. FSTFs and TFs are enriched in the set of target genes ($\chi^2 = 2660$, $p < 0.001$). **(C)** Gene Ontology (GO) analysis of biological processes of enhancer-linked target genes compared to all genes in the planarian genome. **(D)** Schematic overview of ATAC-seq footprinting (Bentsen et al., 2020). Both the fragmenting enzyme Tn5 and TF bind to accessible regions of the genome. Footprints are short and sharp inaccessible segments in otherwise accessible regions. The genomic track figure represents an example of ATAC-seq reads and footprint score overlapping with an sp5 motif in the genome. **(E)** Histogram of the total number of bound motifs in the genome for each TF. The TFs with most bound motifs in the genome are marked. **(F)** Putative gene regulatory network (GRN) of FSTFs in neoblasts. The epidermal FSTFs are labeled in green, the pole-associated FSTFs are labeled in gray, the muscle FSTFs are labeled in red, the intestinal FSTFs are labeled in black, and the neural FSTFs are labeled in blue. The size of the nodes reflects the absolute expression in X1 cells (transcripts per million reads [TPM]), and the color of the node reflects the proportional expression in X1 cells (the more orange, the higher proportional expression). The arrows represent a regulatory link but do not discriminate between positive and negative interactions.

difference in the distance of predicted TF footprints from TSSs for different TFs (Figure 2.4, S2.3).

To test the utility of these data for building and testing GRNs, we focused on the previously defined FSTFs, established the regulatory links between them, and constructed a putative stem cell GRN based on our datasets (Figure 2.4F, Supplementary file 5). In total, we could include 35/43 FSTFs into a GRN prediction and found evidence for multiple cross-regulatory links that may serve to allow stem cells to decide between fates. For example, neural FSTFs regulate enhancers of muscle FSTFs and vice versa (Figure 2.4F). The interactions in this GRN allow specific hypotheses to be formed and tested in the future. As examples, we present in detail the genomic region of the epidermal FSTF *p53*, the muscle FSTF *myoD*, and the neural FSTFs *nkx6-1* and *soxB1-1* (Figure 2.5A–D). The epidermal FSTF *p53* has multiple TFs predicted to bind at the promoter region, including a muscle FSTF (*myoD*) (Scimone et al., 2014; Scimone et al., 2017), and TFs relating to position (*zic-1*) (Vásquez-Doorman and Petersen, 2014) and cell migration (*zeb-1* and *snail-1/2*) (Abnave et al., 2017; Figure 2.5A). An enhancer-like region within an intron of the upstream gene *ythdc23* was predicted to be associated with the downstream muscle FSTF *myoD* as the TSS of *myoD* is closer of the two and *ythdc23* is lowly expressed in X1 and X2 cells. The enhancer-like region includes bound TF motifs such as the neoblast-enriched zinc finger *fir-1*, the muscle segment homeobox gene *msx*, the neural *pou-p1*, *sp6-9*, *lhx2-1*, and position-related *foxA(P)* (Figure 2.5B). A distal enhancer-like region was found downstream of *nkx6-1*, including potentially bound TF motifs such as the neoblast-enriched *egr-1*, the neural *ovo-1* and *fli-1*, the muscle-related *twist*, the position-related *smad4-1*, and the pigmentation-related *ets-1* (Figure 2.5C). A distal enhancer-like region was found upstream of *soxB1-1*, including bound TF motifs such as the position-related *smad1* and *foxA(P)* (Molina et al., 2007; Reddien et al., 2007; Pascual-Carreras et al., 2020), neural *da* (Cowles et al., 2013), intestinal *hmf4* (Wagner et al., 2011; Scimone et al., 2014; van Wolfswinkel et al., 2014), and *zeb-1* and *snail-1/2* relating to cell migration (Abnave et al., 2017; Figure 2.5D). In the future, these putative regulatory links can be verified and studied by experiments using functional

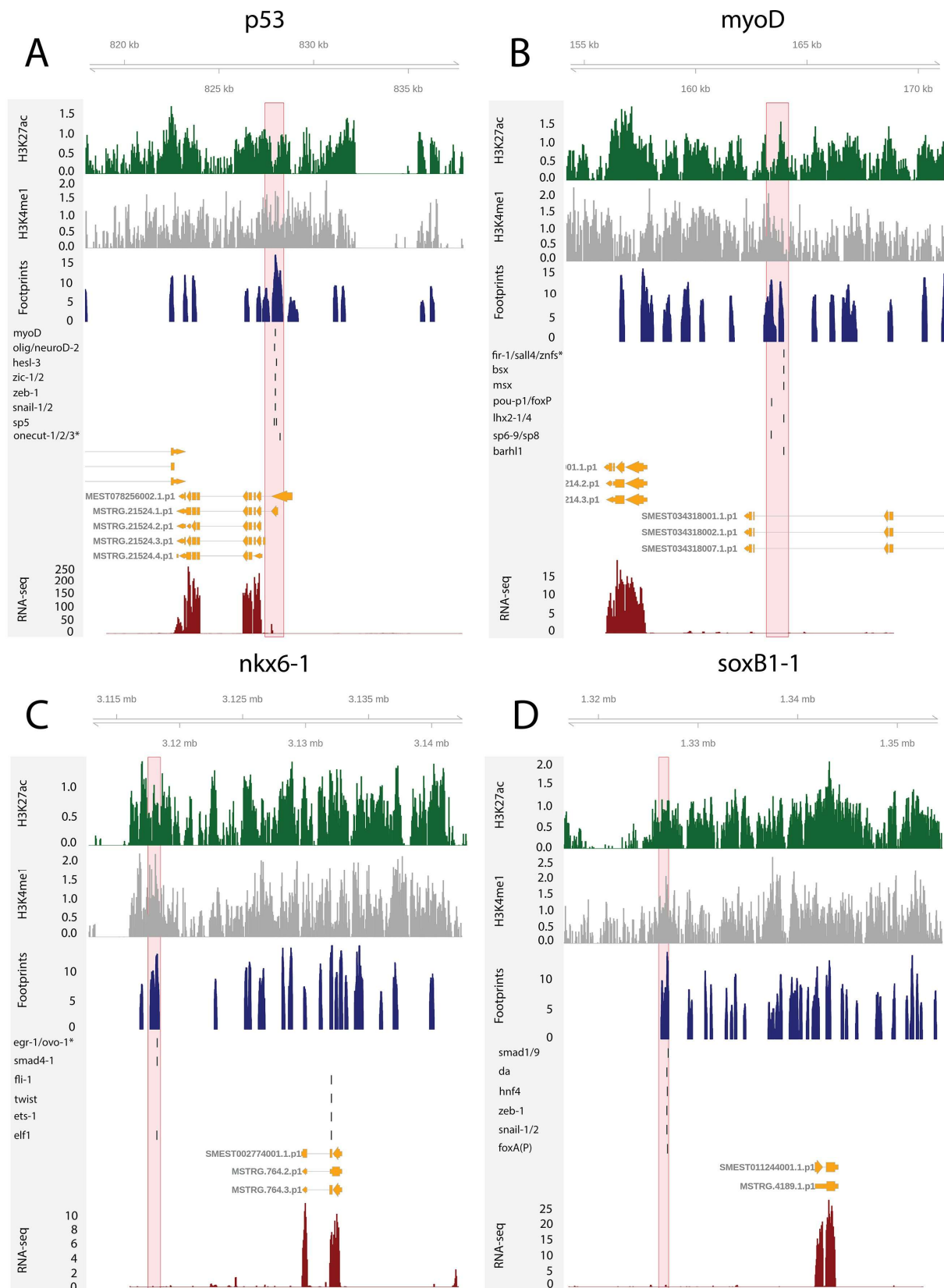


Figure 2.5. The genomic tracks of four fate-specific transcription factors (FSTFs) (A–D).

The first track is the H3K27ac signal (log2 scale compared to the input sample), the second track is H3K4me1 signal (log2 scale compared to the input sample), the third track is the ATAC-seq footprint score (TOBIAS footprint score), the fourth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the fifth track represents the transcripts of the gene in the new annotation, and the sixth track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase of transcript per million [FPKM] mapped reads). **(A)** Genomic track of the epidermal FSTF *p53*. The promoter region marked in red (width 1000 bp) has multiple bound motifs. **(B)** Genomic track of the muscle FSTF *myoD*. An enhancer-like region marked in red (width 1000 bp) is located in the intron of the gene YTH domain containing 2 (*ythdc23*). The enhancer-like region is closer to the *myoD* promoter than the *ythdc23* promoter and *ythdc23* is not expressed almost at all in neoblasts, and hence the enhancer-like region is likely linked to *myoD* rather than *ythdc23*. **(C)** The neural FSTF *nkx6-1* has a distal enhancer-like region marked in red (width 1000 bp) downstream of the gene body. **(D)** The neural FSTF *soxB1-1* has a distal enhancer-like region marked in red (width 1000 bp) upstream of the gene body.

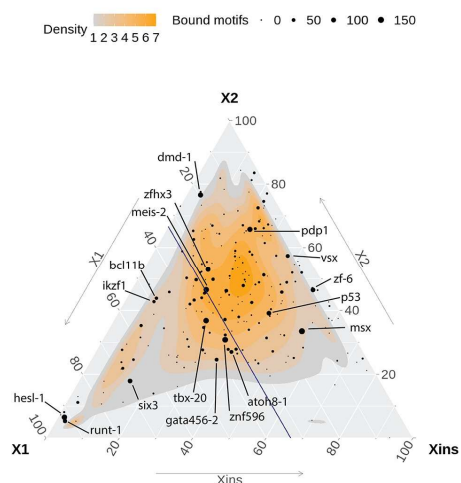
genomics involving the RNAi knockdown of individual genes and subsequent RNA-seq and ATAC-seq analysis.

2.2.5 Multiple enhancers are linked to unstudied transcription factors

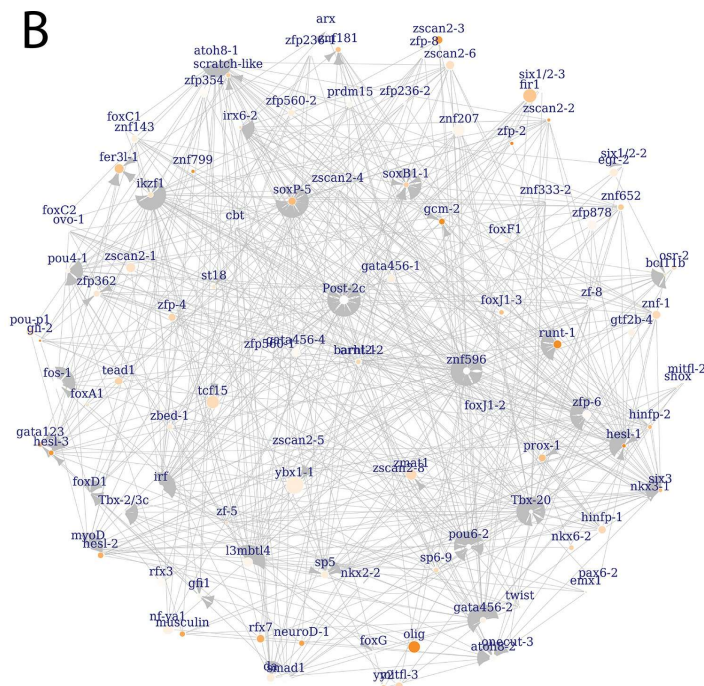
In addition to the relatively well-known planarian FSTFs (Raz et al., 2021), we investigated GRNs relating to less-well-studied TFs with multiple bound motifs in associated enhancer-like regions as these may also have a potentially central role in the neoblast regulation. We defined the number of enhancer-like regions, number of putative bound motifs, and number of putative unbound motifs for each TF (Figure 2.6A, Supplementary file 6). In addition, we selected TFs expressed in the X1 compartment (proportional expression of X1 > 1/3) and constructed the putative GRN summarizing all the 502 regulatory links of these 109 TFs (Figure 2.6B). In the X1 compartment (X1% > 1/3), the least-well-studied TFs with the most bound motifs in enhancer-like regions were *znf596*, *tbx-20*, *hesl-1*, *atoh8-1*, and *ikzf1* (Figure 2.6A, B, Supplementary file 2 and Supplementary file 6). Outside the X1 compartment, *msx*, *zfp-6*, *vsx*, and *pdp-1* had the most bound motifs (Figure 2.6A). The TF with most regulatory interactions, *znf596*, has been characterized to be expressed in neoblasts and more specifically in the neoblasts committed to the neural fate, but otherwise its function is unknown (Figure 2.6A–C, Supplementary file 6; Fincher et al., 2018). We found that numerous FSTFs were predicted to bind to putative intronic enhancers within *znf596*, including neural FSTFs *sp6-9*, *neuroD-1*, and *ovo-1*, the intestinal FSTF *gata456-1* and/or the muscle FSTF *gata456-2*, and the position control genes (PCGs) *smad1/9* (BMP signaling), *isl-1*, and *prep* (Figure 2.6C). In single-cell transcriptomics data, the gene is clearly expressed in a subset of neoblasts, calling for further mechanistic studies of its regulatory function. The binding motif of *znf596* could not be predicted, so its target genes cannot be implicated within the current datasets and approaches.

We identified the same five planarian *tbx* genes that have been previously reported, namely, *tbx-1/10*, *tbx-2/3a*, *tbx-2/3b*, *tbx-2/3c*, and *tbx-20* (Tewari et al., 2019). In mammals, the *tbx* genes have multiple roles in development, and *tbx3* is involved in regulating pluripotency of embryonic stem cells (Baldini et al., 2017; Khan et al., 2020). Nonetheless, the function of planarian *tbx* genes has not been clarified. Here, we found several predicted bound motifs of *tbx* genes in the genome (*tbx-1/10* 91 motifs, *tbx-2/3a/c* 81 motifs, *tbx-20* 62 motifs) and multiple bound motifs at enhancer-like regions linked to *tbx* genes (*tbx-20* 142 motifs, *tbx-2/3c* 34 motifs, *tbx-2/3a* 11 motifs) (Figure 2.6A, B and D, Supplementary file 2 and). *tbx-20* had a clear

A

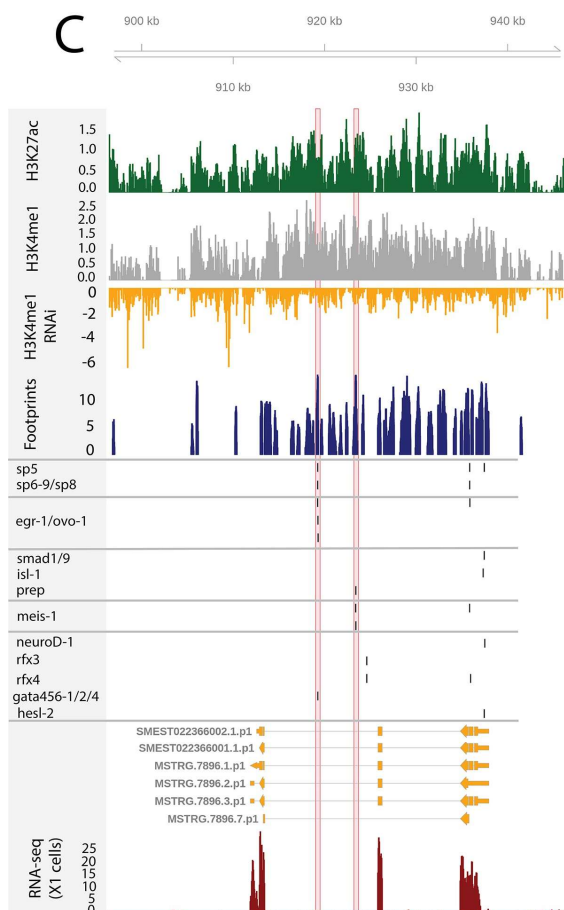


B



znf596

C



tbx-20

D

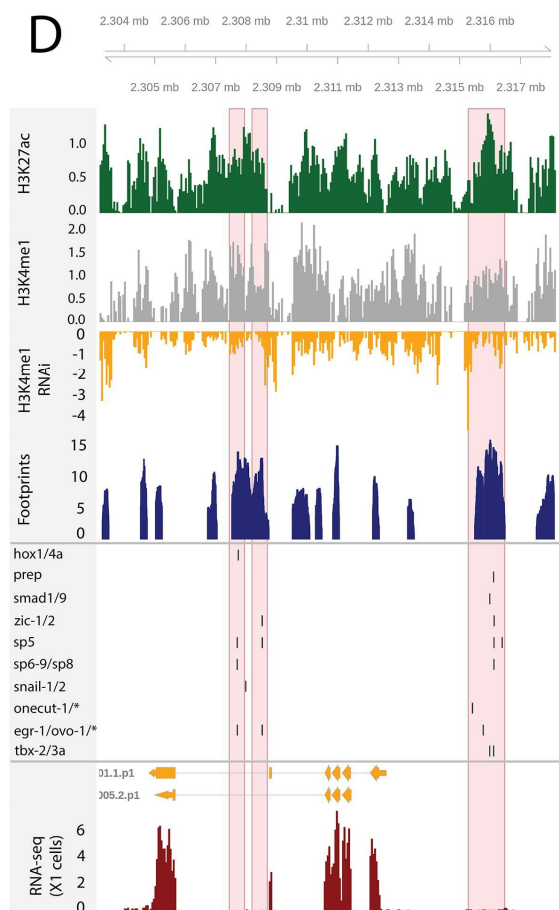


Figure 2.6 Enhancer-like regulatory regions in relation to unstudied planarian transcription factors.

(A) Ternary plot of the proportional expression of all transcription factors (TFs) in the cell compartments (X1, X2, and Xins), and the number of bound motifs linked to the TFs. The size of the dot represents the number of bound motifs. The TFs with most bound motifs in enhancer-like regions are named and marked. **(B)** The putative gene regulatory network (GRN) of all TFs in the X1 compartment. The GRN includes all TFs with an X1 proportional expression higher than 1/3. The size of the nodes reflects the absolute expression in X1 cells (transcripts per million reads [TPM]), and the color of the node reflects the proportional expression in X1 cells (the more orange, the higher proportional expression). The genomic tracks of *zmf596* (C) and *tbx-20* (D). The first track is the H3K27ac signal (log2 scale compared to the input sample), the second track is H3K4me1 signal (log2 scale compared to the input sample), the third track is the log2 fold-change upon *lpt(RNAi)*, the fourth track is ATAC-seq footprint score (TOBIAS footprint score), the fifth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the sixth track represents the transcripts of the gene in the new annotation, and the seventh track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase of transcript per million [FPKM] mapped reads). **(C)** Genomic track of *zmf596*. Two intronic enhancer-like regions are marked in red (width 500 bp). **(D)** Genomic track of *tbx-20*. Two intronic (width 500 bp) and one distal enhancer-like region (width 1200 bp) are marked in red.

intronic enhancer-like region and a distal enhancer-like region, both containing motifs for TFs implicated as positional control genes (PCGs) (*box1*, *box4a*, *prep*, *smad1*, *smad9*, *zic-1*, *sp5*) (Molina et al., 2007; Reddien et al., 2007; Felix and Aboobaker, 2010; Vásquez-Doorman and Petersen, 2014; Tewari et al., 2019), TFs involved in neoblast migration (*snail-1* and *snail-2*) (Abnave et al., 2017), and various other TFs (Figure 2.6D).

2.2.6 Elements of the planarian positional gene regulatory network are active in neoblasts

Next, we studied enhancer-like regions and potential regulatory links associated with the well-known planarian PCGs. Planarians are a primary model system to understand how positional information guides and directs stem cell function during regeneration (Reddien, 2018), and therefore some functional genomics data exists with regard to PCGs, enabling the comparison with our findings (Tewari et al., 2019). A constitutive positional information system is established by the regional expression pattern of PCGs that pattern the anterior-posterior (AP), dorsal-ventral (DV), and medial-lateral (ML) axes (Reddien, 2018).

The AP axis is patterned by the Wnt signaling pathway: high Wnt activity specifies posterior identity, while low activity specifies anterior identity (Reddien, 2018). Upon β -catenin knockdown, posteriorly expressed *sp5* and the Hox genes *post-2c*, *post-2d*, *lox5a*, and *box4b* are rapidly downregulated (Tewari et al., 2019). In multiple vertebrate species, *sp5* is known to be a direct target of Wnt signaling (Weidinger et al., 2005; Fujimura et al., 2007), and this regulatory link seems to be conserved in planarians (Tewari et al., 2019). Here, we found a putative bound *sp5* footprint in the vicinity of the *post-2c* and *lox5a* promoters, suggesting and further supporting that *sp5* regulates the expression of these posterior PCGs (Figure 2.7A). *sp5* had 431 putative bound motifs and was the fourth most bound motif in X1 cells overall, further suggesting that *sp5* mediates the broad positional information provided by Wnt signaling in planarians (Supplementary file 2).

Through ATAC-seq and ChIPmentation techniques, Pascual-Carreras et al., 2020 screened for cis-regulatory elements in planarian tissues in notum and *wnt1* (RNAi) animals. Upon *wnt1* knockdown,

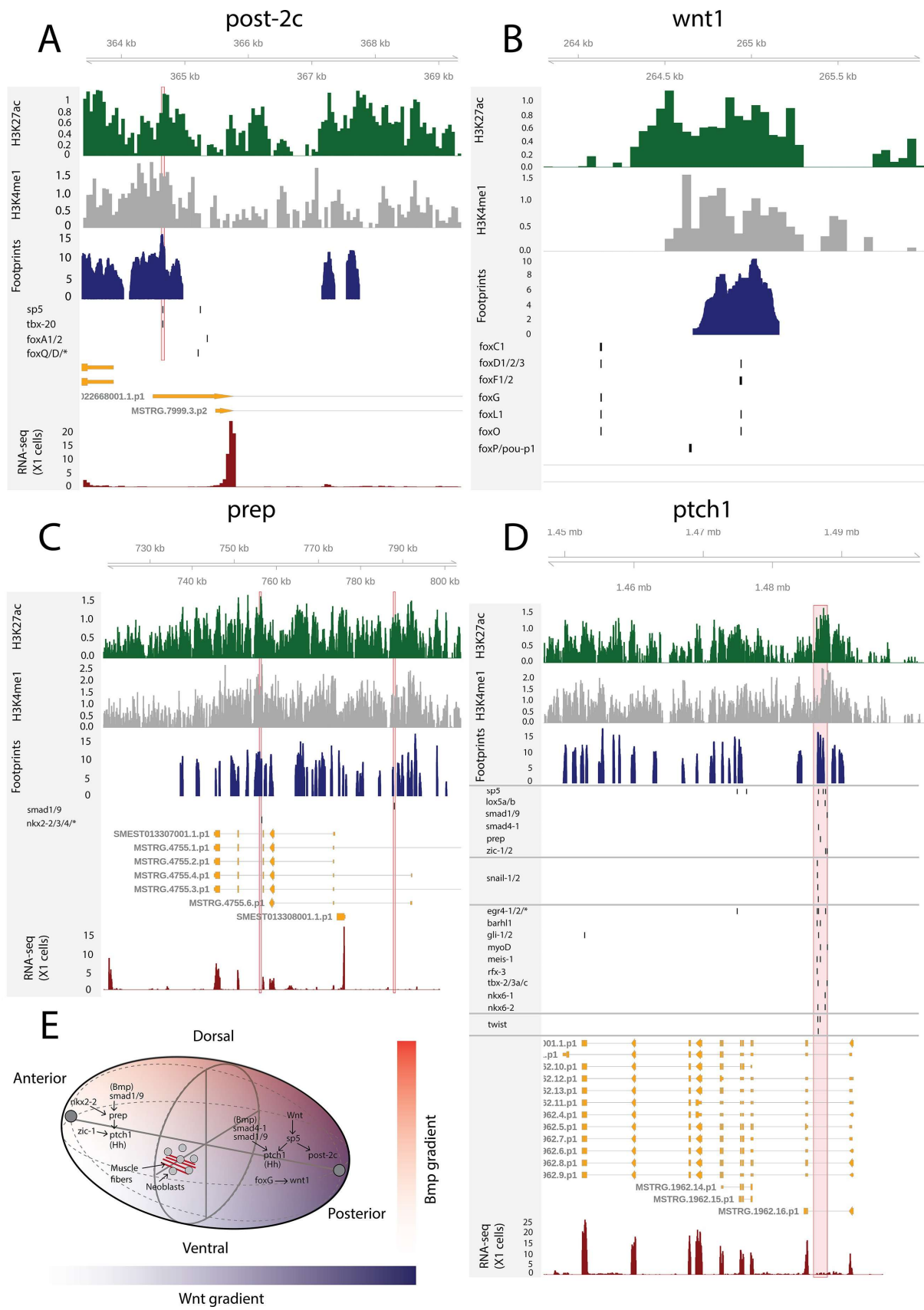


Figure 2.7 The genomic tracks of *post-2c* (A), *wnt1* (B), *prep* (C), and *ptch1* (D).

The first track is the H3K27ac signal (log2 scale compared to the input sample), the second track is H3K4me1 signal (log2 scale compared to the input sample), the third track is ATAC-seq footprint score (TOBIAS footprint score), the fourth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the fifth track represents the transcripts of the gene in the new annotation, and the sixth track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase of transcript per million [FPKM] mapped reads). **(A)** Genomic track of *post-2c*. One upstream cis-regulatory regions is marked in red (width 50 bp). The foxQ/D* footprint represents motif MA0851.1 corresponding to TFs foxQ/D, foxJ1-1, foxJ1-2, foxJ1-3, foxJ1-4, foxJ1-5, foxN2/3-1, foxN2/3-2, and foxN2/3-3. **(B)** Genomic track of *wnt1*. One intronic enhancer-like region with one potential Fox family TF binding motif with evidence of binding from footprinting analysis, other Fox TF motifs lie outside the enhancer predicted by ChIP-seq data (Pascual-Carreras et al., 2020). **(C)** Genomic track of *prep*. Two intronic enhancer-like regions are marked in red (width 500 bp). **(D)** Genomic track of *ptch1*. One intronic enhancer-like region (width 2000 bp) is marked in red. **(E)** Schematic of position control gene (PCG) regulatory links discussed.

posterior Hox genes *box4b*, *post-2c*, *post-2b*, *lox5a*, *lox5b* and *wnt11-1*, *wnt11-2*, *foxd4*, and *sp5* were downregulated, replicating the results of β -catenin knockdown (Pascual-Carreras et al., 2020). In addition, two foxG binding sites were found in the first intron of *wnt1*, and foxG knockdown was found to phenocopy *wnt1* knockdown, supporting the hypothesis that foxG is an upstream regulator of *wnt1* (Pascual-Carreras et al., 2020). Here, we found one enhancer-like region in the first intron of *wnt1* with a high level of H3K27ac, H3K4me1, and ATAC-seq footprinting scores for Fox family TFs (Figure 2.7B). This motif implicated by footprinting analysis is the same as one of the motifs described by previous work (Pascual-Carreras et al., 2020). We did not see evidence of binding at the second motif, and neither was this motif within one of our predicted enhancers (Figure 2.7B).

Although the planarian Hox genes are expressed in a regionalized manner along the AP axis, the knockdown of the genes apart from *post-2d* does not result in homeostatic or regeneration-associated phenotypic changes (Currie et al., 2016; Arnold et al., 2021). Instead, the five Hox genes *box1*, *box3a*, *box3b*, *lox5b*, and *post2b* have been shown to be involved in asexual reproduction by regulating fission at potential cryptic segment boundaries (Arnold et al., 2021). Here, we did not find any predicted bound motifs implicated by footprinting associated with *box1*, *box3a*, and *box3b* (Supplementary file 2), and only a few at the promoters of *lox5b* and one at *post2b* (Supplementary file 2, Figure 2.7, S2.4A and B). The proportional expression values of these Hox genes in X1 cells are 4% (*hox1*), 3% (*hox3a*), 63% (*hox3b*), 8% (*lox5b*), and 31% (*post2b*). Altogether, this suggests that these Hox genes are not driving fission through their regulatory activity in adult stem cells.

Prep, *zic-1*, *isl-1*, and *foxD1* are PCGs that are expressed in the anterior pole of both intact and regenerating planarians (Felix and Aboobaker, 2010; Vásquez-Doorman and Petersen, 2014; Vogg et al., 2014). Interestingly, we found that these TFs are bound to several motifs in enhancer-like regions of X1 cells (Supplementary file 1 enhancers, *prep* 132 motifs, *zic-1* 2 330 motifs, *isl-1* 42 motifs, *foxD1* 2/3 42 motifs) and a few motifs were found to bind to the enhancer-like regions linked to *prep*, *zic-1*, *isl-1*, and *foxD1* (Figure 2.7C, Supplementary file 2, Supplementary file 5, and Supplementary file 6). *Prep* has enhancer-like regions that have bound motifs of smad factors (Bmp signaling components) and *nkx2-2/3/4* (Figure 2.7C).

Knockdown of *nkx2-2* (also known as DTH-1 or *nkx2.2*) causes blastemal defects both at the anterior and posterior ends, while knockdown of *prep* leads to defects in the anterior, suggesting that *nkx2-2* might work upstream of *prep* at the anterior end (Felix and Aboobaker, 2010; Forsthoefel et al., 2020).

Lastly, we studied the Hedgehog signaling receptor *patched-1* (*ptch-1*) and found an intronic enhancer-like region containing multiple bound motifs, including the Hedgehog pathway TFs *gli-1/2* as would be expected, for the Wnt effector *sp5*, the Hox genes *lox5a/b*, BMP signaling components (*smad4-1*, *smad1/9*), the anterior pole TFs *prep* and *zic-1*, the cell migration factors *snail-1/2*, and the muscle TFs *myoD* and *twist* (Figure 2.7D). Hedgehog signaling in planarians is known to have pleiotropic functions, including interaction with the Wnt signaling pathway and glial cell function (Rink et al., 2009, Yazawa et al., 2009, Wang et al., 2016). Taken together, we found that elements of positional control gene regulatory network are active in planarian S/G2/M neoblasts and an integral part of the overall GRN determining neoblast behavior (Figure 2.7E).

2.2.7 Available transcriptomic data supports putative gene regulatory networks

As an independent test of the regulatory links between TFs predicted by our GRNs, we used transcriptomic data from published RNAi knockdown experiments for TFs that had predicted footprints. We analyzed transcriptomic data for *coe*(RNAi) (Cowles et al., 2014), *foxD1*(RNAi) (Vogg et al., 2014), *pax2/5/8-1* (Stower's Institute, 2016), *myoD*(RNAi) (Scimone et al., 2017), *nkx1-1* (Scimone et al., 2017), *box1*, *box3a*, and *lox5b* (Arnold et al., 2021). These available datasets had been generated from experiments using whole worms or blastemal tissues rather than the stem cell population. Nonetheless, in all cases both stem cells (X1 and X2), recent stem cell progeny (X2), and differentiated tissues (Xins) will have been present in these samples. We performed differential gene expression (DGE) analysis (Supplementary file 7) and obtained differentially expressed transcription factors (DETFs) as an independently derived set of functionally validated potential target TFs (Supplementary file 8).

To investigate whether the predicted regulatory links of our putative GRNs were supported by this collection of RNAi knockdown transcriptomic data, we compared the frequency of observed DETFs with predicted regulation by the RNAi-targeted TF in our GRNs to the expected frequency with random assignment (Figure 2.8A–C). The null hypothesis assumes that differential expression and enhancers are randomly assigned to genes and that footprints are randomly assigned to the enhancers (for details, see 'Materials and methods'). On average across these eight TFs with RNAi transcriptome data, we discovered 10.38 times more regulatory links between TFs than expected supported by transcriptome data with unbound footprints in a predicted enhancer. For TFs with predicted bound footprints, we found on average 21.09 times more regulatory links than expected (unbound p-value<0.0001, bound p-value<0.0001; Figure 2.8A–C). Results varied between knockdown experiments, indicating that the whole worm transcriptomic data reflects stem cell transcriptomics better for some TFs than others (Figure 2.8A–D). The predictive power of the GRNs developed in our study correlates with the proportional expression of the knockdown TF in the X1 compartment, which only contains stem cells, and inversely correlates with the level of expression in the Xins compartment, which only contains differentiated cells (Figure 2.8D, S2.5A, B), suggesting that our GRNs are indeed reflective of activity in stem cells but not necessarily in postmitotic differentiating or differentiated cells.

For instance, we examined the potential regulatory links of *coe* in our GRN model and found eight putative target TFs associated with an enhancer containing a potentially bound footprint for *coe*. Upon *coe*(RNAi) of whole worms, we observed eight DETFs, and by collating these two datasets, we obtained *pou4-1* as the only DETF containing a putative bound *coe* footprint (Figure 2.8E). This putative footprint is located within an enhancer downstream of the gene body (Figure 2.8E). Indeed, Cowles et al., 2014 tested DETFs in their transcriptomic analysis and observed that both *coe*(RNAi) and *pou4-1*(RNAi) lead to neural defects, demonstrating that the regulatory link corroborated by our GRN could be functionally validated within the

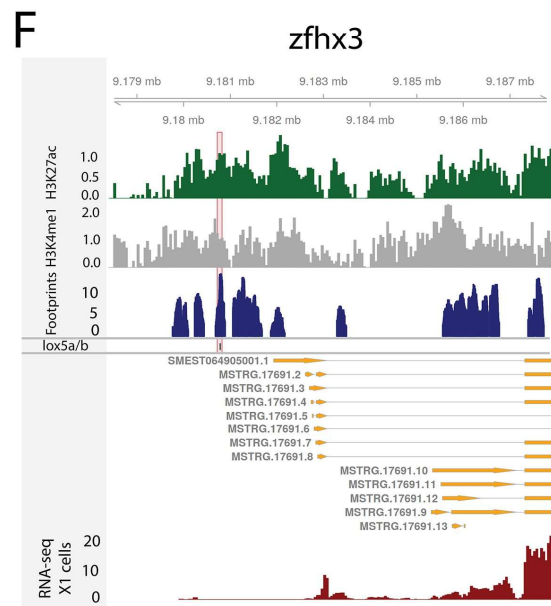
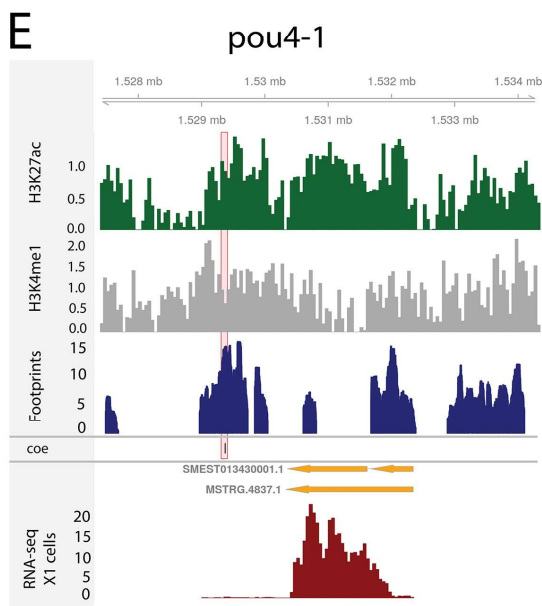
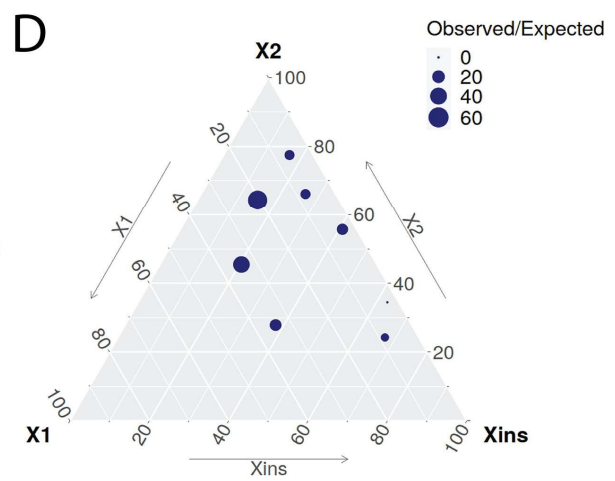
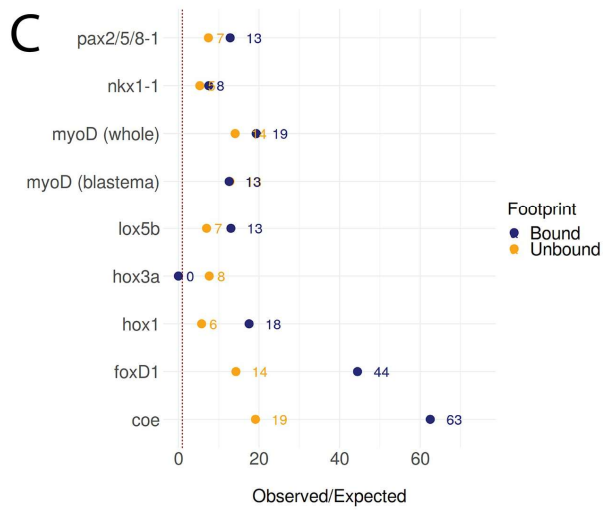
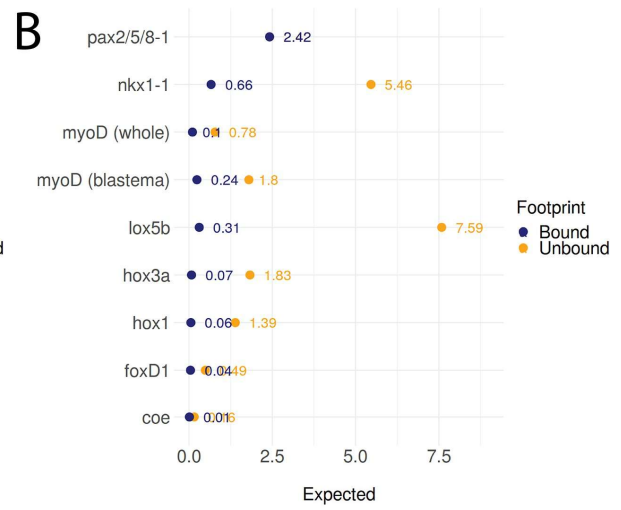
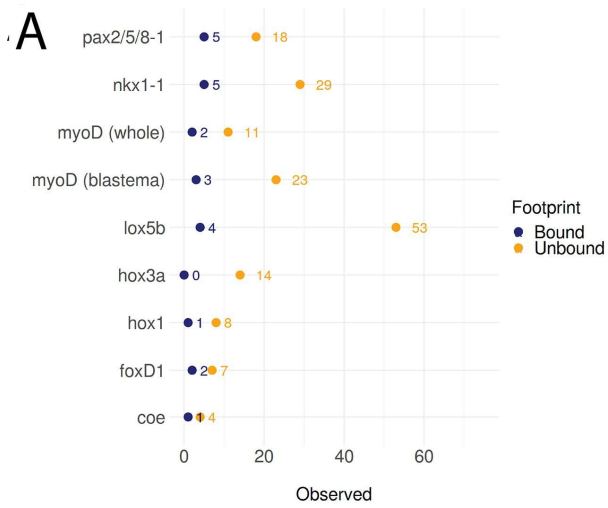


Figure 2.8 Transcriptomic data from knockdown experiments supports predicted regulatory links.

(A) The number of observed regulatory links between the knocked down transcription factor (TF) (y-axis) and differentially expressed TFs in our gene regulatory network (GRN) model. The regulatory links with unbound and bound footprints are displayed separately. **(B)** The number of expected regulatory links between the knocked down TF (y-axis) and differentially expressed TFs. **(C)** The ratio of observed and expected regulatory links as obtained from transcriptomic data for knocked down TFs. The regulatory links with unbound and bound footprints are displayed separately. **(D)** A ternary plot displaying the observed/expected ratio alongside the proportional expression values of the X1, X2, and Xins compartments. The size of the dot corresponds to the observed/expected ratio. The ratio is higher closer to the X1 and X2 compartment maxima, indicating that the GRN model predicts links better for TFs with higher expression in stem cells. **(E)** Genomic track of *pou4-1*. One coe bound footprint within a downstream enhancer-like region (width 100 bp) is marked in red. **(F)** Genomic track of *zfhx3*. One *lox5a/b* bound footprint within an upstream enhancer-like region (width 100 bp) is marked in red.

limits of the experimental toolkit of planarians. As for unstudied zinc fingers, we found that *zfhx3* is differentially expressed in *lox5b*(RNAi) whole worms and has a putative bound footprint of *lox5b* within an enhancer upstream of the gene body (Figure 2.8F). These analyses overall independently validate some of our predictions and provides a set of high-confidence predictions for further studies of TFs that were found to be differentially expressed after RNAi of TFs predicted to bind nearby enhancers.

2.3 Discussion

In this study, we improved the current genome annotation of *S. mediterranea* by integrating available RNA-seq data and identified new coding and non-coding transcripts. We reviewed the literature on both computationally and experimentally derived planarian TFs and defined a high-confidence set of TFs. If possible, we also predicted binding motifs for the planarian TFs. We developed and performed ATAC-seq on the proliferating stem cell compartment and determined genomic regions of open chromatin. We analyzed genome-wide profiles of the histone modifications H3K27ac and H3K4me1, along with chromatin accessibility data, and used these epigenetic datasets to delineate putative active enhancers in planarian stem cells. Lastly, we identified TFs binding to the enhancers of potential target genes and constructed hypothetical GRNs active in the stem cells.

We found more than 50,000 new isoforms to known transcripts in our annotation. This information has already helped to improve the clustering of single-cell RNA-seq data (García-Castro et al., 2021). Our approach demonstrated that the principle of integrating available RNA-seq data into a comprehensive expression-driven annotation can significantly improve genome annotation. We were also able to annotate more than 7000 new loci, with over 3000 predicted to be coding, increasing the number of protein coding genes by more than 10%. These new proteins tend to be shorter, are more likely to be proteins specific to planarians, but are expressed at similar levels to the rest of the coding transcriptome. This, together with the extensive annotation of alternate splice forms, gives a more complete picture of the genome of the model planarian.

Altogether, we annotated 551 TFs that were distributed evenly with respect to enrichment across all three FACS cell compartments in planarians (X1, X2, and Xins). In planarians, most studied functionally have been involved in regulating differentiation. Here, we define many neoblast-enriched TFs that have not been formally studied, with potential roles in maintaining neoblast pluripotency and potentially early lineage

commitment. A future systematic functional screen of these may help to uncover the GRN network that maintains pluripotency analogous to that in mammals and vertebrates (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). In particular, we found a large number of uncharacterized planarian zinc finger TFs, and this large and diverse family of TFs is still relatively poorly studied, perhaps because they evolve relatively rapidly in metazoans in general and often homology cannot be confidently assigned across phyla (Albà, 2017; Cassandri et al., 2017; Najafabadi et al., 2017). Currently little is known about transcriptional control of pluripotency in adult stem cells across animals, outside of vertebrates. In planarians, both the NuRD (Jaber-Hijazi et al., 2013) and BRG/Brahma complexes (Onal et al., 2012) have been implicated in regulating pluripotency through their RNAi phenotypes that block differentiation without affecting stem cell maintenance/self-renewal. By analogy with vertebrates, these chromatin remodeling complexes may directly regulate the activity of pluripotency TFs, which may include some unstudied TFs in our predicted GRNs.

Overall, for the main goals of this study, our analysis and identification of TFs allowed us to confidently assign likely binding motifs to just under half of the annotated planarian TFs (Figure 2.2E, Supplementary file 1). In the future as the number of planarians studies using ATAC data increases, and by looking at the actual sequence motifs implicated by footprinting analyses, it should be possible to refine motifs for planarian TFs and even define motifs for some TFs to which motifs could not be assigned, for example, in the diverse zinc finger TF group.

Based on combining epigenomic experiments and data types, we could identify putative intergenic and intronic enhancers in the planarian genome of proliferating stem cells. The combined use of ChIP-seq data, RNAi of a histone methyltransferase combined with other ChIP-seq, ATAC-seq data, and footprinting analyses together provided strong evidence for the identification of bona fide planarians stem cell enhancers. In the future, genome-wide ChIP-seq against the transcriptional cofactor p300 could serve as a complementary high-throughput approach to further substantiate these enhancers (Visel et al., 2009; Schwaiger et al., 2014), if available antibodies recognize the planarian ortholog of this protein (Fraguas et al., 2021; Stelman et al., 2021).

Verification of the function and targets of planarian enhancers is currently not possible using traditional approaches as no transgenic reporter technologies enabling enhancer-reporter constructs to be assayed are available. Therefore, we are bound to rely on less direct evidence from genome-wide sequencing technologies. Here, we assigned enhancers to target genes based on distance and the expression of the target gene, similar to established protocols used by others (Duren et al., 2017). We find cases where regulatory interactions suggested by previous expression-based studies of neoblast fate control and RNAi-based studies of gene function are supported by the GRNs interactions uncovered by our analyses. We find evidence that regulators of one differentiation lineage bind the enhancers of genes that regulate another, presumably acting as repressors. Our predicted GRNs were statistically supported by differential expression analyses of RNA-seq sets collected after RNAi of specific TFs (Figure 2.8), suggesting that many of the predicted regulatory interactions may be real. Experiments based on the predictions from our data combining RNAi against TFs with RNA-seq and ATAC-seq approaches will allow these GRNs to be studied further to help us understand how stem cells drive regeneration and homeostasis in planarians. In the future, both promoter-capture HiC and the co-accessibility of putative enhancers and promoters in scATAC-seq data would offer further computational possibilities to study these promoter-enhancer interactions (Schoenfelder et al., 2015; Pliner et al., 2018).

Taken together, our definition of enhancers in stem cells genome wide creates a foundation for constructing detailed GRNs to help understand regenerative mechanisms driven by stem cells in planarians.

2.4 Methods

2.4.1 Reference assembly and annotations

The sexual genome SMESG.1 genome, the SMESG high-confidence annotation (SMESG-high release 1), and the SMESG annotation filtered for repeats (SMESG-repeat release 2) were downloaded from PlanMine (Brandl et al., 2016; Grohme et al., 2018). Available RNA-seq datasets were aligned to the genome with HISAT2 (version 2.1.0) using default parameter settings and providing known splice sites from the SMESG-high annotation (Kim et al., 2015). Transcripts were assembled and merged from the alignments with StringTie using the SMESG-high annotation as a reference (Pertea et al., 2015; Pertea et al., 2016).

The new expression-driven annotation (García-Castro et al., 2021, https://github.com/jakke-neiro/Oxplatys/raw/gh-pages/Schmidtea_mediterranea_Oxford_v1.gtf.zip) was compared to the SMESG-high annotation with gffcompare (Pertea et al., 2015; Pertea et al., 2016). Transcripts labeled with the class code “=” were classified as full matches corresponding to the transcripts in the SMESG.1 genome, while transcripts labeled with the class codes “c”, “k”, “m”, “n”, “j”, “e”, and “o” were classified as isoforms. Lastly, transcripts labeled with class codes “u” (unknown or intergenic), “i” (fully contained within a reference intron), and “x” (exonic overlap on the opposite strand) were selected as potential candidates for new coding and non-coding transcripts (Pertea et al., 2016; Wu et al., 2016, Azlan et al., 2019). Transdecoder was used to identify ORFs in these new transcripts, and transcripts longer than 100 amino acids were classified as putative coding transcripts (Haas et al., 2013; Wu et al., 2016, Azlan et al., 2019). Subsequently, InteProScan and BlastX against UniProt were used to look for protein-like structures in the translated ORFs of the remaining transcripts, and transcripts without any hits were retained as putative non-coding transcripts (Jones et al., 2014; Wu et al., 2016).

2.4.2 Expression values, homology, and GO

To obtain FACS-specific expression values, the same FACS RNA-seq datasets used by Dattani et al., 2018 were used. The RNA-seq datasets were pseudo-aligned with Salmon using selective alignment, k-mer size 31 and 50 bootstrap iterations (Patro et al., 2017). The abundance estimates were converted to the Kallisto compatible format with wasabi, and the transcript counts were normalized with Sleuth (Pimentel et al., 2017). The TPM values of individual transcripts were summed to calculate TPM values for each gene. The mean TPM value across all samples in a FACS cell compartment was calculated for each gene to get the final absolute TPM value. The proportional expression values ($Xk\%$) were calculated by dividing the compartment-specific TPM value by the sum of TPM values of all compartments:

$$Xk\% = \frac{Xk(TPM)}{X1(TPM) + X2(TPM) + Xins(TPM)}, Xk \in \{X1, X2, Xins\}$$

Transcripts were categorized into FACS enrichment categories as follows: X1 when $X1\% > 0.5$, X2 when $X2\% > 0.5$, Xins when $Xins\% > 0.5$, X1 and X2 when $X1\% + X2\% > 0.75$ and neither enriched in X1 nor X2, X1, and Xins when $X1\% + Xins\% > 0.75$ and neither enriched in X1 nor Xins, X2, and Xins when $X2\% + Xins\% > 0.75$ and neither enriched in X2 nor Xins, and ubiquitous when not enriched in any of the above categories (Dattani et al., 2018). Transcript lengths and expression values were compared with Kruskal–Wallis test. Ternary plots and kernel density estimation of the proportional expression values were generated with the package ggtern (Hamilton and ggtern, 2018). Information content (IC) was defined to represent as a singular scalar the divergence of a combination of proportional expression values from the even distribution point ($X1\% = X2\% = Xins\%$):

$$IC = 0.01 \cdot p_1 \cdot \log(p_1) + 0.01 \cdot p_2 \cdot \log(p_2) + 0.01 \cdot p_3 \cdot \log(p_3)$$

where $p_1 = X1\%$, $p_2 = X2\%$, $p_3 = Xins\%$, and $\min(\{p_1, p_2, p_3\}) = 0.01$, meaning that proportional expression values below 0.01 were assigned the value 0.01.

The homology of the transcripts was investigated by using Blastx (Altschul et al., 1990). The non-coding and coding transcripts were aligned to the transcriptomes of humans (GRCh38.p13) and *D. japonica* (assembled in García-Castro et al., 2021). The threshold for non-coding transcripts was $e\text{-value} = 10^{-5}$ and for coding transcripts $e\text{-value} = 10^{-10}$. GO enrichment analysis of new coding transcripts with respect to known coding transcripts was performed using topGO with Fisher's exact test (Alexa and Rahnenfuhrer, 2020). The data for proportional expression and homology assigned by BLAST of all transcripts is available at <https://jakke-neiro.github.io/Oxplatys/>; Neiro, 2020.

2.4.3 Identification and characterization of transcription factors

We conducted a systematic domain annotation of all transcripts by using the InterProScan resource (Jones et al., 2014). Transcripts with an InterProScan description 'transcription factor', a Pfam hit to the Pfam families listed in the TF database DNA-binding domain (DBD) v2.0, or a SUPERFAMILY hit listed in the SUPERFAMILY families in DBD were classified as TFs (Wilson et al., 2008). We used Blast to align the potential planarian TFs to TFs in humans and fruit flies, and potential TFs without hits were filtered out. The planarian literature was systematically reviewed (Supplementary file 1), and if a TF was mentioned, the GenBank accession number or primer information was retrieved to establish the exact sequence in the literature. This sequence from the literature was aligned to the potential TFs (default parameters for GenBank accession entries and word size 10 for primers) to correctly assign the TFs in the literature to our transcripts (Supplementary file 1). If a planarian TF had previously been described in the literature under a certain name, one name was chosen as the primary name, while all other alternative synonyms used for the same TF were listed as secondary names (Supplementary file 1). If a potential TF had not been described in the planarian literature, the name of the best human or fruit fly Blast hit was used (Supplementary file 1). The TFs were categorized into four main groups (Basic domain, Zinc fingers, Helix-turn-helix, Alpha-helical, Immunoglobulin-fold, Beta-hairpin, Beta-sheet, Beta-barrel, Others) according to the TRANSFAC database (Stegmaier et al., 2004; Supplementary file 1).

Ternary plots, kernel density estimation, and the information content of the proportional expression values of TFs were calculated as for other transcripts (see 'Expression values, homology, and GO'). Hierarchical clustering of FACS proportions was performed using Euclidean distance and Ward's method with the package hclust (van den Boogaart and Tolosana-Delgado, 2008). The frequencies of different TF domains among the cellular enrichment classes X1, X2, and Xins (for which FACS proportional expression is higher or equal to 50%) were compared with the chi-square test using Yates correction.

2.4.4. Motifs of transcription factors

The motifs of the TFs (Supplementary file 1) were predicted by using the JASPAR profile inference tool (Fornes et al., 2020). The protein sequences were retrieved with Transdecoder (see 'Reference assembly and annotations'). The information content for each motif was determined by calculating the mean of Shannon's entropy at each position in the position weight matrix (PWM). The motifs were visualized with seqLogo (Bembom, 2019).

2.4.5 ATAC-seq library preparation

A standard protocol was used for preparing the ATAC-seq library (Buenrostro et al., 2013). The X1 cell compartments were isolated by FACS (Romero et al., 2012), and two replicates with 120,000–250,000 cells were collected from each compartment. The cells were washed and centrifuged (1 XPBS, 1200 RPM). Lysis buffer (10 mM Tris-Cl [pH 7.5], 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40) was added, the cells were centrifuged (500 RPM, 10 min, 4°C), and the nuclei pellet was collected. Then, the transposase mix (25 µl 2X TD Buffer, 2.5 µl Tn5 Transposase, 22.5 µl of nuclease-free water) was added, and the cells were resuspended and incubated (37°C, 60 min). Finally, DNA was isolated using the ZymoGen Clean & Concentrator Kit, and eluted in EB buffer.

Subsequently, the eluted DNA was used for PCR amplification and library purification. DNA was amplified using a standard reaction (10 µl of purified transposed DNA, 10 µl of nuclease-free water, 15 µl Nextera PCR Master Mix, 5 µl of PCR primer cocktail, 5 µl Index Primer 1, 5 µl Index Primer 2, 72°C, 3 min, 98°C, 30 s, 14 cycles of 98°C, 10 s, 63°C, 30 s, 72°C for 1 min). Finally, the libraries were cleaned with the AMPure bead purification kit. Samples were paired-end sequenced on the Illumina NextSeq. The sample replicates can be found on Sequence Read Archive (SRR18923214 and SRR18923213).

2.4.6 ChIP-seq library preparation and sequencing

The H3K27ac ChIP-seq library was prepared and sequenced using established protocols (Mihaylova et al., 2018; Dattani et al., 2018). A total of 600,000–700,000 planarian X1 cells were isolated for each experimental replicate (actual sample and input control). Two sample replicates (Sequence Read Archive SRR18925505 and SRR18925504) and two input replicates (Sequence Read Archive SRR18925503 and SRR18925502) were prepared. The H3K27ac Abcam ab4729 antibody was used for immunoprecipitation.

2.4.7 ChIP-seq data analysis

The ChIP-seq data with respect to H3K4me1 and H3K4me1 after lpt(RNAi) was reanalyzed (Mihaylova et al., 2018, PRJNA338116) alongside the H3K27ac data prepared here. The reads were quality-checked with FastQC and trimmed with Trimmomatic (Andrews et al., 2010; Bolger et al., 2014). The reads were aligned to the SMESG.1 genome with Bowtie2 (Langmead et al., 2009). Only uniquely mapped reads and reads with a quality score greater than 10 were retained with Samtools (Li et al., 2009). Peaks were called with MACS2 with default parameters for H3K4me1 and the broad peak parameter for H3K27ac (Zhang et al., 2008). Coverage tracks were generated with deepTools bamCoverage, and the mean coverage track was determined with wiggletools (Zerbino et al., 2014; Ramírez et al., 2016). Heatmaps of ChIP-seq profiles were generated with deepTools computeMatrix and plotHeatmap 1000 bp upstream and downstream of the peak center (Ramírez et al., 2016). The mean peak value was calculated as the mean of the ChIP-seq signal 1000 bp upstream and downstream of the peak center. The H3K27ac peaks that were at most 500 bp from the nearest H3K4me1 peak were selected as putative enhancers. Random genomic regions were generated with bedtools random (Quinlan, 2014). The difference of mean peak value at enhancers and random regions was tested with Wilcoxon's test. Detailed analysis and code are available at <https://jakke-neiro.github.io/Oxplatys/>; Neiro, 2020.

2.4.8 ATAC-seq data analysis and motif footprinting

The quality of paired-end reads was assessed with FastQC and adapter sequences were removed with Trimmomatic (Andrews et al., 2010; Bolger et al., 2014). The reads were aligned to the SMESG.1 genome with Bowtie2 (Langmead et al., 2009). Only uniquely mapped reads and reads with a quality score greater than 10 were retained with Samtools (Li et al., 2009). Duplicates were removed with Samtools (Li et al., 2009). Insert sizes were calculated with Picard CollectInsertSizeMetrics (<http://broadinstitute.github.io/picard/>; Magner, 2022).

For coverage track generation and peak calling, samples were downsampled to a standard of ~20M reads using Picard DownsampleSam (<http://broadinstitute.github.io/picard/>; Magner, 2022). DeepTools2 bamCoverage was used to generate coverage tracks for each sample in bigwig and bedgraph formats by normalizing with respect to reads million mapped reads (RPKM) (Ramírez et al., 2016). TOBIAS was used to look for footprints in the ATAC-seq by using the motifs predicted by JASPAR (see 'Motifs of transcription factors'; Bentsen et al., 2020). BAM files of mapped reads are available at <https://jakke-neiro.github.io/Oxplatys/>; Neiro, 2020.

2.4.9 Enhancer targets and gene regulatory networks

The putative enhancers were assigned to the nearest TSS with ChIPSeeker (Yu et al., 2015). The footprints in enhancers were used to create links between TFs and target genes for network construction using a

Python script (available at <https://jakke-neiro.github.io/Oxplatys/>; Neiro, 2020). Networks were visualized with the R package Igraph (<https://github.com/igraph/rigraph>; Nepusz, 2022) while the genomic regions were visualized with Gviz (Hahne and Ivanek, 2016).

2.4.10 Functional transcriptomics and validation of gene regulatory network predictions

To calculate the odds ratio of our approach, we defined the expected value of the number regulatory links for a given TF. A regulatory link for a given TF assumes that the target gene is linked to an enhancer and that the enhancer contains a footprint for the given TF. In addition, if we assume that all differentially expressed genes are real direct target genes for the given TF, then the target gene of the regulatory link is also differentially expressed upon perturbation of TF activity. First, we define A as the event that a TF is differentially expressed. The probability of this event is

$$P(A) = \frac{n_{dge}}{n_{ge}}$$

where n_{ge} is the total number of TFs in the genome, and n_{dge} is the number of differentially expressed TFs. Second, we define B as the event that a given TF is linked to an enhancer. The probability of this event is

$$P(B) = \frac{n_{cis}}{n_{ge}}$$

where n_{cis} is the total number of TFs linked to an enhancer. Third, we define C as the event that an enhancer contains a footprint (bound motif) of the given TF. The probability of this event is

$$P(C) = \frac{n_{cisTF}}{n_{cis}}$$

where n_{cisTF} is the number of enhancer-linked genes containing the footprint of the given TF in their enhancers.

We assume that the events A , B , and C are independent. The expected number of differentially expressed genes linked to enhancers with the footprint of a given TF is equal to

$$P(A \cap B \cap C)n_{ge} = P(ABC)n_{ge} = P(A)P(B)P(C)n_{ge}$$

2.4.11 Code availability

All codes are available in a notebook provided as a zip file (Source code 1).

2.4.12 Data availability

Data and analyses are available at <https://jakke-neiro.github.io/Oxplatys/>, (copy archived at swh:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25). All analysis code is provided in supplementary file New sequence data are available at the NCBI under Bioproject ID PRJNA832235, <https://www.ncbi.nlm.nih.gov/bioproject/832235>.

2.5 Supplementary figures

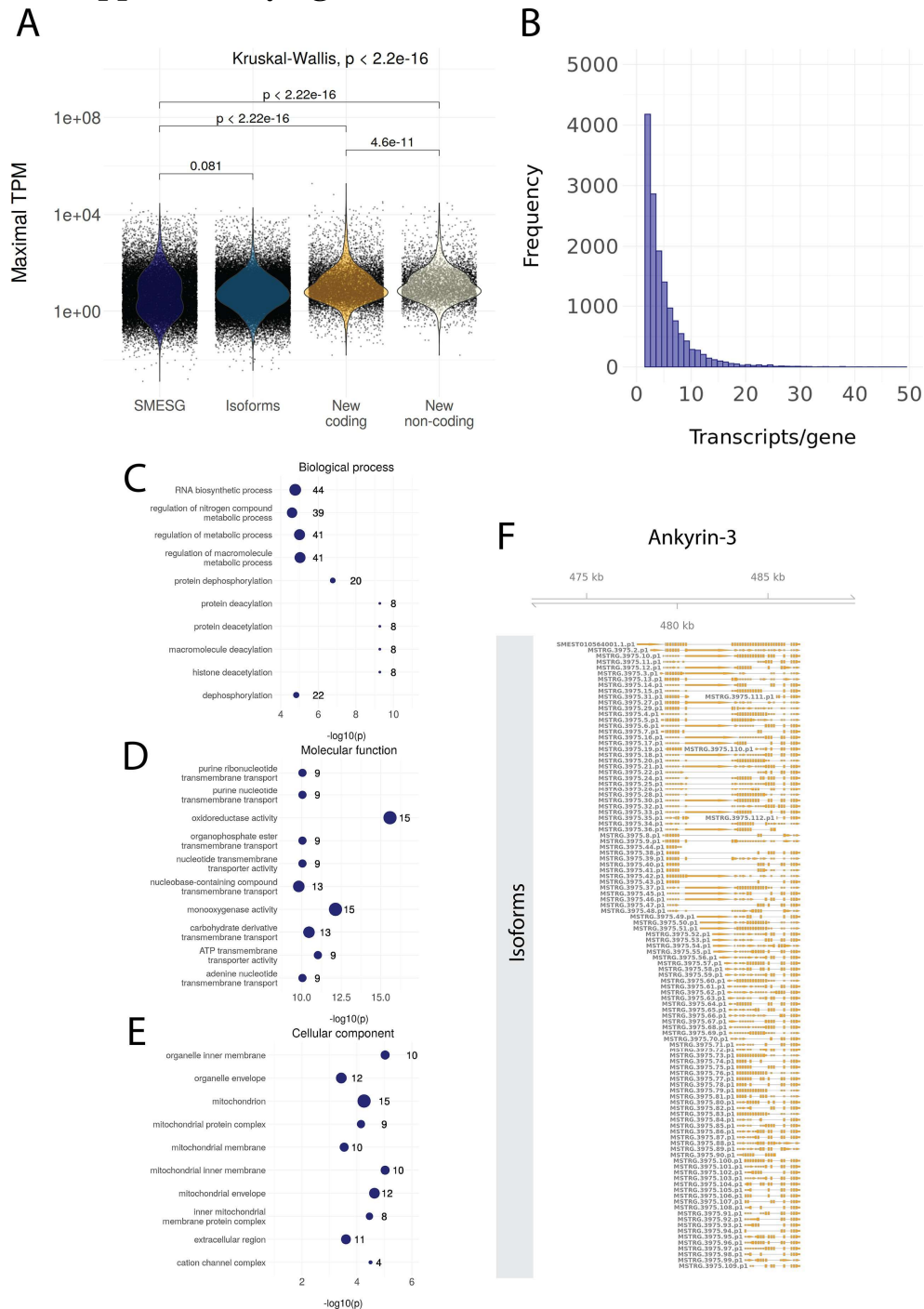


Figure S2.1 Information on the genome annotation of *Schmidtea mediterranea*. (A) Maximal transcripts per kilobase million (TPM) of known transcripts (SMESG, $n = 28624$), Isoforms ($n = 50213$), new coding transcripts ($n = 4752$), and new non-coding transcripts ($n = 5884$). Maximal TPM is defined as the highest TPM value in the 183 RNA-seq samples used for reannotation. Kruskal-Wallis test is used to compare all groups and Wilcoxon rank-sum test is used for pairwise comparisons. (B) The number of transcripts per gene (distribution up to 50 transcripts/gene). (C) The Gene Ontology (GO) terms for biological processes enriched in the new coding transcripts. (D) The GO terms for molecular functions enriched in the new coding transcripts. (E) The GO terms for cellular components enriched in the new coding transcripts. (F) All isoforms for ankyrin-3 (ankyrin-G).

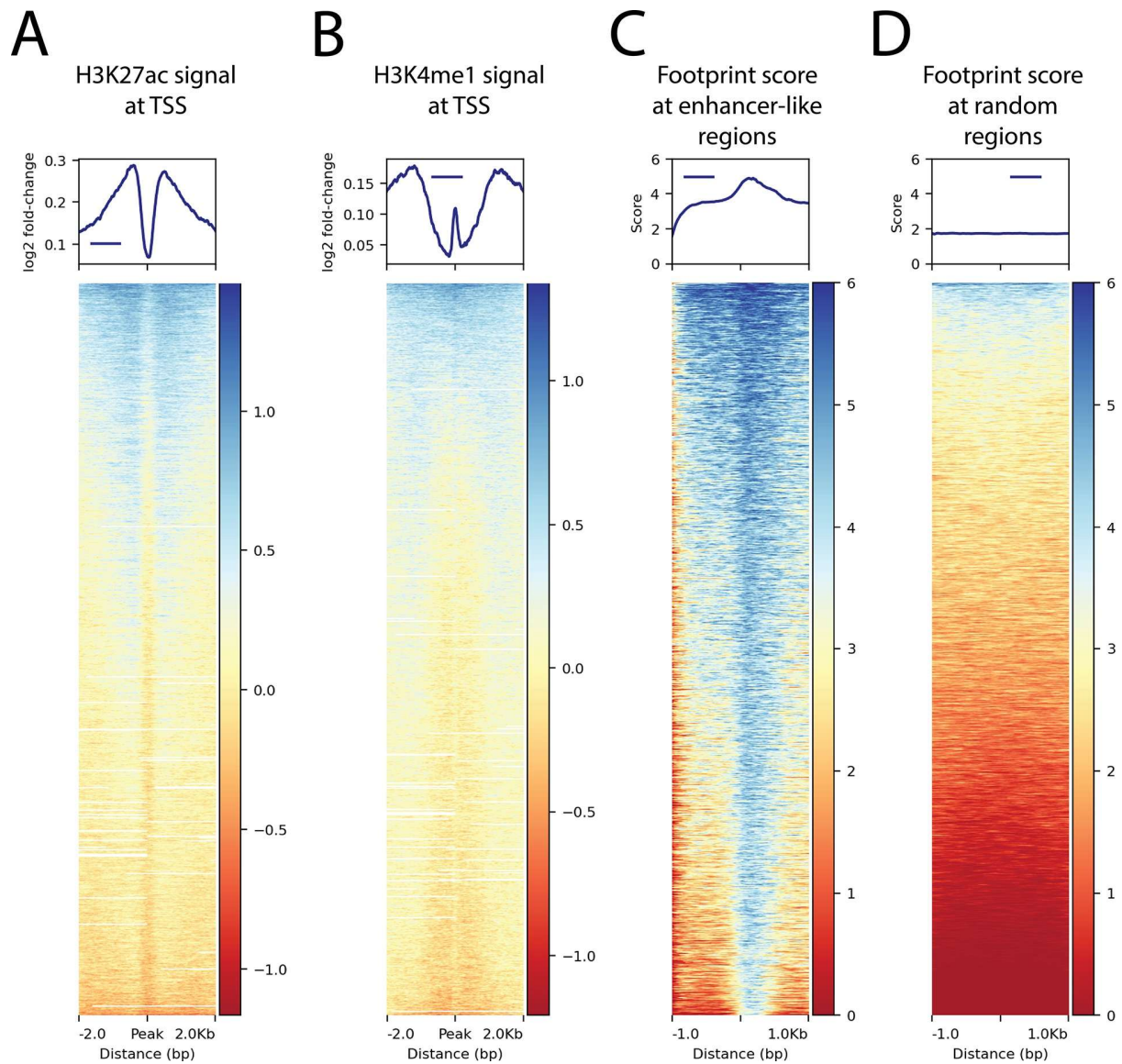
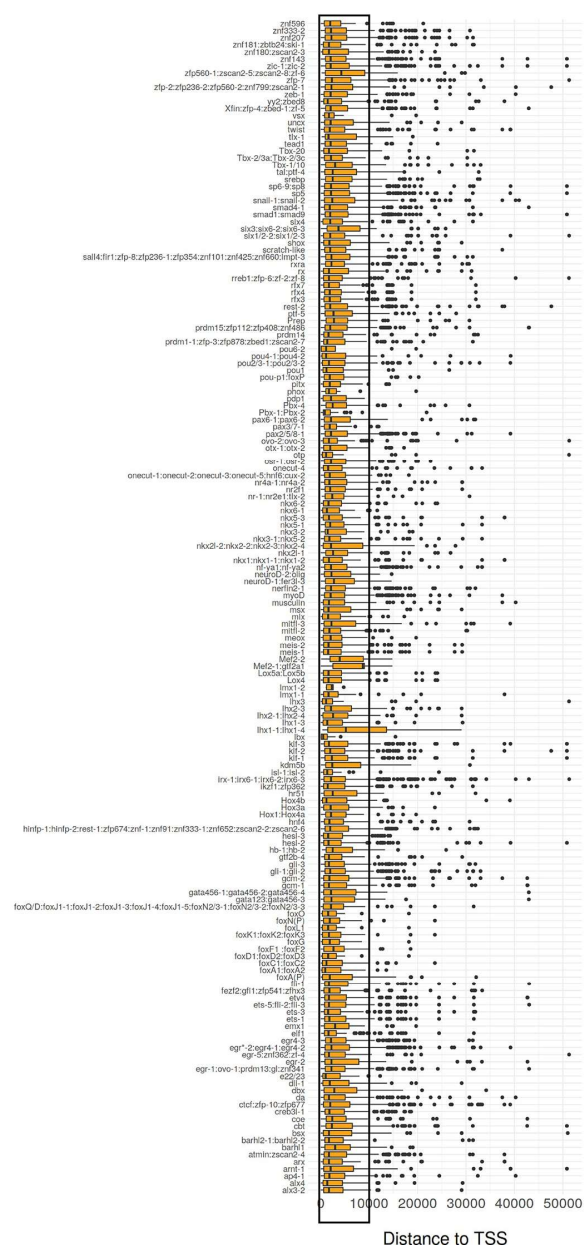


Figure S2.2 Histone modifications at transcription start sites and footprinting scores in relation to enhancer-like regions. (A) H3K27ac ChIP-seq signal around transcription start sites (TSS) (B) H3K4me1 ChIP-seq signal around TSS. (C) The ATAC-seq footprint score at the combined peaks or putative enhancer-like regions. (D) The ATAC-seq footprint score at 10,000 random regions in the genome.

A



B

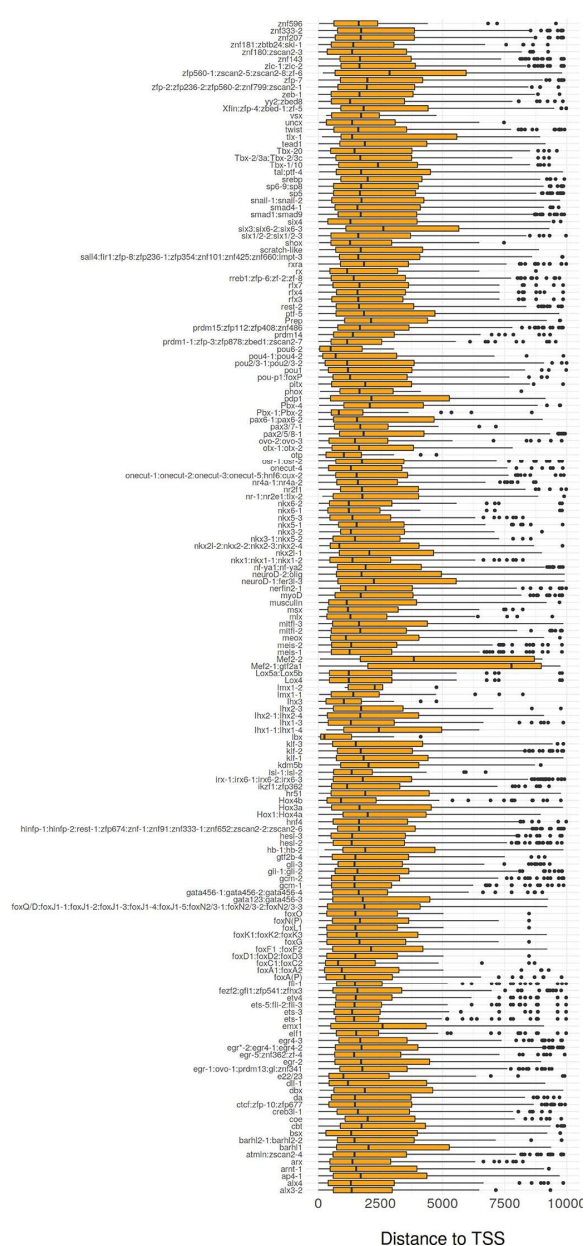


Figure S2.3 Distance of bound transcription factor footprints to transcriptional start sites. Boxplot of distance to transcriptional start site (TSS) for bound transcription factor footprints. Transcription factors predicted to have the same motif are grouped together. **(B)** Detailed view of **(A)** for distances from 0 bp to 10,000 bp.

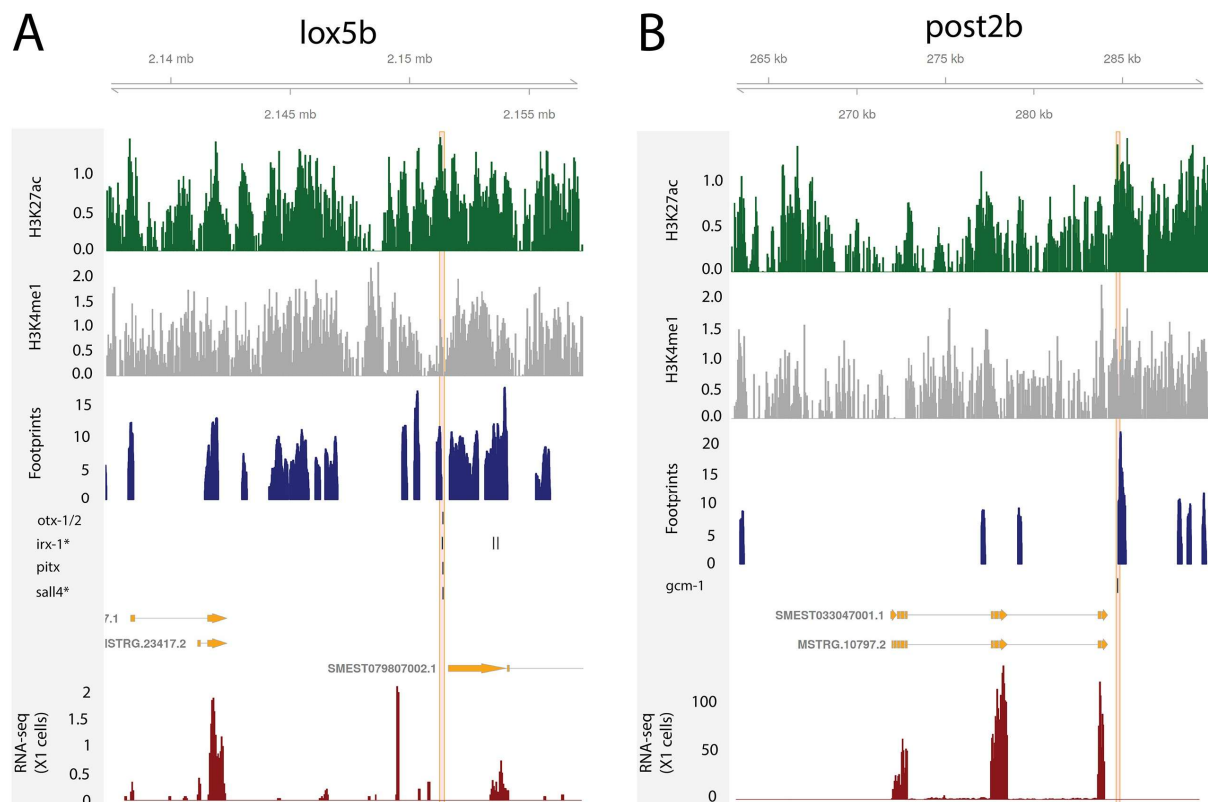


Figure S2.4 Regulatory regions in relation to Hox genes. (A) Predicted enhancers binding sites at the lox5b locus. **(B)** Predicted bound enhancer at the post2B locus.

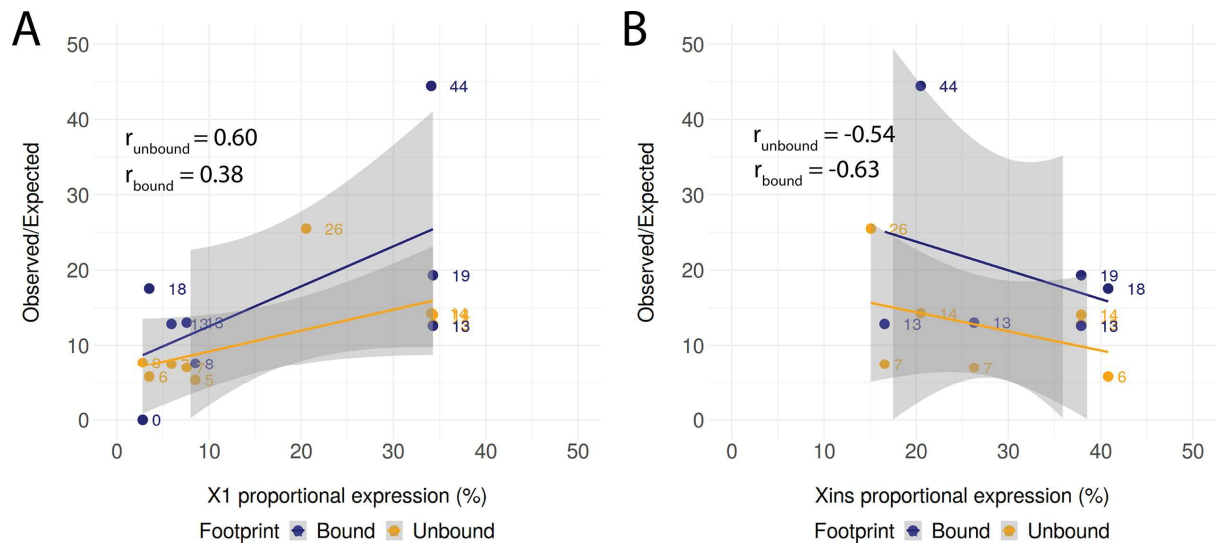


Figure S2.5 Correlation of compartment-specific proportional expression and the enrichment of observed regulatory links. (A) Positive correlation of X1 proportional expression and the enrichment of observed regulatory links as compared to the expected number of regulatory links. The shaded area represents the confidence interval (0.95) for the linear regression lines. (B) Negative correlation of Xins proportional expression and the enrichment of observed regulatory links as compared to the expected number of regulatory links. The shaded area represents the confidence interval (0.95) for the linear regression lines.

2.6 Supplementary material

Supplementary file 1

Full list of transcription factors (TFs) annotated in the *S. mediterranea* genome.

TFs are named according to existing names in the literature or according to the currently agreed convention. The absolute (transcripts per million) or proportional expression in the fluorescence-activated cell sorting (FACS)-associated cell compartments X1, X2, and Xins is also recorded.

Supplementary file 2

Full list of putative enhancers in adult stem cells defined by ChIP-seq and ATAC-seq analyses.

Includes the position in the genome and the identify of the predicted target gene based on proximity to the predicted transcriptional start site.

Supplementary file 3

Full list of predicted ATAC-seq footprints in the genome, including predicted transcription factor (TF)/TF family, position of binding site in the genome, and position of the peak/enhancer in the genome.

Supplementary file 4

Full list of motifs with and without footprints organized by transcription factor (TF)/TF family in predicted enhancer regions.

Supplementary file 5

Overview of number of enhancers, predicted bound motifs (footprints), and unbound motifs for fate-specific transcription factors (FSTF) expressed in adult stem cells.

Supplementary file 6

Full list of transcription factors (TFs), involved in a putative stem cell gene regulatory network (GRN).

Includes TF with expression $>1/3$ in the X1 compartment and includes other TFs implicated in regulating or being regulated by these TFs.

Supplementary file 7

Differential expression analysis of transcription factor (TF) RNAi with RNA-seq datasets.

Supplementary file 8

Differentially expressed transcription factor (TFs) after TF RNAi.

2.7 References

- Abnave P, Aboukhatwa E, Kosaka N, Thompson J, Hill MA, Aboobaker AA. 2017. Epithelial-mesenchymal transition transcription factors control pluripotent adult stem cell migration in vivo in planarians. *Development* 144:3440–3453. DOI: <https://doi.org/10.1242/dev.154971>, PMID: 28893948
- Aboobaker AA. 2011. Planarian stem cells: a simple paradigm for regeneration. *Trends in Cell Biology* 21:304–311. DOI: <https://doi.org/10.1016/j.tcb.2011.01.005>, PMID: 21353778
- Adler CE, Sánchez Alvarado A. 2015. Types or states? cellular dynamics and regenerative potential. *Trends in Cell Biology* 25:687–696. DOI: <https://doi.org/10.1016/j.tcb.2015.07.008>, PMID: 26437587
- Albà MM. 2017. Zinc-finger domains in metazoans: evolution gone wild. *Genome Biology* 18:1–3. DOI: <https://doi.org/10.1186/s13059-017-1307-y>, PMID: 28877710
- Alexa A, Rahnenfuhrer J. 2020. TopGO: enrichment analysis for gene ontology. R Package Version 2.40.0.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2), PMID: 2231712
- An Y, Kawaguchi A, Zhao C, Toyoda A, Sharifi-Zarchi A, Mousavi SA, Bagherzadeh R, Inoue T, Ogino H, Fujiyama A, Chitsaz H, Baharvand H, Agata K. 2018. Draft genome of *Dugesia japonica* provides insights into conserved regulatory elements of the brain restriction gene *nou-darake* in planarians. *Zoological Letters* 4:24. DOI: <https://doi.org/10.1186/s40851-018-0102-2>, PMID: 30181897
- Andrews S, Krueger F, Secondi-Pichon A, Biggins F, Wingett S. 2010. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*.
- Arnold CD, Gerlach D, Stelzer C, Boryń LM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339:1074–1077. DOI: <https://doi.org/10.1126/science.1232542>, PMID: 23328393
- Arnold CP, Lozano AM, Mann FG, Nowotarski SH, Haug JO, Lange JJ, Seidel CW, Alvarado AS. 2021. Hox genes regulate asexual reproductive behavior and tissue segmentation in adult animals. *Nature Communications* 12:6706. DOI: <https://doi.org/10.1038/s41467-021-26986-2>, PMID: 34795249
- Azlan A, Obeidat SM, Yunus MA, Azzam G. 2019. Systematic identification and characterization of aedes aegypti long noncoding rnas (lncrnas). *Scientific Reports* 9:1–9. DOI: <https://doi.org/10.1038/s41598-019-47506-9>, PMID: 31434910
- Baldini A, Fulcoli FG, Illingworth E. 2017. Tbx1: transcriptional and developmental functions. *Current Topics in Developmental Biology* 122:223–243. DOI: <https://doi.org/10.1016/bs.ctdb.2016.08.002>, PMID: 28057265
- Bembom O. 2019. SeqLogo: sequence logos for DNA sequence alignments. R Package Version 1.52.0.
- Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. 2020. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nature Communications* 11:1–11. DOI: <https://doi.org/10.1038/s41467-020-18035-1>, PMID: 32848148
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>, PMID: 24695404

- Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* 44:148–156. DOI: <https://doi.org/10.1038/ng.1064>, PMID: 22231485
- Brandl H, Moon H, Vila-Farré M, Liu SY, Henry I, Rink JC. 2016. PlanMine--a mineable resource of planarian biology and biodiversity. *Nucleic Acids Research* 44:D764–D773. DOI: <https://doi.org/10.1093/nar/gkv1148>, PMID: 26578570
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* 10:1213–1218. DOI: <https://doi.org/10.1038/nmeth.2688>, PMID: 24097267
- Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Molecular Cell* 49:825–837. DOI: <https://doi.org/10.1016/j.molcel.2013.01.038>, PMID: 23473601
- Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G, Raschellà G. 2017. Zinc-finger proteins in health and disease. *Cell Death Discovery* 3:17071. DOI: <https://doi.org/10.1038/cddiscovery.2017.71>, PMID: 29152378
- Cowles MW, Brown DDR, Nisperos SV, Stanley BN, Pearson BJ, Zayas RM. 2013. Genome-wide analysis of the bhlh gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. *Development* 140:4691–4702. DOI: <https://doi.org/10.1242/dev.098616>, PMID: 24173799
- Cowles MW, Omuro KC, Stanley BN, Quintanilla CG, Zayas RM. 2014. COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians. *PLOS Genetics* 10:e1004746. DOI: <https://doi.org/10.1371/journal.pgen.1004746>, PMID: 25356635
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone h3k27ac separates active from poised enhancers and predicts developmental state. *PNAS* 107:21931–21936. DOI: <https://doi.org/10.1073/pnas.1016071107>, PMID: 21106759
- Currie KW, Brown DDR, Zhu S, Xu C, Voisin V, Bader GD, Pearson BJ. 2016. HOX gene complement and expression in the planarian *Schmidtea mediterranea*. *EvoDevo* 7:1–11. DOI: <https://doi.org/10.1186/s13227-016-0044-8>, PMID: 27034770
- Dattani A, Kao D, Mihaylova Y, Abnave P, Hughes S, Lai A, Sahu S, Aboobaker AA. 2018. Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells. *Genome Research* 28:1543–1554. DOI: <https://doi.org/10.1101/gr.239848.118>, PMID: 30143598
- Dattani A, Sridhar D, Aziz Aboobaker A. 2019. Planarian flatworms as a new model system for understanding the epigenetic regulation of stem cell pluripotency and differentiation. *Seminars in Cell & Developmental Biology* 87:79–94. DOI: <https://doi.org/10.1016/j.semcdb.2018.04.007>, PMID: 29694837
- Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. 2017. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Research* 27:2096–2107. DOI: <https://doi.org/10.1101/gr.226233.117>, PMID: 29141961

- Duncan EM, Chitsazan AD, Seidel CW, Sánchez Alvarado A. 2015. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. *Cell Reports* 13:2741–2755. DOI: <https://doi.org/10.1016/j.celrep.2015.11.059>, PMID: 26711341
- Duren Z, Chen X, Jiang R, Wang Y, Wong WH. 2017. Modeling gene regulation from paired expression and chromatin accessibility data. *PNAS* 114:E4914–E4923. DOI: <https://doi.org/10.1073/pnas.1704553114>
- Duren Z, Chen X, Xin J, Wang Y, Wong WH. 2020. Time course regulatory analysis based on paired expression and chromatin accessibility data. *Genome Research* 30:622–634. DOI: <https://doi.org/10.1101/gr.257063.119>, PMID: 32188700
- Emerson RO, Thomas JH. 2011. Gypsy and the birth of the SCAN domain. *Journal of Virology* 85:12043–12052. DOI: <https://doi.org/10.1128/JVI.00867-11>, PMID: 21865395
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49. DOI: <https://doi.org/10.1038/nature09906>, PMID: 21441907
- Felix DA, Aboobaker AA. 2010. The TALE class homeobox gene *smad-prep* defines the anterior compartment for head regeneration. *PLOS Genetics* 6:e1000915. DOI: <https://doi.org/10.1371/journal.pgen.1000915>, PMID: 20422023
- Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. 2018. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* 360:eaq1736. DOI: <https://doi.org/10.1126/science.aq1736>, PMID: 29674431
- Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* 48:D87–D92. DOI: <https://doi.org/10.1093/nar/gkz1001>, PMID: 31701148
- Forsthoefel DJ, Cejda NI, Khan UW, Newmark PA. 2020. Cell-type diversity and regionalized gene expression in the planarian intestine. *eLife* 9:e52613. DOI: <https://doi.org/10.7554/eLife.52613>, PMID: 32240093
- Fraguas S, Cárcel S, Vivancos C, Molina MD, Ginés J, Mazariegos J, Sekaran T, Bartscherer K, Romero R, Cebrià F. 2021. CREB-binding protein (CBP) gene family regulates planarian survival and stem cell differentiation. *Developmental Biology* 476:53–67. DOI: <https://doi.org/10.1016/j.ydbio.2021.02.008>, PMID: 33774010
- Fujimura N, Vacik T, Machon O, Vlcek C, Scalabrin S, Speth M, Diep D, Krauss S, Kozmik Z. 2007. Wnt-mediated down-regulation of *sp1* target genes by a transcriptional repressor *sp5*. *The Journal of Biological Chemistry* 282:1225–1237. DOI: <https://doi.org/10.1074/jbc.M605851200>, PMID: 17090534
- Gaiti F, Calcino AD, Tanurdžić M, Degnan BM. 2017a. Origin and evolution of the metazoan non-coding regulatory genome. *Developmental Biology* 427:193–202. DOI: <https://doi.org/10.1016/j.ydbio.2016.11.013>, PMID: 27880868

- Gaiti F, Jindrich K, Fernandez-Valverde SL, Roper KE, Degnan BM, Tanurdžić M. 2017b. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *eLife* 6:e22194. DOI: <https://doi.org/10.7554/eLife.22194>, PMID: 28395144
- Gao T, Qian J. 2020a. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research* 48:D58–D64. DOI: <https://doi.org/10.1093/nar/gkz980>, PMID: 31740966
- Gao B, Wang Y, Diaby M, Zong W, Shen D, Wang S, Chen C, Wang X, Song C. 2020b. Evolution of pogo, a separate superfamily of IS630-tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mobile DNA* 11:1–15. DOI: <https://doi.org/10.1186/s13100-020-00220-0>, PMID: 32742312
- García-Castro H, Kenny NJ, Iglesias M, Álvarez-Campos P, Mason V, Elek A, Schönauer A, Sleight VA, Neiro J, Aboobaker A, Permanyer J, Irimia M, Sebé-Pedrós A, Solana J. 2021. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biology* 22:1–34. DOI: <https://doi.org/10.1186/s13059-021-02302-5>, PMID: 33827654
- Gehrke AR, Neverett E, Luo YJ, Brandt A, Ricci L, Hulett RE, Gompers A, Ruby JG, Rokhsar DS, Reddien PW, Srivastava M. 2019. Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* 363:eaau6173. DOI: <https://doi.org/10.1126/science.aau6173>, PMID: 30872491
- Goldman JA, Kuzu G, Lee N, Karasik J, Gemberling M, Foglia MJ, Karra R, Dickson AL, Sun F, Tolstorukov MY, Poss KD. 2017. Resolving heart regeneration by replacement histone profiling. *Developmental Cell* 40:392–404. DOI: <https://doi.org/10.1016/j.devcel.2017.01.013>, PMID: 28245924
- Gorkin DU, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, Beer MA, Pavan WJ, McCallion AS. 2012. Integration of chip-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Research* 22:2290–2301. DOI: <https://doi.org/10.1101/gr.139360.112>, PMID: 23019145
- Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, Brandl H, Henry I, Dahl A, Powell S, Hiller M, Myers E, Rink JC. 2018. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* 554:56–61. DOI: <https://doi.org/10.1038/nature25473>, PMID: 29364871
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* 8:1494–1512. DOI: <https://doi.org/10.1038/nprot.2013.084>, PMID: 23845962
- Hahne F, Ivanek R. 2016. Visualizing genomic data using gviz and bioconductor. *Methods in Molecular Biology* 1418:335–351. DOI: https://doi.org/10.1007/978-1-4939-3578-9_16, PMID: 27008022
- Hamilton NE, ggtern FM. 2018. Ternary diagrams using ggplot2. *Journal of Statistical Software* 87:1–7. DOI: <https://doi.org/10.18637/jss.v087.c03>
- Harris RE, Setiawan L, Saul J, Hariharan IK. 2016. Localized epigenetic silencing of a damage-activated Wnt enhancer limits regeneration in mature *Drosophila* imaginal discs. *eLife* 5:e11588. DOI: <https://doi.org/10.7554/eLife.11588>, PMID: 26840050

- Hayashi T, Asami M, Higuchi S, Shibata N, Agata K. 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development, Growth & Differentiation* 48:371–380. DOI: <https://doi.org/10.1111/j.1440-169X.2006.00876.x>, PMID: 16872450
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* 39:311–318. DOI: <https://doi.org/10.1038/ng1966>, PMID: 17277777
- Hoff KJ, Stanke M. 2013. WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* 41:W123–W128. DOI: <https://doi.org/10.1093/nar/gkt418>, PMID: 23700307
- Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with augustus. *Current Protocols in Bioinformatics* 65:e57. DOI: <https://doi.org/10.1002/cpbi.57>, PMID: 30466165
- Hopitzan AA, Baines AJ, Ludosky MA, Recouvreur M, Kordeli E. 2005. Ankyrin-G in skeletal muscle: tissuespecific alternative splicing contributes to the complexity of the sarcolemmal cytoskeleton. *Experimental Cell Research* 309:86–98. DOI: <https://doi.org/10.1016/j.yexcr.2005.04.013>, PMID: 15953600
- Jaber-Hijazi F, Lo PJKP, Mihaylova Y, Foster JM, Benner JS, Tejada Romero B, Chen C, Malla S, Solana J, Ruzov A, Aziz Aboobaker A. 2013. Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Developmental Biology* 384:141–153. DOI: <https://doi.org/10.1016/j.ydbio.2013.09.020>, PMID: 24063805
- Jänes J, Dong Y, Schoof M, Serizay J, Appert A, Cerrato C, Woodbury C, Chen R, Gemma C, Huang N, Kissiov D, Stempor P, Steward A, Zeiser E, Sauer S, Ahringer J. 2018. Chromatin accessibility dynamics across *C. elegans* development and ageing. *eLife* 7:e37344. DOI: <https://doi.org/10.7554/eLife.37344>, PMID: 30362940
- Janssens J, Aibar S, Taskiran II, Ismail JN, Gomez AE, Aughey G, Spanier KI, De Rop FV, González-Blas CB, Dionne M, Grimes K, Quan XJ, Papasokrati D, Hulselmans G, Makhzami S, De Waegeneer M, Christiaens V, Southall T, Aerts S. 2022. Decoding gene regulation in the fly brain. *Nature* 601:630–636. DOI: <https://doi.org/10.1038/s41586-021-04262-z>
- Jegla T, Nguyen MM, Feng C, Goetschius DJ, Luna E, van Rossum DB, Kamel B, Pisupati A, Milner ES, Rolls MM. 2016. Bilaterian giant ankyrins have a common evolutionary origin and play a conserved role in patterning the axon initial segment. *PLOS Genetics* 12:e1006457. DOI: <https://doi.org/10.1371/journal.pgen.1006457>, PMID: 27911898
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genomescale protein function classification. *Bioinformatics* 30:1236–1240. DOI: <https://doi.org/10.1093/bioinformatics/btu031>, PMID: 24451626
- Kang J, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, Gemberling M, Goldman JA, Black BL, Poss KD. 2016. Modulation of tissue repair by regeneration enhancer elements. *Nature* 532:201–206. DOI: <https://doi.org/10.1038/nature17644>, PMID: 27049946
- Khan SF, Damerell V, Omar R, Du Toit M, Khan M, Maranyane HM, Mlaza M, Bleloch J, Bellis C, Sahm BDB, Peres J, ArulJothi KN, Prince S. 2020. The roles and regulation of TBX3 in development and disease. *Gene* 726:144223. DOI: <https://doi.org/10.1016/j.gene.2019.144223>, PMID: 31669645

- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12:357–360. DOI: <https://doi.org/10.1038/nmeth.3317>, PMID: 25751142
- Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nature Reviews. Genetics* 20:207–220. DOI: <https://doi.org/10.1038/s41576-018-0089-8>, PMID: 30675018
- Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J. 2017. *Drosophila* poised enhancers are generated during tissue patterning with the help of repression. *Genome Research* 27:64–74. DOI: <https://doi.org/10.1101/gr.209486.116>
- Labbé RM, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DDR, Ross EJ, Voisin V, Bader GD, Blencowe BJ, Pearson BJ. 2012. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* 30:1734–1745. DOI: <https://doi.org/10.1002/stem.1144>, PMID: 22696458
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25. DOI: <https://doi.org/10.1186/gb-2009-10-3-r25>, PMID: 19261174
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25:2078–2079. DOI: <https://doi.org/10.1093/bioinformatics/btp352>, PMID: 19505943
- Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. 2019. Identification of transcription factor binding sites using ATAC-seq. *Genome Biology* 20:45. DOI: <https://doi.org/10.1186/s13059-019-1642-2>, PMID: 30808370
- Lopez AY, Wang X, Xu M, Maheshwari A, Curry D, Lam S, Adesina AM, Noebels JL, Sun Q-Q, Cooper EC. 2017. Ankyrin-G isoform imbalance and interneuronopathy link epilepsy and bipolar disorder. *Molecular Psychiatry* 22:1464–1472. DOI: <https://doi.org/10.1038/mp.2016.233>, PMID: 27956739
- Lowe EK, Cuomo C, Voronov D, Arnone MI. 2019. Using ATAC-seq and RNA-seq to increase resolution in GRN connectivity. *Methods in Cell Biology* 151:115–126. DOI: <https://doi.org/10.1016/bs.mcb.2018.11.001>, PMID: 30948003
- Magner R. 2022. Picard. swb:1:rev:fc87aed37f657bc844c7271a91b190ac1a77a9ac. Software Heritage. <https://archive.softwareheritage.org/swb:1:dir:c4a976f952df723d9f815ff16d8f6de52d45a1ce;origin=https://github.com/broadinstitute/picard;visit=swb:1:snp:4ab926082574490562e2205c801f7e9ec2260463;anchor=swb:1:rev:fc87aed37f657bc844c7271a91b190ac1a77a9ac>
- Mihaylova Y, Abnave P, Kao D, Hughes S, Lai A, Jaber-Hijazi F, Kosaka N, Aboobaker AA. 2018. Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Nature Communications* 9:1–7. DOI: <https://doi.org/10.1038/s41467-018-06092-6>, PMID: 30194301
- Miraldi ER, Pokrovskii M, Watters A, Castro DM, De Veaux N, Hall JA, Lee J-Y, Ciofani M, Madar A, Carriero N, Littman DR, Bonneau R. 2019. Leveraging chromatin accessibility for transcriptional regulatory network inference in T helper 17 cells. *Genome Research* 29:449–463. DOI: <https://doi.org/10.1101/gr.238253.118>, PMID: 30696696
- Molina MD, Saló E, Cebrià F. 2007. The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Developmental Biology* 311:79–94. DOI: <https://doi.org/10.1016/j.ydbio.2007.08.019>, PMID: 17905225

- Murad R, Macias-Muñoz A, Wong A, Ma X, Mortazavi A. 2019. Integrative Analysis of Hydra Head Regeneration Reveals Activation of Distal Enhancer-like Elements. *bioRxiv*. DOI: <https://doi.org/10.1101/544049>
- Najafabadi HS, Garton M, Weirauch MT, Mnaimneh S, Yang A, Kim PM, Hughes TR. 2017. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biology* 18:1–15. DOI: <https://doi.org/10.1186/s13059-017-1287-y>, PMID: 28877740
- Neiro J. 2020. Oxplatys. swl:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25. Software Heritage. <https://archive.softwareheritage.org/swl:1:dir:d0b31ae353719430dbcd0785d1a62a2ce5297a6d;origin=https://github.com/jakke-neiro/Oxplatys;visit=swl:1:snp:c490a9c5c87bc69eb23a62fd05f80b6d5b201a51;anchor=swl:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25>
- Nelson AD, Jenkins PM. 2017. Axonal membranes and their domains: assembly and function of the axon initial segment and node of ranvier. *Frontiers in Cellular Neuroscience* 11:136. DOI: <https://doi.org/10.3389/fncel.2017.00136>, PMID: 28536506
- Nepusz T. 2022. Rigraph. swl:1:rev:eda5db6952d19d4043522ff1daf82ccff8e71d84. Software Heritage. <https://archive.softwareheritage.org/swl:1:dir:88c2167c0a67ce01671a7b4783a4e7282163d3e1;origin=https://github.com/igraph/igraph;visit=swl:1:snp:d5e6cdd440f706979d912e610cebecf1a09ea99;anchor=swl:1:rev:eda5db6952d19d4043522ff1daf82ccff8e71d84>
- Onal P, Grün D, Adamidi C, Rybak A, Solana J, Mastrobuoni G, Wang Y, Rahn H-P, Chen W, Kempa S, Ziebold U, Rajewsky N. 2012. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *The EMBO Journal* 31:2755–2769. DOI: <https://doi.org/10.1038/emboj.2012.110>, PMID: 22543868
- Pascual-Carreras E, Marín-Barba M, Castillo-Lara S, Coronel-Córdoba P, Magri MS, Wheeler GN, Abril JF, Gomez-Skarmeta JL, Saló E, Adell T. 2020. Genomic Analyses Reveal FoxG as an Upstream Regulator of Wnt1 Required for Posterior Identity Specification in Planarians. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.12.08.416008>
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14:417–419. DOI: <https://doi.org/10.1038/nmeth.4197>, PMID: 28263959
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33:290–295. DOI: <https://doi.org/10.1038/nbt.3122>, PMID: 25690850
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. *Nature Protocols* 11:1650–1667. DOI: <https://doi.org/10.1038/nprot.2016.095>, PMID: 27560171
- Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* 14:687. DOI: <https://doi.org/10.1038/nmeth.4324>
- Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glázar P, Obermayer B, Theis FJ, Kocks C, Rajewsky N. 2018. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* 360:eaq1723. DOI: <https://doi.org/10.1126/science.aq1723>, PMID: 29674432

- Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, Adey AC, Steemers FJ, Shendure J, Trapnell C. 2018. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular Cell* 71:858–871. DOI: <https://doi.org/10.1016/j.molcel.2018.06.044>, PMID: 30078726
- Pundhir S, Bagger FO, Lauridsen FB, Rapin N, Porse BT. 2016. Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic Acids Research* 44:4037–4051. DOI: <https://doi.org/10.1093/nar/gkw250>, PMID: 27095194
- Quinlan AR. 2014. BEDTools: the swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics* 47:11–12. DOI: <https://doi.org/10.1002/0471250953.bi1112s47>
- Rada-Iglesias A, Baipai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470:279–283. DOI: <https://doi.org/10.1038/nature09692>, PMID: 21160473
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* 44:W160–W165. DOI: <https://doi.org/10.1093/nar/gkw257>, PMID: 27079975
- Raz AA, Wurtzel O, Reddien PW. 2021. Planarian stem cells specify fate yet retain potency during the cell cycle. *Cell Stem Cell* 28:1307–1322. DOI: <https://doi.org/10.1016/j.stem.2021.03.021>, PMID: 33882291
- Reddien PW, Bermange AL, Kicza AM, Sánchez Alvarado A. 2007. BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. *Development* 134:4043–4051. DOI: <https://doi.org/10.1242/dev.007138>, PMID: 17942485
- Reddien PW. 2018. The cellular and molecular basis for planarian regeneration. *Cell* 175:327–345. DOI: <https://doi.org/10.1016/j.cell.2018.09.021>, PMID: 30290140
- Rink JC, Gurley KA, Elliott SA, Sánchez Alvarado A. 2009. Planarian HH signaling regulates regeneration polarity and links HH pathway evolution to cilia. *Science* 326:1406–1410. DOI: <https://doi.org/10.1126/science.1178712>, PMID: 19933103
- Rink JC. 2013. Stem cell systems and regeneration in planaria. *Development Genes and Evolution* 223:67–84. DOI: <https://doi.org/10.1007/s00427-012-0426-4>, PMID: 23138344
- Romero BT, Evans DJ, Aboobaker AA. 2012. Facs analysis of the planarian stem cell compartment as a tool to understand regenerative mechanisms. *Progenitor Cells* 916:167–179. DOI: https://doi.org/10.1007/978-1-61779-980-8_13
- Schoenfelder S, Sugar R, Dimond A, Javierre B-M, Armstrong H, Mifsud B, Dimitrova E, Matheson L, Tavares-Cadete F, Furlan-Magaril M, Segonds-Pichon A, Jurkowski W, Wingett SW, Tabbada K, Andrews S, Herman B, LeProust E, Osborne CS, Koseki H, Fraser P, et al. 2015. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nature Genetics* 47:1179–1186. DOI: <https://doi.org/10.1038/ng.3393>, PMID: 26323060
- Schwaiger M, Schönauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D, Technau U. 2014. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Research* 24:639–650. DOI: <https://doi.org/10.1101/gr.162529.113>, PMID: 24642862
- Scimone ML, Kravarik KM, Lapan SW, Reddien PW. 2014. Neoblast specialization in regeneration of the planarian *Schmidtea mediterranea*. *Stem Cell Reports* 3:339–352. DOI: <https://doi.org/10.1016/j.stemcr.2014.06.001>, PMID: 25254346

- Scimone ML, Cote LE, Reddien PW. 2017. Orthogonal muscle fibres have different instructive roles in planarian regeneration. *Nature* 551:623–628. DOI: <https://doi.org/10.1038/nature24660>, PMID: 29168507
- Sebé-Pedrós A, Ballaré C, Parra-Acero H, Chiva C, Tena JJ, Sabidó E, Gómez-Skarmeta JL, Di Croce L, Ruiz-Trillo I. 2016. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* 165:1224–1237. DOI: <https://doi.org/10.1016/j.cell.2016.03.034>, PMID: 27114036
- Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews. Genetics* 15:272–286. DOI: <https://doi.org/10.1038/nrg3682>, PMID: 24614317
- Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019. Stem cell differentiation trajectories in hydra resolved at single-cell resolution. *Science* 365:eaav9314. DOI: <https://doi.org/10.1126/science.aav9314>, PMID: 31346039
- Solana J, Kao D, Mihaylova Y, Jaber-Hijazi F, Malla S, Wilson R, Aboobaker A. 2012. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial rnaseq, RNA interference and irradiation approach. *Genome Biology* 13:R19. DOI: <https://doi.org/10.1186/gb-2012-13-3-r19>, PMID: 22439894
- Spicuglia S, Vanhille L. 2012. Chromatin signatures of active enhancers. *Nucleus* 3:126–131. DOI: <https://doi.org/10.4161/nucl.19232>, PMID: 22555596
- Stegmaier P, Kel AE, Wingender E. 2004. Systematic DNA-binding domain classification of transcription factors. *Genome Informatics* 15:276–286 PMID: 15706513.
- Stelman CR, Smith BM, Chandra B, Roberts-Galbraith RH. 2021. CBP/p300 homologs CBP2 and CBP3 play distinct roles in planarian stem cell function. *Developmental Biology* 473:130–143. DOI: <https://doi.org/10.1016/j.ydbio.2021.02.004>, PMID: 33607113
- Stower's Institute. 2016. RNA seq analysis in flatworm following 6, 12, or 18 day treatment with rnai against pax5, sox8, or zfp1, relative to unc22 to identify transcriptional targets.
- Stower's Institute. Swapna LS, Molinaro AM, Lindsay-Mosher N, Pearson BJ, Parkinson J. 2018. Comparative transcriptomic analyses and single-cell RNA sequencing of the freshwater planarian *Schmidtea mediterranea* identify major cell types and pathway conservation. *Genome Biology* 19:1–22. DOI: <https://doi.org/10.1186/s13059-018-1498-x>
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676. DOI: <https://doi.org/10.1016/j.cell.2006.07.024>, PMID: 16904174
- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872. DOI: <https://doi.org/10.1016/j.cell.2007.11.019>, PMID: 18035408
- Tewari AG, Owen JH, Petersen CP, Wagner DE, Reddien PW. 2019. A small set of conserved genes, including sp5 and hox, are activated by wnt signaling in the posterior of planarians and acoels. *PLOS Genetics* 15:e1008401. DOI: <https://doi.org/10.1371/journal.pgen.1008401>, PMID: 31626630
- Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, Biggin MD, Stamatoyannopoulos JA. 2011. Dynamic reprogramming of chromatin

- accessibility during *Drosophila* embryo development. *Genome Biology* 12:R43. DOI: <https://doi.org/10.1186/gb-2011-12-5-r43>, PMID: 21569360
- Tomoyasu Y, Halfon MS. 2020. How to study enhancers in non-traditional insect models. *The Journal of Experimental Biology* 223:jeb212241. DOI: <https://doi.org/10.1242/jeb.212241>, PMID: 32034049
- van den Boogaart KG, Tolosana-Delgado R. 2008. “Compositions”: a unified R package to analyze compositional data. *Comput Geosci* 34:320–338. DOI: <https://doi.org/10.1016/j.cageo.2006.11.017>
- van Wolfswinkel JC, Wagner DE, Reddien PW. 2014. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell* 15:326–339. DOI: <https://doi.org/10.1016/j.stem.2014.06.007>, PMID: 25017721
- Vásquez-Doorman C, Petersen CP. 2014. Zic-1 expression in planarian neoblasts after injury controls anterior pole regeneration. *PLOS Genetics* 10:e1004452. DOI: <https://doi.org/10.1371/journal.pgen.1004452>, PMID: 24992682
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858. DOI: <https://doi.org/10.1038/nature07730>, PMID: 19212405
- Vogg MC, Owlarn S, Pérez Rico YA, Xie J, Suzuki Y, Gentile L, Wu W, Bartscherer K. 2014. Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires zic and forkhead transcription factors. *Developmental Biology* 390:136–148. DOI: <https://doi.org/10.1016/j.ydbio.2014.03.016>, PMID: 24704339
- Wagner DE, Wang IE, Reddien PW. 2011. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* 332:811–816. DOI: <https://doi.org/10.1126/science.1203983>, PMID: 21566185
- Wang IE, Lapan SW, Scimone ML, Clandinin TR, Reddien PW. 2016. Hedgehog signaling regulates gene expression in planarian glia. *eLife* 5:e16996. DOI: <https://doi.org/10.7554/eLife.16996>, PMID: 27612382
- Wang W, Hu C-K, Zeng A, Alegre D, Hu D, Gotting K, Ortega Granillo A, Wang Y, Robb S, Schnittker R, Zhang S, Alegre D, Li H, Ross E, Zhang N, Brunet A, Sánchez Alvarado A. 2020. Changes in regeneration-responsive enhancers shape regenerative capacities in vertebrates. *Science* 369:eaaz3090. DOI: <https://doi.org/10.1126/science.aaz3090>, PMID: 32883834
- Wang LH, Aberin MAE, Wu S, Wang SP. 2021. The mll3/4 h3k4 methyltransferase complex in establishing an active enhancer landscape. *Biochemical Society Transactions* 49:1041–1054. DOI: <https://doi.org/10.1042/BST20191164>, PMID: 34156443
- Weidinger G, Thorpe CJ, Wuennenberg-Stapleton K, Ngai J, Moon RT. 2005. The sp1-related transcription factors sp5 and sp5-like act downstream of wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. *Current Biology* 15:489–500. DOI: <https://doi.org/10.1016/j.cub.2005.01.041>, PMID: 15797017
- West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. 2014. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nature Communications* 5:1–2. DOI: <https://doi.org/10.1038/ncomms5719>

- Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. 2008. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Research* 36:D88–D92. DOI: <https://doi.org/10.1093/nar/gkm964>
- Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, Zhao P, Xia Q, Ling E. 2016. Systematic identification and characterization of long non-coding rnas in the silkworm, *bombyx mori*. *PLOS ONE* 11:e0147147. DOI: <https://doi.org/10.1371/journal.pone.0147147>
- Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. *Genome Biology* 21:22. DOI: <https://doi.org/10.1186/s13059-020-1929-3>, PMID: 32014034
- Yang K, Kang J. 2019. Tissue regeneration enhancer elements: a way to unlock endogenous healing power. *Developmental Dynamics* 248:34–42. DOI: <https://doi.org/10.1002/dvdy.24676>, PMID: 30291668
- Yazawa S, Umesono Y, Hayashi T, Tarui H, Agata K. 2009. Planarian hedgehog/patched establishes anteriorposterior polarity by regulating Wnt signaling. *PNAS* 106:22329–22334. DOI: <https://doi.org/10.1073/pnas.0907464106>, PMID: 20018728
- Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics* 31:2382–2383. DOI: <https://doi.org/10.1093/bioinformatics/btv145>, PMID: 25765347
- Zerbino DR, Johnson N, Juettemann T, Wilder SP, Flicek P. 2014. WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* 30:1008–1009. DOI: <https://doi.org/10.1093/bioinformatics/btt737>, PMID: 24363377
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of chip-seq (MACS). *Genome Biology* 9:R137. DOI: <https://doi.org/10.1186/gb-2008-9-9-r137>, PMID: 18798982
- Zhu B, Zhang W, Zhang T, Liu B, Jiang J. 2015. Genome-wide prediction and validation of intergenic enhancers in arabidopsis using open chromatin signatures. *The Plant Cell* 27:2415–2426. DOI: <https://doi.org/10.1105/tpc.15.00537>, PMID: 26373455
- Zhu SJ, Pearson BJ. 2016. (Neo)blast from the past: new insights into planarian stem cell lineages. *Current Opinion in Genetics & Development* 40:74–80. DOI: <https://doi.org/10.1016/j.gde.2016.06.007>, PMID: 27379899

Chapter 3. Phylogenetic analysis of TALE homeobox genes in *Schmidtea mediterranea* reveals conservation and lineage-specific evolution

TALE homeobox genes are ancestral to eukaryotes and belong to the conserved molecular toolkit that underlies metazoan development and cell fate specification. However, TALE genes have diversified in spiralian, giving rise to a spiralian-specific clade (SPILE) alongside other divergent TALE genes. In early mollusc embryos, SPILE genes are expressed in blastomere-specific patterns and regulate cell specification. Planarians (Tricladida) are highly regenerative flatworms that belong to the spiralian superphylum, yet the presence or function of divergent TALE genes has not been studied in this group. To establish the TALE gene repertoire in the planarian *Schmidtea mediterranea*, we revisit the metazoan TALE phylogeny and discover both conserved and lineage-specific SPILE genes among 15 TALE genes in this species. Additionally, our phylogenetic analysis reconstructs a previously undescribed spiralian-specific clade, which we name LOPHILE (LOPHotrochozon taLE), and a platyhelminth-specific TALE clade, which we name PALE (PLatyhelmiht-specific taLE). Although spiralian-specific TALE genes have been primarily implicated in early development, we uncover a role for spiralian-specific *SPILE-Z1* in adult planarian biology and a role for the platyhelminth-specific *PALE* in regenerative patterning. Altogether, our findings reveal the importance of lineage-specific gene evolution in the biology of regenerative planarians.

3.1 Introduction

Morphological evolution has been proposed to be partly driven by alterations in cis-regulatory sequences that modify the expression patterns of a conserved set of developmental genes, referred to as a conserved “molecular toolkit” (Davidson and Erwin, 2006; Carroll, 2008). As a part of this toolkit, the conserved class of TALE homeobox genes predates metazoans and are involved in several conserved developmental trajectories in both vertebrates and invertebrates (Selleri et al., 2019; Schulte et al., 2019; Bobola and Sagerström, 2024). However, modern genomics has also uncovered many lineage-specific transcription factors in development, and the diversification of TALE genes in spiralian constitutes one of the largest known expansions of homeobox genes (Paps et al., 2015; Holland et al., 2017; Morino et al., 2017; Phuangphong et al., 2021; Morino et al., 2022; Morino and Yoshikawa, 2023). Spiralian comprise nearly half of metazoan phyla, and many spiralian share the conserved spiralian cleavage program of early development (Martín-Durán and Marlétaz, 2020). Interestingly, several genes in a spiralian-specific clade of TALE factors, the SPILE genes, are expressed in a blastomere-specific manner in both molluscs and annelids, regulating the activity of other transcription factors (Morino et al., 2017; Phuangphong et al., 2021; Morino and Yoshikawa, 2023). Therefore, diversification of TALE genes in spiralian has contributed to the evolution of fate specification in blastomeres during early spiralian development and has directed the evolution of the spiralian body plan (Morino et al., 2017; Morino and Yoshikawa, 2023).

Among spiralian are flatworms that include both parasitic and non-parasitic species. Previous phylogenetic studies have identified spiralian-specific genes in the parasitic flatworms *Echinococcus multilocularis*, *Echinococcus granulosus*, and *Schistosoma mansoni* (Paps et al., 2015; Morino et al., 2017). However, the presence of divergent TALE genes in the clade of highly regenerative planarians (Tricladida) has remained unexamined. Among planarians, the species *Schmidtea mediterranea* is capable of regenerating any missing body part, thanks to a population of adult somatic pluripotent stem cells and instructions from a body-wide positional information

system (Reddien et al., 2022). In planarian regeneration, the TALE factors *prep* and *pbx* are necessary for axial regeneration, while *meis-1* is involved in the regeneration of eyes (Felix and Aboobaker, 2010; Lapan and Reddien, 2012; Blassberg et al., 2013; Chen et al., 2013). Additionally, the expression patterns and phenotypes of some other TALE factors have been sporadically reported, along with recent studies on the involvement of *irx* genes in neuronal regeneration (van Wolfswinkel et al., 2014; Rodríguez-Esteban et al., 2015; King et al., 2024). In our previous work, we performed a comprehensive annotation of planarian transcription factors and identified several additional uncharacterised TALE factors in the genome of *S. mediterranea*. Consequently, this raises the possibility of lineage-specific TALE factors in planarians and necessitates a clarification of the phylogeny of planarian TALE factors. Given their potential for regulating regenerative events this may pinpoint to new lineage specific gene evolution contributing to regeneration.

Here, we revise the metazoan phylogeny of TALE factors by incorporating publicly available sequences from additional spiralian phyla. As earlier studies, we find support for a spiralian-specific SPILE clade; however, denser taxon sampling is still required to corroborate its status as a monophyletic clade. Additionally, we reconstruct another spiralian-specific clade of TALE factors as a sister branch to the MEINOX clade (LOPHILE, LOPHotochozon taLE), and a platyhelminth-specific TALE clade (PALE, PLatyhelmiht-specific taLE). In *S. mediterranea*, we find two *pbx* genes, four *irx* genes, a single *prep* gene, and three *meis* genes along with divergent TALE genes in the SPILE clade and a platyhelminth-specific PALE gene. Gene expression analysis throughout development reveals that TALE genes are expressed in both early development and during organogenesis. Intriguingly, we expand the functional repertoire of spiralian-specific TALE genes and find evidence for roles in both adult planarian physiology and regenerative patterning.

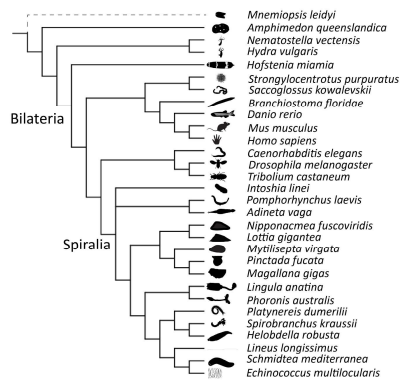
3.2 Results

3.2.1 *S. mediterranea* possesses 15 TALE genes from both conserved and divergent families

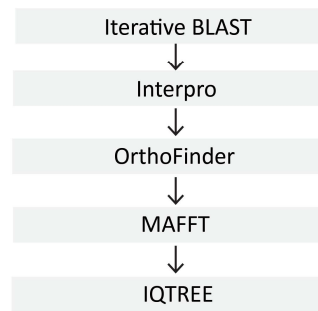
To find the orthologues of TALE genes in *S. mediterranea*, we reconstructed a phylogeny of TALE genes based on sequences from 30 metazoan species representative of different phyla (Figure 3.1A, Table S3.1). Earlier phylogenetic studies on spiralian-specific genes have been based on sequences from molluscs, annelids, platyhelminths, and brachiopods (Paps et al., 2015; Morino et al., 2017; Barton-Owen et al., 2018). We included these species in our analysis to guide comparisons. Moreover, we integrated sequences from additional spiralian phyla with representative genomes that had not been included in earlier phylogenies, namely Acanthocephala (*Pomphorhynchus laevis*) (Mauer et al., 2020), Phoronida (*Phoronis australis*) (Luo et al., 2018), Orthonectida (*Intoshia linei*) (Mikhailov et al., 2016), and Nemertea (*Lineus longissimus*) (Kwiatkowski et al., 2021) (Figure 3.1A, Table S3.1). We used reciprocal BLAST to recover an initial set of homologues, verified the presence of a homeodomain with InterProScan, and utilised orthogroups in Orthofinder to enhance our set of orthologues into the final collection (Figure 3.1B, Methods). The recovery of TALE genes was further validated by the presence of the conserved atypical homeodomain, which contains the insertion of three amino acids (Figure 3.1C). We used maximum likelihood inference implemented in

Figure 3.1 Discovery of metazoan TALE genes. (A) Species used in the construction of the phylogenetic tree. Phylogenetic relationships are based on ctenophores as sister to other animals and spiralian phylogenies (Laumer et al., 2015; Marlétaz et al., 2019; Schultz et al., 2023). **(B)** The computational pipeline for phylogenetic inference. TALE sequences were discovered through a reciprocal best-blast hit approach; the presence of a homeodomain was validated using Interproscan, and the presence of orthologues was further assessed with Orthofinder. The protein sequences were aligned with MAFFT, and the trees were constructed using IQTREE. **(C)** Multiple sequence alignment of TALE genes reveals the conservation of the homeodomain (HD) and its atypical insertion of three amino acids (the Three Amino acids Loop Extension) (subregion 1, PYP).

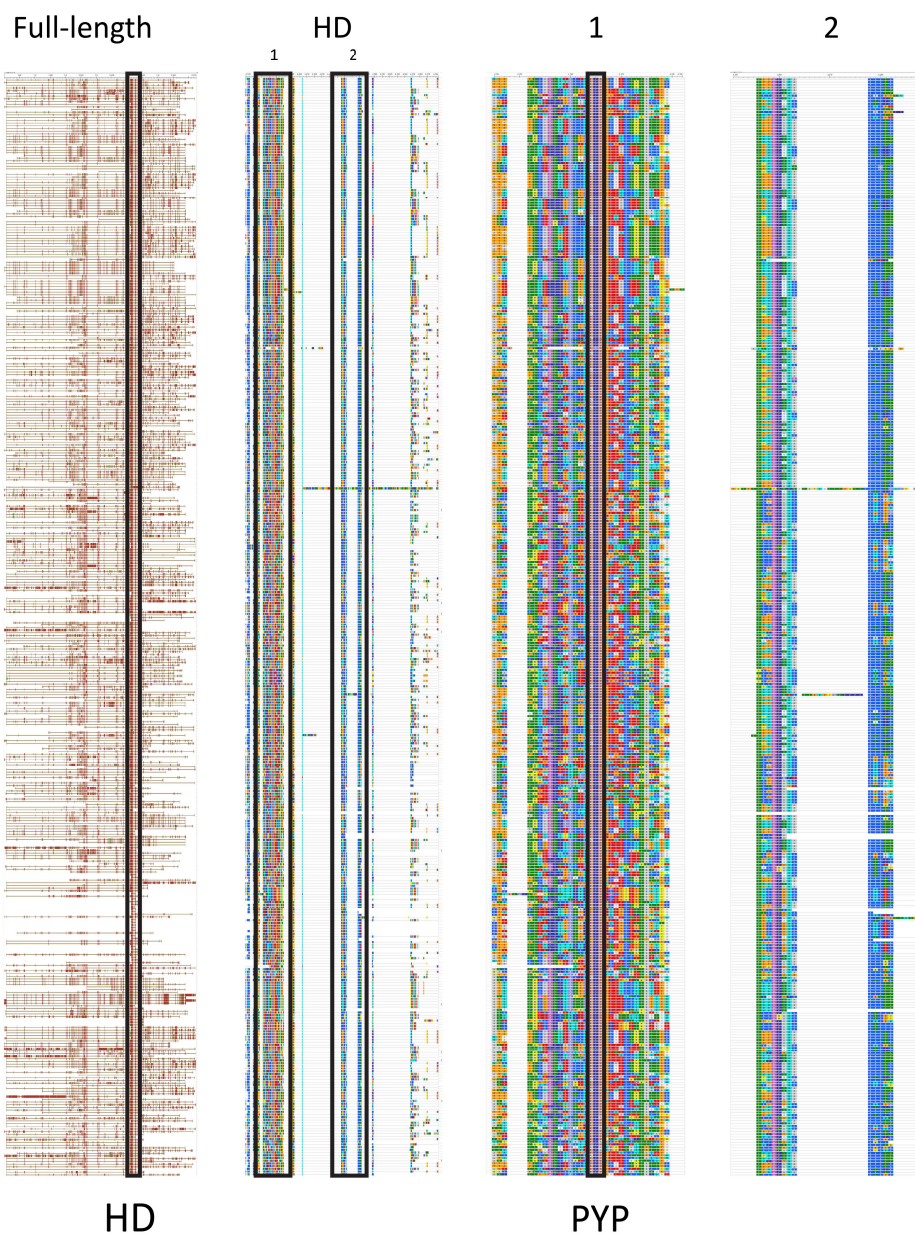
A



B



C



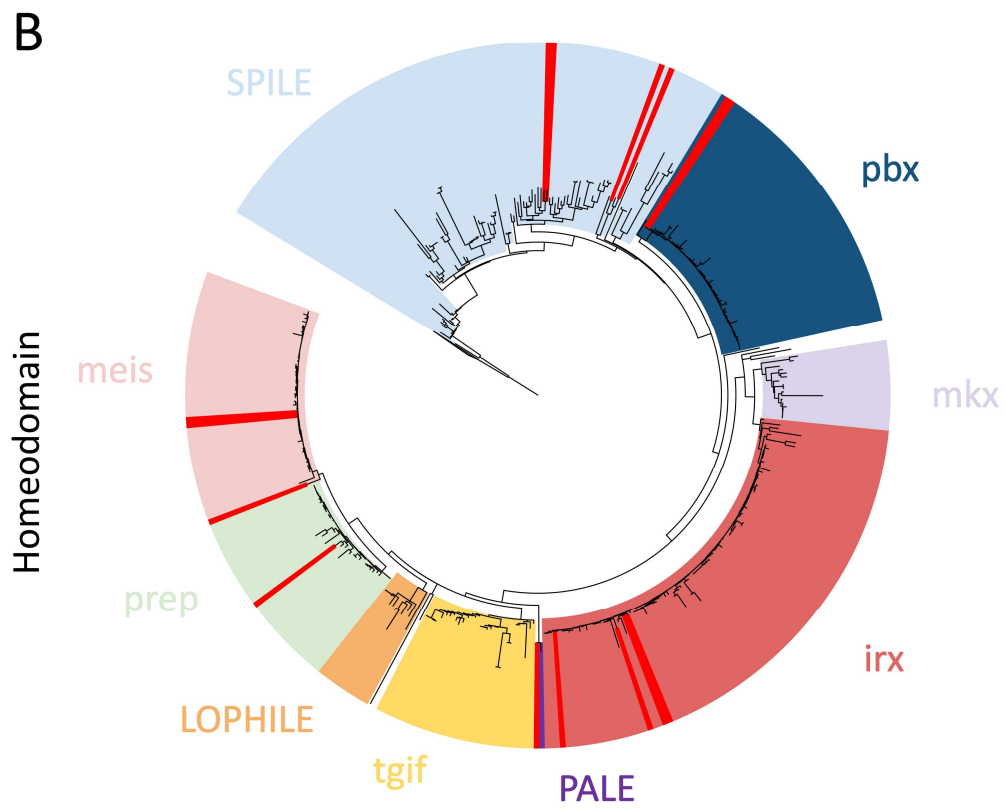
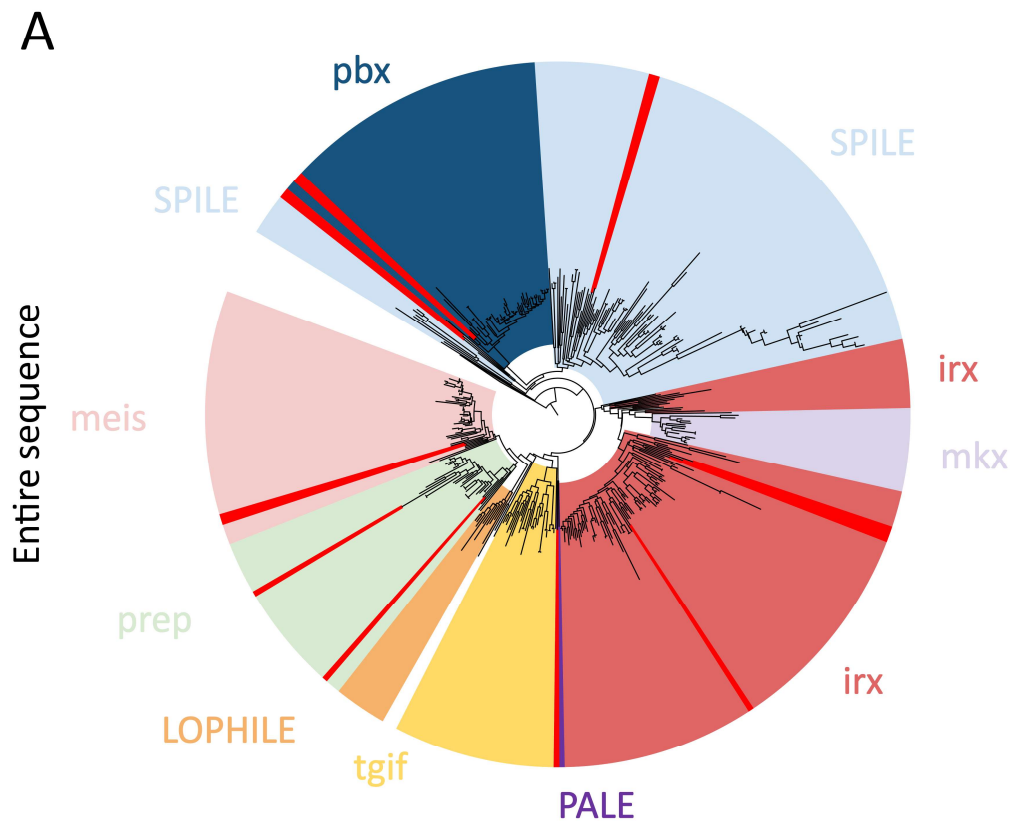


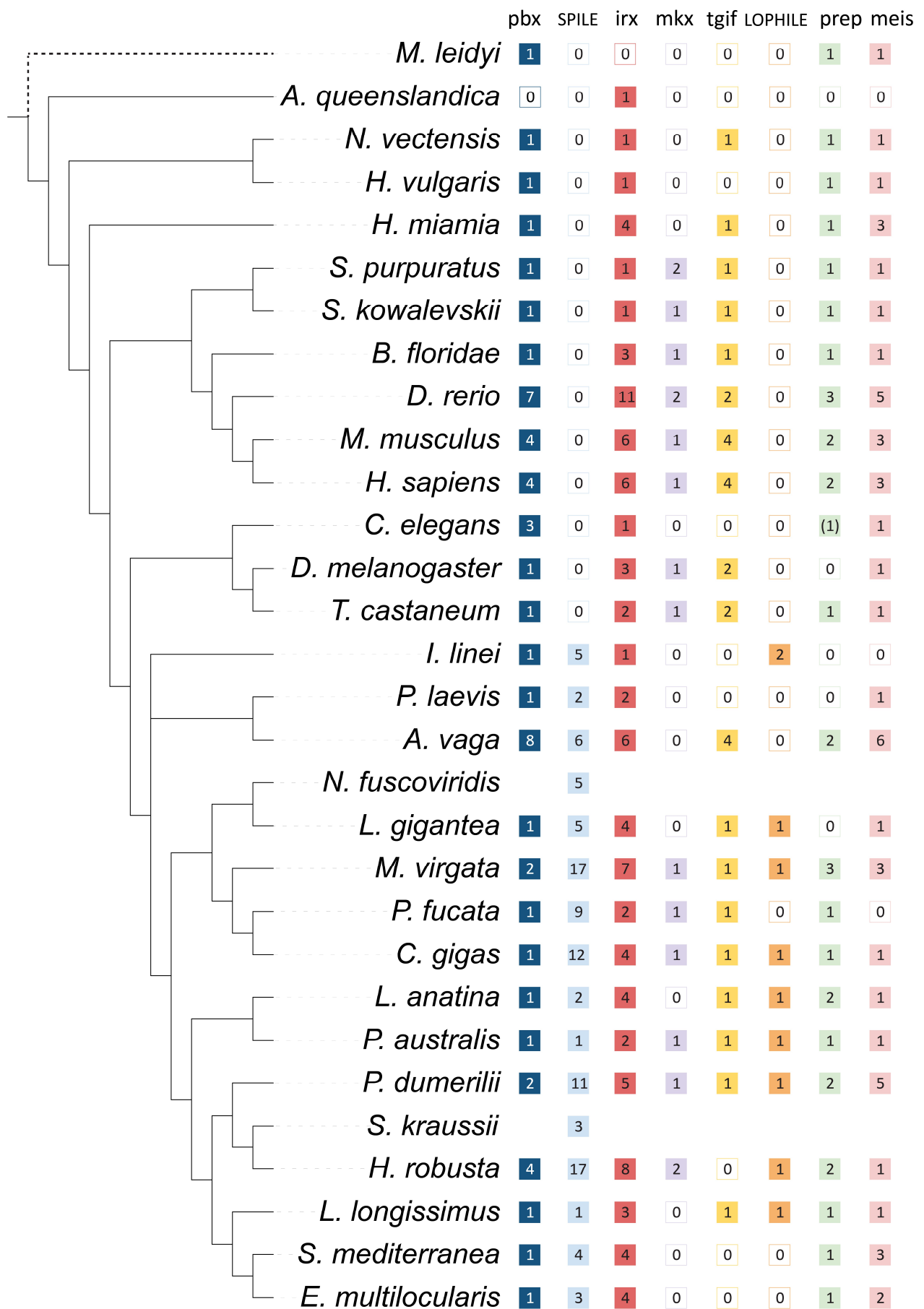
Figure 3.2 Phylogeny of metazoan TALE genes. A maximum likelihood tree constructed based on the entire amino acid sequences **(A)** or solely the homeodomain **(B)**. The Pbx clade is coloured in blue, the Irx clade in red, the Mlx clade in violet, the Tgif clade in yellow, the Prep clade in olive green, and the Meis clade in light red. Additionally, the spiralian-specific clade SPILE is coloured in light blue, the spiralian-specific LOPHILE clade is coloured in orange, and the platyhelminth-specific clade PALE is coloured in purple. Sequences of *S. mediterranea* are marked with red. Human Hox genes have been used as an outgroup for rooting the trees.

IQTREE to construct two complementary trees (Minh et al., 2020): one based on the entire amino acid sequence (Figure 3.2A) and the other solely on the homeodomain (Figure 3.2B).

As expected, our phylogenetic analysis recovered the monophyly of conserved subfamilies with high support: *Pbx* (bootstrap value 100 in Figure 3.2A, 75 in Figure 3.2B), *Irx* and *Mlx* (bootstrap value 91 in Figure 3.2A, 68 in Figure 3.2B), *Tgif* (bootstrap value 99 in Figure 3.2A, 90 in Figure 3.2B), *Prep* (bootstrap value 75 in Figure 3.2A, 92 in Figure 3.2B), and *Meis* (bootstrap value 94 in Figure 3.2A, 94 in Figure 3.2B) (Mukherjee and Bürglin, 2007; Bürglin and Affolter, 2016; Morino et al., 2017). Furthermore, our phylogeny based on entire sequences reproduces *Pbx* as the most basal clade (Figure 3.2A), and the relationship between *Irx* and *Mlx*, and the common ancestry of *Meis*, *Prep*, and *Tgif* is replicated in both trees (Figure 3.2A, B) (Mukherjee and Bürglin, 2007; Bürglin and Affolter, 2016). In our phylogeny, the *Pbx* class is present in *Nematostella vectensis* and *Mnemiopsis leidyi* as previously reported (Figure 3.3) (Mukherjee and Bürglin, 2007; Ryan et al., 2010). This is also in agreement with earlier studies indicating that the *Pbx* class predates the origin of metazoans (Joo et al., 2018). An *Irx* gene is present in the genome of *N. vectensis*, but our analysis did not recover an *Mlx* gene in basal metazoans, supporting that the *Mlx* family emerged in bilaterians (Figure 3.2) (Mukherjee and Bürglin, 2007; Kerner et al., 2009). Lastly, both *Prep* and *Meis* are present in the genome of *M. leidyi* (Figure 3.3), suggesting that an ancestral MEINOX gene duplicated in the metazoan ancestor, as *Prep* has not been found in opisthokonts (Joo et al., 2018).

Altogether, we found 15 TALE genes in *S. mediterranea* (Figure 3.3). We identified two copies of the *pbx* gene family in the genome of *S. mediterranea*, which appears to be the result of a recent duplication (Figure 3.3, S3.1). Both flatworms, *E. multilocularis* and *S. mediterranea*, have four *irx* genes but have lost the *mlx* family (Figure 3.3, S3.2, S3.3). Similarly, the *tgif* family seems to have been lost in flatworms, while being present in basal metazoans and other spiralian, in accordance with earlier phylogenies (Figure 3.3, S3.3) (Mukherjee and Bürglin, 2007). In *S. mediterranea*, we found a single canonical *prep* and two canonical *meis* genes, but an additional gene was classified as either a divergent *prep* (Figure 3.2A), a divergent *meis* (Figure 3.2B), or a divergent *prep* in a separate phylogeny based on sequences solely from the MEINOX family (Figure S3.4).

Figure 3.3 TALE gene repertoires of 30 metazoans. Coloured boxes indicate the presence of an orthologue in TALE subclasses. Membership is defined based on the homeodomain-based phylogenetic analysis (Figure 3.2A). The phylogeny is identical to that in Figure 3.1A. A number inside a box indicates copies per family and species. Abbreviations for species names correspond to full names in Figure 3.1A. *C. elegans* has a divergent homeodomain-less *prep* gene, *psa-3*, which was excluded from our analysis; this is indicated with brackets in the column.



3.2.2 Analysis of the SPILE clade reveals the presence of SPILE-E and SPILE-Z orthologues in *S. mediterranea*

As earlier phylogenies, we found several spiralian TALE genes that do not belong to any of the conserved classes (Figure 3.2A, B). In their phylogenetic analysis, Morino et al. (2017) were able to reconstruct a monophyletic spiralian-specific clade named SPILE, with a sufficiently high bootstrap value; however, not all divergent TALE genes belonged to this clade. Other phylogenetic analyses have also identified spiralian-specific clades (Paps et al., 2015; Barton-Owen et al., 2018). In our analysis, we found spiralian-specific genes diverging from the canonical Pbx clade in both our phylogeny based on the entire sequence (Figure 3.2A) and one based solely the homeodomain (Figure 3.2B). However, these spiralian-specific TALE genes constituted a paraphyletic clade in the phylogeny restricted to the homeodomain (Figure 3.2B), whereas our phylogenetic tree based on entire protein sequences reproduced a spiralian-specific clade with high support (Figure 3.2B, bootstrap value 95). To resolve the phylogeny further, we extracted Pbx and associated spiralian-specific sequences and were able to reproduce a monophyletic clade that includes genes from the previously reported SPILE clade (Figure 3.3) (Morino et al., 2017; Phuangphong et al., 2021; Morino, 2022). Furthermore, our phylogenetic analysis was able to reproduce the clades SPILE-A/C (bootstrap value 87), SPILE-B (bootstrap value 84), SPILE-D (bootstrap value 100), and SPILE-E (bootstrap value 90), which have been constructed based on mollusc SPILE genes and named after the defining sequences of *Nipponacmea fuscoviridis* (Morino et al., 2017; Phuangphong et al., 2021; Morino, 2022).

In the SPILE phylogeny, we observe that the SPILE-A/C clade includes previously reported members from the mollusc species *N. fuscoviridis*, *Crassostrea gigas*, *Lottia gigantea*, and *Pinctada fucata* (Figure 3.4) (Morino et al., 2022). Moreover, the SPILE-A/C clade contains genes from the annelids *Platynereis dumerilii* and *Helobdella robusta*, but no SPILE genes from other spiralian phyla (Figure 3.4). Given the high copy number in *H. robusta*, some of these SPILE-A/C genes could also be pseudogenes (Figure 3.4). As previously reported, the SPILE-B clade is characterised by the presence of two homeodomains (Figure S3.10). In *S. mediterranea*, we identified two orthologues of the SPILE-E gene found in *N. fuscoviridis*, which we designate as *SPILE-E1* (SMESG000005966.1) and *SPILE-E2* (SMESG000008497.1) (Figure 3.4). Furthermore, we found two planarian TALE genes in a clade with the previously characterised *SPILE-Z* gene in the annelid *Spirobranchus kraussii* (bootstrap value 71). Following the nomenclature devised by Morino et al. (2022), we named this clade SPILE-Z and the genes in *S. mediterranea* as *SPILE-Z1* (SMESG000049852.1) and *SPILE-Z2* (SMESG000014778.1).

Interestingly, we also found non-SPILE spiralian-specific clades in the metazoan phylogeny of TALE genes (Figure 3.2A, B). Basal to the branch containing Tgif, Meis, and Prep is a platyhelminth-specific clade (bootstrap value 100 in both Figure 3.2A and 3.2B) that includes a single sequence from *S. mediterranea* (SMESG000071247.1) and a single sequence from *E. multilocularis* (EmuJ 000903300.1). Following the nomenclature for SPILE, we designated this clade PALE (PLatyhelminth-specific tALE). Furthermore, we were able to construct another well-supported spiralian-specific clade that is a basal branch to the canonical Meis and Prep clades (bootstrap value 100 in Figure 3.2A, 96 in Figure 3.2B). This clade comprises 10 sequences from spiralian species, including the previously described *TALE2* in *L. gigantea* and *TALE2* in *C. gigas* that were placed as basal to the Tgif clade in the phylogeny of Morino et al. (2017) (Figure S3.3). Thus, we conclude that there is a reproducible spiralian-specific clade of non-SPILE divergent TALE factors associated with the MEINOX clade, which we designate as LOPHILE (LOPhotrochozoa-specific tALE factors).

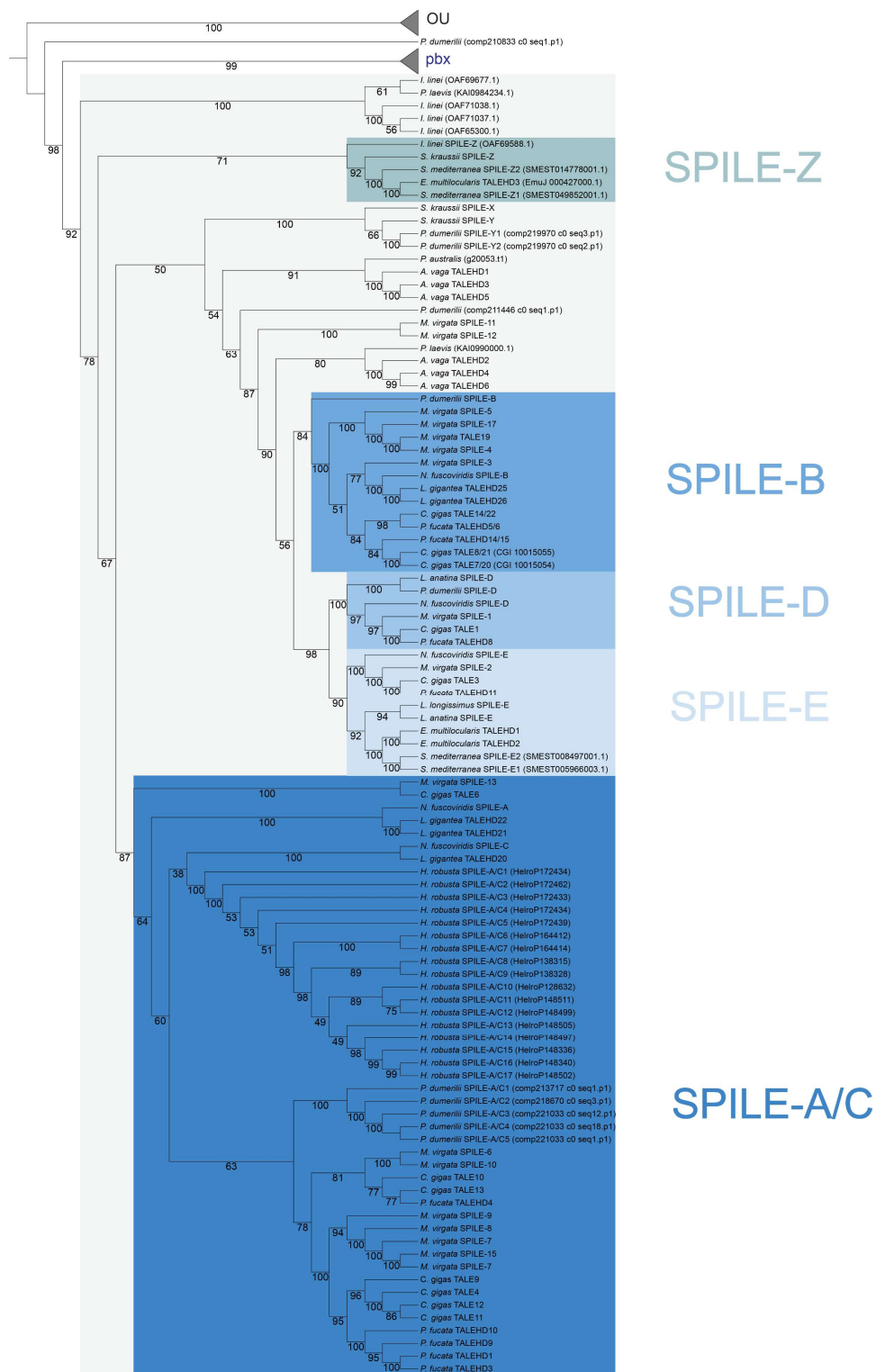


Figure 3.4 Phylogenetic tree of the spiralian-specific SPILE clade. The phylogeny was constructed using entire protein sequences and inferred by maximum likelihood (IQTREE). The numbers at nodes indicate bootstrap values (ultrafast bootstrap approximations) (Minh et al., 2020; Hoang et al., 2018). The clades SPILE-A/C, SPILE-B, SPILE-D, and SPILE-E are labelled (Morino, 2022). Established gene names are shown for *C. gigas*, *M. virgata*, *P. fucata*, *N. fuscoviridis*, and *L. gigantea* (Morino et al., 2017; Phuangphong et al., 2021; Morino, 2022), while names are proposed for other taxa. Human Hox genes were used as an outgroup for rooting the tree, and the Pbx clade is represented as a collapsed branch.

3.2.3 Lineage-specific TALE factors are widely present in platyhelminth species

To further understand the evolution of TALE genes in *S. mediterranea*, we conducted an additional phylogenetic analysis based on platyhelminth sequences. In our analysis, we took advantage of the recently published collection of 50 high-quality de novo assembled transcriptomes along with sequences from *E. multilocularis* and *S. mansoni* (Vila-Farré et al., 2023) (Figure 3.5A, B). As expected, the phylogeny reconstructed the canonical classes Pbx (bootstrap value 100), Irx (bootstrap value 100), Prep (bootstrap value 99.7), and Meis (bootstrap value 100) along with the divergent SPILE-Z (bootstrap value 99.2), SPILE-E (bootstrap value 62.6), and PALE (bootstrap value 100) (Figure 3.5A, B). In the SPILE-Z clade, we observe that orthologues of *SPILE-Z1* are present in the genomes of *E. multilocularis* and *Schistosoma mansoni*, while the *SPILE-Z2* gene appears to have duplicated before the emergence of the Triclad (Figure S3.12). In the SPILE-E clade, we note that *S. mediterranea* shares orthologues of *SPILE-E1* and *SPILE-E2* with other planarian species, but the timing of the duplication is difficult to interpret in this sample of taxa (Figure S3.13). Furthermore, we observe that the *PALE* gene is present in the parasitic flatworms and in most analysed planarian species (Figure 3.5A,B). The planarian phylogeny allowed us to properly classify the divergent MEINOX as a Meis, but given its high divergence, we designate it as *meis-like* (Figure 3.5A,B, S3.16).

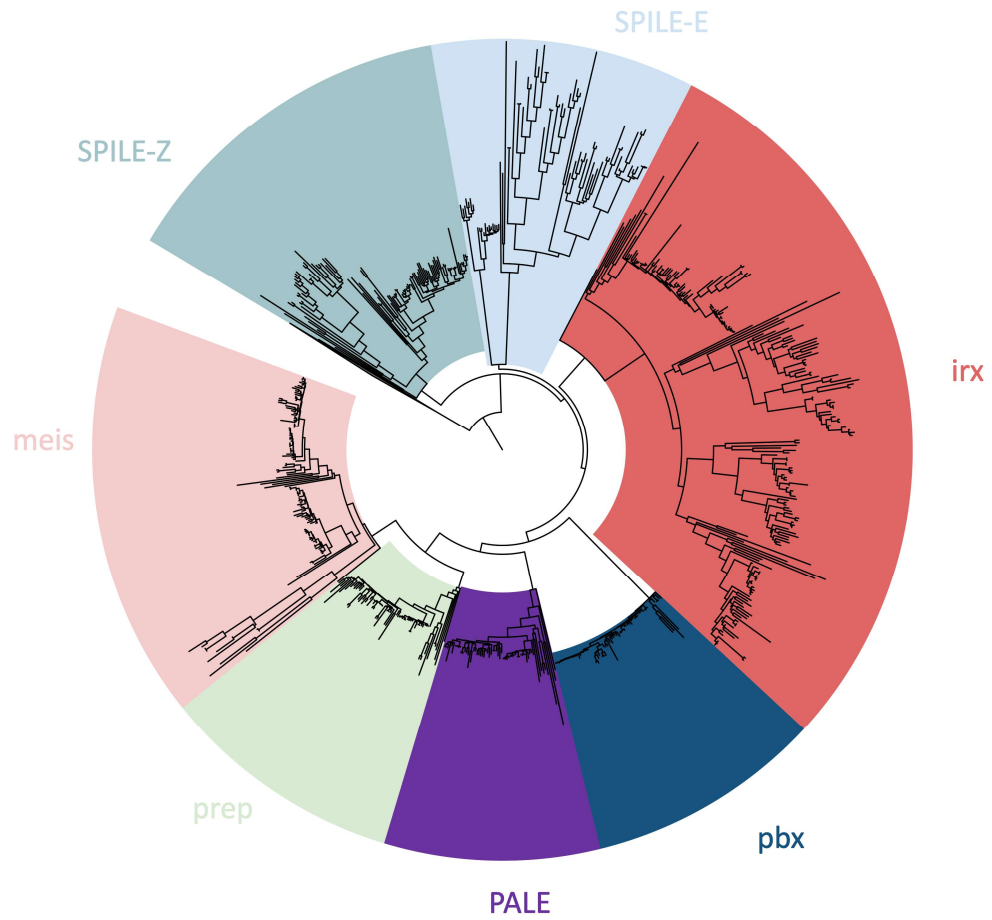
As a number of TALE genes have been sporadically characterised in *S. mediterranea*, we revised and expanded our previous review of transcription factors to rigorously assign members of the TALE gene repertoire to those reported in the literature (Neiro et al., 2022). Recently, three *irx* genes named *IRX1*, *IRX2*, and *IRX6*, have been characterised in a transcription factor atlas of neoblast differentiation (King et al., 2024). In our phylogeny, sequence alignment of reported transcripts with BLAST indicate that *IRX1* corresponds to *irx-1*, *IRX2* corresponds to *irx-4*, and *IRX6* corresponds to *irx-2* (Table 3.1). Thus, *irx-3* has not yet been characterised (Table 3.1). As the planarian *irx* genes are not orthologues of mammalian *Irx* genes, we name the planarian *irx* genes by arbitrarily numbering them (Figure S3.2, 3.4, S3.14).

Previously, a single *pbx* gene has been implicated in both anterior and posterior regeneration (Blassberg et al., 2013; Chen et al., 2013). We assigned both original clones (KC353351 by Blassberg et al. (2013) and H.110.1c by Chen et al. (2013)) to the sequence of *pbx-1*. Additionally, van Wolfswinkel et al. (2014) had reported the expression pattern of an additional *pbx* gene, “*pbx4*”. However, we were able to definitely assign the original sequence to the *SPILE-Z1* gene (BLAST alignment 1610/1610 identities, 0 gaps) (Table 3.1).

In the MEINOX family, the single *prep* gene can be straightforwardly assigned to the previously characterised *prep* gene (Felix and Aboobaker, 2010). In contrast, different *meis* genes have been reported in the literature. Lapan and Reddien (2012) identified a *meis* gene (initially named “*meis*”) involved in eye differentiation and morphogenesis; the reported sequence unambiguously aligns with *meis-1*. Scimone et al. (2014) also report this gene. van Wolfswinkel et al. (2014) reported the expression pattern of another *meis* gene, named *meis-2*, and the reported sequence corresponds to the *meis-2* gene in our phylogeny. Lastly, Rodríguez-Esteban et al. (2015) characterised a gene termed *meis-like* that is expressed in e.g. the nervous system and neoblasts, and is required for anterior regeneration. However, the reported GenBank sequence KM981934.1 clearly aligns to the lineage-specific *PALE* gene rather than a Meis family member (Table 3.1).

Figure 3.5 Figure 4. Phylogenetic analysis of planarian TALE factors. (A). A maximum likelihood phylogeny of planarian TALE factors. Human Hox genes were used as an outgroup for rooting the tree. **(B)** TALE gene repertoire in planarian species. The number in the column indicates the number of gene copies in the gene family. The species phylogeny is inferred using Orthofinder. Taxon names are referenced from Vila-Farré et al. (2023).

A



B

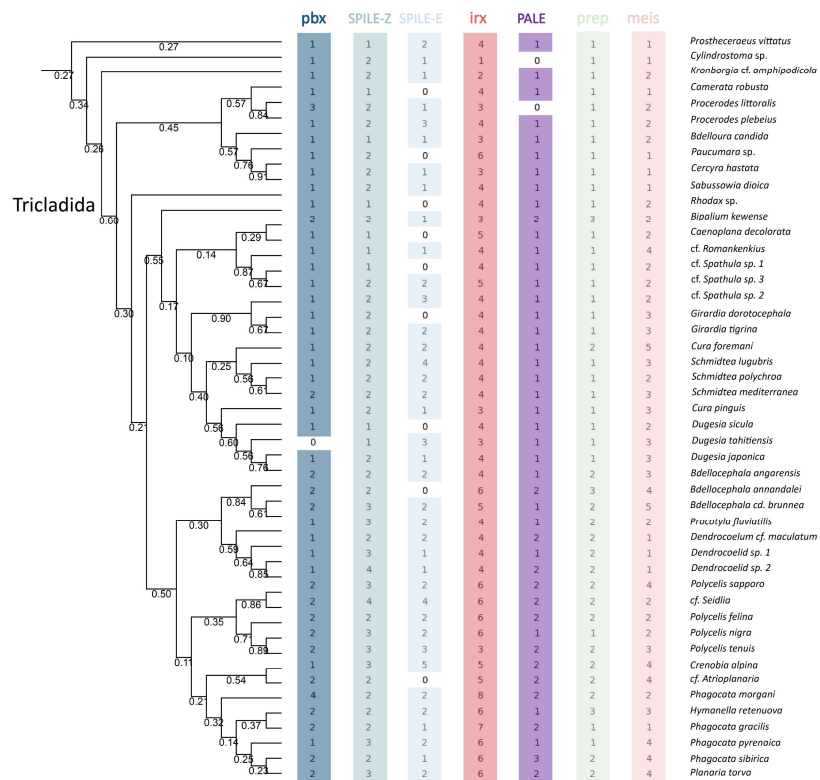


Table 3.1 The TALE gene repertoire in *S. mediterranea*.

Name	Old names	PlanMine ID	Neiro ID	References
<i>irx-1</i>	<i>IRX1, irx-1</i>	SMESG000022899.1	MSTRG.7450	Neiro et al. (2022), King et al. (2024)
<i>irx-2</i>	<i>IRX6, irx6-2</i>	SMESG000026100.1	MSTRG.572	Neiro et al. (2022), King et al. (2024)
<i>irx-3</i>	<i>irx6-1</i>	SMESG000015279.1	MSTRG.5557	Neiro et al. (2022)
<i>irx-4</i>	<i>IRX2, irx6-3</i>	SMESG000061184.1	MSTRG.18286	Neiro et al. (2022), King et al. (2024)
<i>PALE</i>	<i>meis-4, meis-like</i>	SMESG000071247.1	MSTRG.20883	Rodríguez-Esteban et al. (2015), Neiro et al. (2022)
<i>prep</i>	<i>prep</i>	SMESG000013307.1	MSTRG.4755	Felix and Aboobaker (2010), Neiro et al. (2022)
<i>meis-like</i>	<i>pbx-4</i>	SMESG000062973.1	MSTRG.18649	Neiro et al. (2022)
<i>meis-1</i>	<i>meis-1, meis</i>	SMESG000006530.1	MSTRG.1897	Lapan and Reddien (2012), Scimone et al. (2014), Neiro et al. (2022)
<i>meis-2</i>	<i>meis-2</i>	SMESG000078684.1	MSTRG.23026	van Wolfswinkel et al. (2014), Neiro et al. (2022)
<i>pbx-1</i>	<i>pbx-1, pbx</i>	SMESG000001913.1	MSTRG.1510	Blassberg et al. (2013), Chen et al. (2013), Neiro et al. (2022)
<i>pbx-2</i>	<i>pbx-2</i>	SMESG000022232.1	MSTRG.7854	Neiro et al. (2022)
<i>SPILE-Z1</i>	<i>meis-3, pbx4</i>	SMESG000049852.1	MSTRG.15193	van Wolfswinkel et al. (2014), Neiro et al. (2022)
<i>SPILE-Z2</i>	<i>Pbx-5</i>	SMESG000014778.1	MSTRG.4467	Neiro et al. (2022)
<i>SPILE-E1</i>	<i>Pbx-3</i>	SMESG000005966.1	MSTRG.1741	Neiro et al. (2022)
<i>SPILE-E2</i>	-	SMESG000008497.1	MSTRG.3285	-

3.2.4 TALE factors of *S. mediterranea* contain both conserved and novel protein domains

To begin to understand the putative functional role of TALE factors in *S. mediterranea*, we investigated the genic structures, the domain architectures, and the three-dimensional protein structures as predicted by ColabFold (Jumper et al., 2021; Mirdita et al., 2022). In *S. mediterranea*, the exon number varies from 3 (*irx-1*, *SPILE-E2*) to 13 (*meis-2*), while the length of the proteins varies from 217 residues (*pbx-2*) to 773 residues (*PALE*) (Figure 3.6).

Pbx-1 contains a full homeodomain, the conserved PBC-A and PBC-B domains that mediate dimerisation (Longobardi et al., 2014), and nuclear localisation signals NLS1 and NLS2 (Figure 3.6, S3.5). In contrast, pbx-2 is truncated and contains a homeodomain with NLS1 flanking it, but it lacks the PBC-A and PBC-B domains (Figure 3.6, S3.5). The genomic sequence of pbx-2 appears to have a splice site mutation that results in a shorter variant of the Pbx protein (Figure S3.19). Similarly, in the MEINOX clade the canonical meis-1, meis-2, and prep contain the MH-A and MH-B domains involved in dimerisation, while the divergent meis-like is truncated and includes only the homeodomain (Figure 3.6, S3.7). Nonetheless, all Meis genes contain the conserved NLS sequence flanking the homeodomain.

Strikingly, we find that both the PALE and SPILE-Z1 proteins contain zinc finger motifs in the protein sequences (Figure 3.6). PALE contains three putative zinc fingers (ZNF1 residues 15 to 38, ZNF2 residues 48 to 71, ZNF3 residues 160 to 182). Although Interproscan detects the zinc fingers with low confidence (ZNF1 evalue 13, ZNF2 evalue 13, ZNF3 evalue 110), the putative zinc finger regions produce three-dimensional predictions with relatively high confidence, as indicated by pIDDT in ColabFold (Figure 3.6). Moreover, the second zinc finger is independently predicted by ZFPrinceton (Persikov et al., 2009; Persikov and Singh, 2014) (Figure S3.20). Similarly, SPILE-Z1 includes two putative zinc finger motifs (ZNF1 residues 57 to 79, ZNF2 residues 93 to 116), as detected by Interproscan, ColabFold, and ZFPrinceton (Figure 3.6, S3.20). Although the binding sites predicted for the zinc fingers are uncertain, the physical association between the zinc fingers and the homeodomain may influence DNA binding properties.

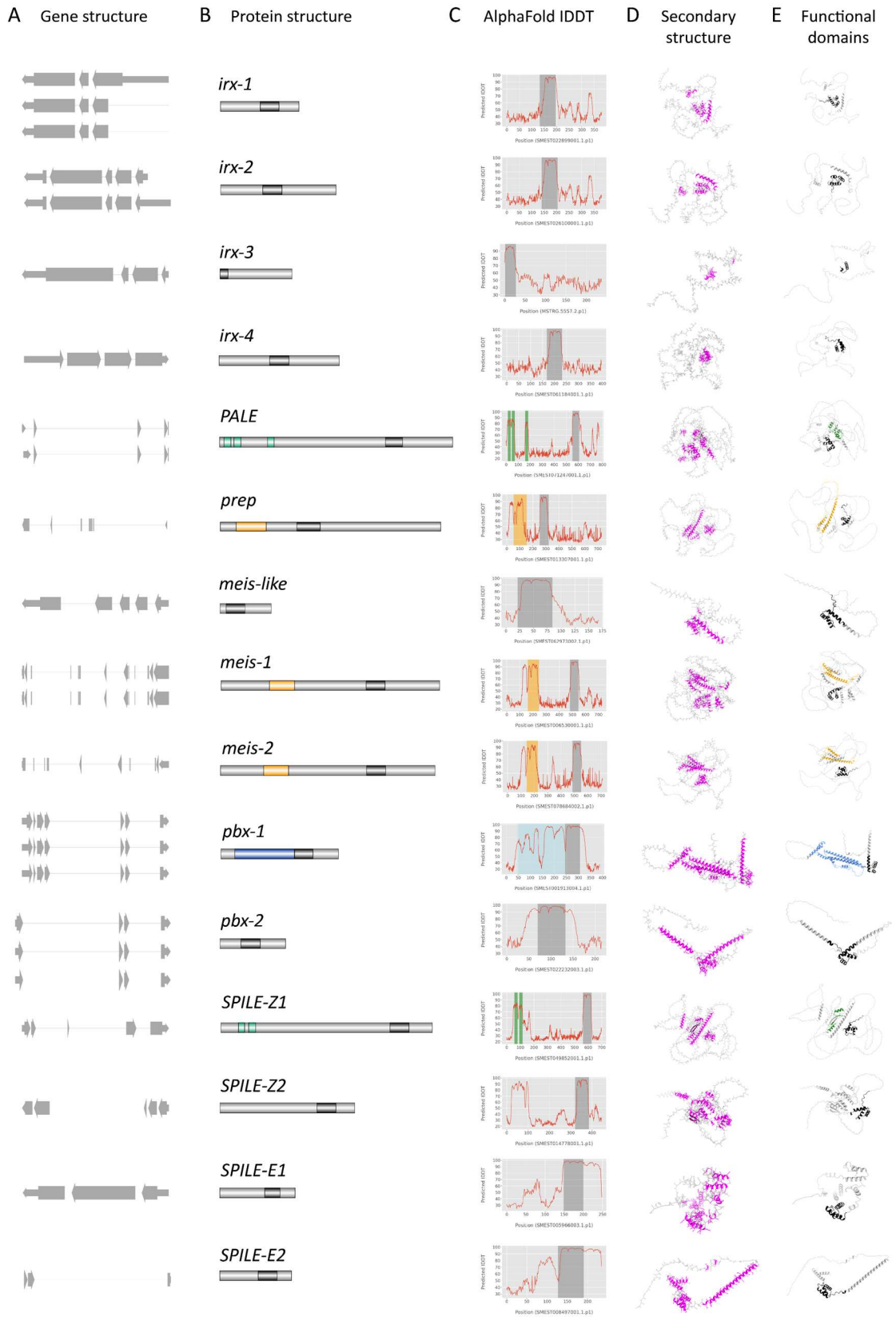


Figure 3.6 The structure and functional domains of TALE genes in *S. mediterranea*. (A) Gene structures including exon-intron boundaries. Grey boxes indicate coding regions and the interspersed intronic regions are shown as grey lines. (B) Schematic representation of the protein domain architecture in TALE genes. The homeodomain is shown in black, MEINOX domains (MH-A and MH-B) in orange, PBC domains (PBC-A and PBC-B) in blue, and zinc fingers in green. (C) The pLLDT metric (predicted local distance difference test, a pre-residue confidence metric) along the length of protein sequence as predicted by ColabFold (Mirdita et al., 2022). The metric ranges from 0 to 100, and values above 85 are considered predictions of high certainty. The homeodomain is coloured in black, the MEINOX domain (MH-A and MH-B) is coloured in yellow, the PBC domain (PBC-A and PBC-B) is coloured in blue, and zinc fingers are coloured in green. (D) Three-dimensional protein foldings as predicted by ColabFold. Secondary structures including coils and helices are coloured in purple. (E) Homeodomains (black), MEINOX domains (yellow), PBC domains (blue), and zinc fingers (green) labelled in the three-dimensional structure.

3.2.5 TALE factors are expressed during both early development and organogenesis

As earlier studies of spiralian-specific gene expression indicated that several of these genes are expressed during early development, we sought to understand the expression patterns of TALE factors in *S. mediterranea*. For this purpose, we analysed publicly available data from embryonic development (Davies et al., 2017), adult worms, expression values along the AP axis (Stückemann et al., 2017), and scRNA-seq atlases (Fincher et al., 2018; Plass et al., 2018).

The study of embryonic development in *S. mediterranea* has been relatively overlooked compared to research on regeneration in adult asexual worms. This oversight is largely due to several factors complicating embryonic manipulation: internal fertilisation, ectolecithal eggs, i.e. the yolk is separated from egg cells into distinct yolk cells, a dark and compact eggshell, and variability in the timing of embryonic development (Davies et al., 2017). Furthermore, planarians do not undergo stereotypic spiralian cleavage but display an evolutionarily divergent cleavage pattern termed dispersed cleavage or blastomere anarchy (Cardona et al., 2005; Martín-Durán et al., 2012). In dispersed cleavage, the blastomeres detach from one another and migrate freely within the yolk during early development before forming the transient yolk-feeding embryo composed of the primary epidermis, the embryonic pharynx, and an embryonic intestine (Martín-Durán et al., 2012). After feeding yolk, organogenesis begins and the embryo gradually flattens, adopts the shape and body plan of a flatworm, and hatches from the egg shell (Martín-Durán et al., 2012). Nonetheless, we took advantage of a stage-specific transcriptomic dataset in *S. mediterranea*, which includes samples from yolk cells in unfertilised embryos, dispersely cleaving and yolk-feeding embryos (Stages 2 to 5), and embryos undergoing organogenesis and maturation (Stages 6 to 8).

In our analysis of expression patterns in developing *S. mediterranea*, we observe that the expression of TALE genes clusters into two main categories: TALE genes with enriched expression during early development (*irx-1*, *PALE*, *SPILE-Z2*, *SPILE-E1*, and *SPILE-E2*), and TALE genes that are expressed from stage 5 and 6 onwards, reflecting the onset of organogenesis (*irx-2*, *irx-3*, *irx-4*, *prep*, *meis-like*, *meis-1*, *meis-2*, *pbx-1*, *pbx-2*, and *SPILE-Z1*) (Figure 3.7A). Within the early embryonic expression cluster, the strongest expression was observed in yolk cells, while the expression of genes in the organogenetic cluster peaked at stages 5, 6, and 7 (Figure 3.7A).

Interestingly, although *irx-1* is expressed in and required for the formation of *ZAN6+* parenchymal cells in the adult worm (King et al. 2024), the developmental expression profile suggests that *irx-1* could also play a pleiotropic role in early development (Figure 3.7A). In contrast, *irx-2*, *irx-3*, and *irx-4* are part of the organogenetic cluster and expressed from stage 5 onwards (Figure 3.7A). As *irx-2* is expressed in and

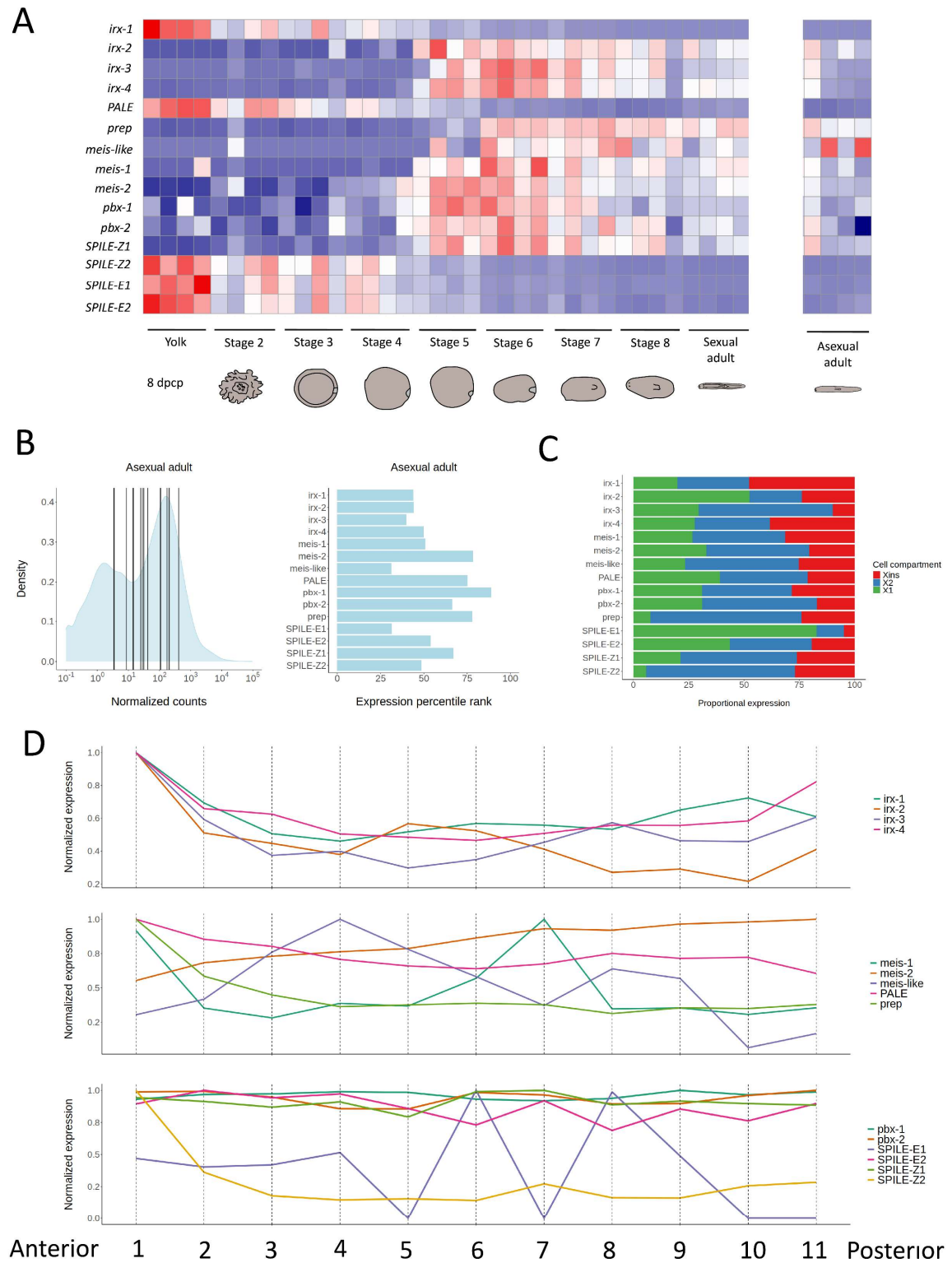
required for the formation of *npp-18*⁺ neurons during adult regeneration (King et al., 2024), this may reflect a recapitulation of a neuronal specification program also observed during embryonic organogenesis. Similarly, scRNA-seq atlases (Fincher et al., 2018; Plass et al., 2018) reveal that *irx-3* is expressed mainly in neural progenitors and ChAT neurons, while *irx-4* is expressed in a neoblast subpopulation and *spp-11*⁺ neurons. Furthermore, RNAi of *irx-4* abolishes dopaminergic neurons (King et al., 2024). The neuronal involvement of *Ir*x genes is further reflected in head-specific expression along the AP axis of the worm (Figure 3.7B).

The expression of the platyhelminth-specific *PAL*E gene is enriched in yolk cells and early development (Figure 3.7A). Furthermore, *PAL*E is expressed more or less equally in all FACS cell compartments (X1, X2, and Xins), with a slight bias towards the head in the adult worm (Figure 3.7B, C, D). scRNA-seq atlases reveal that *PAL*E is expressed broadly in different cell types, including neoblast subpopulations, neuronal progenitors and neuronal subtypes, as well as in the muscle lineage and *cathepsin*⁺ cells (Fincher et al., 2018; Plass et al., 2018).

The planarian *prep* gene has been well characterised and is known to be involved in the anterior regeneration of adult planarians (Felix and Aboobaker, 2010; Scimone et al., 2014a). Our expression analysis further suggest that *prep* is also involved in organogenesis during embryonic development (Figure 3.7A). Interestingly, the expression of *prep* begins at stage 6 rather than at stage 5, as is the case for the majority of genes in the organogenetic expression cluster, suggesting that *prep* is not involved in the initial organogenetic events during embryogenesis. Furthermore, the eye-associated *meis-1* and its paralogue *meis-2* are expressed from stage 6 onwards, while the uncharacterised *meis-like* is expressed from stage 6 (Figure 3.7A). Along the AP axis, *meis-2* displays a slight posterior bias in expression, whereas *meis-like* exhibits a shifting expression pattern, possibly due to its low absolute expression levels (Figure 3.7B,D). In addition, scRNA-seq atlases suggest that *meis-2* is ubiquitously expressed and present in neoblast subpopulations, muscles, epidermal cells, neuronal subpopulations, and the pharynx (Fincher et al., 2018; Plass et al., 2018). Lastly, the lowly expressed *meis-like* transcript is primarily present in epidermal cell types (Fincher et al., 2018; Plass et al., 2018).

Planarian *pbx-1* plays a role in axial regeneration in adults (Blassberg et al., 2013; Chen et al., 2013), and the developmental expression pattern indicates that *pbx-1* expression begins at stage 5, suggesting it may also be involved in embryonic patterning (Figure 3.7A). The uncharacterised *SPILE-Z1* is similarly expressed from stage 5 onwards, and is expressed ubiquitously along the AP axis in the asexual adults (Figure 3.7A,D). In the scRNA-seq atlas, *SPILE-Z1* is expressed mainly in *aquaporin*⁺ cells, *ldlrr-1*⁺ parenchymal cells, different parenchymal subpopulations, muscles, and protonephridia (Fincher et al., 2018; Plass et al., 2018). In contrast, its paralogue *SPILE-Z2* is expressed in yolk cells and during early development, and its expression is biased towards the anterior in adult worms (Figure 3.7A, D). In scRNA-seq datasets, *SPILE-Z2* is lowly expressed in secretory cells, the pharynx cell progenitors, and GABA neurons. Lastly, members of the *SPILE-E* clade, *SPILE-E1* and *SPILE-E2*, are expressed in yolk cells and throughout early development, and expressed ubiquitously along the AP axis (Figure 3.7 A,D).

Figure 3.7 Expression profiling of TALE genes in *S. mediterranea*. (A). Expression of TALE genes in planarian embryos (Davies et al. 2017). The heatmap shows standardised gene expression values (z-scores) across developmental stages for each gene. The values of different genes are not at the same scale. 8 dpcp = 8 days post capsule deposition. (B). Percentile rank for absolute expression value (normalised counts) of TALE factors in relation to the distribution of all expressed genes in an asexual adult worm. (C). Proportional expression values of TALE genes in the X1, X2, and Xins cell compartments. (D). The expression of TALE genes along the anterior-posterior axis (Stückemann et al., 2017). Expression values have been scaled so that the maximal expression value equals one.



3.2.6 Functional characterisation of the platyhelminth-specific *PALE* and spiralian-specific *SPILE-Z1* uncovers roles in regenerative patterning and adult biology

To understand the functional role of TALE genes in planarians, we performed a preliminary RNA interference (RNAi) screen. Recently, the neuronal involvement of Irx factors has been studied by King et al. (2024) in a scRNA-seq atlas of planarian transcription factors. Therefore, we focussed our efforts on knocking down genes of the canonical Pbx, Meis, and Prep families, along with the lineage-specific factors in the PALE and SPILE clades.

As anticipated, knockdown of *meis-1* replicated earlier findings and revealed a defect in eye regeneration (Lapan and Reddien, 2012). Previous knockdowns had reduced the area of rhabdomeric photoreceptor neuron aggregates, while we observed a visible reduction in eye size even under brightfield conditions (Figure 3.8B, C). Knockdown of *meis-2* resulted in a posterior regeneration defect and eyeless worms, and we continued to pursue this finding in greater detail (see Chapter 4). Interestingly, the knockdown of the MH-A- and MH-B-less *meis-like* gave rise to a cyclopic worm, similar to what has been observed after *prep*(RNAi) (Figure 3.8D). However, as the number of replicates is low (1 worm out of 5), further knockdown experiments are required to corroborate this finding.

Intriguingly, the knockdown of the platyhelminth-specific factor *PALE* produced worms with a “snouty” appearance in the stronger phenotype, or a blunted head tip and a widened overall morphology of the head in the milder phenotype (Figure 3.8E). Therefore, our knockdown experiments reveal that *PALE* plays a role in patterning the anterior region, despite our analysis indicating that *PALE* expression is enriched during early development (Figure 3.7A).

Furthermore, knockdown of *SPILE-Z1* caused a surprisingly severe phenotype (Figure 3.8F). In the standard ten-day injection regime, amputated fragments began to degrade as early as 4 days post-amputation (dpa), with continued degradation over time (Figure 3.8F). Additionally, some worms developed lesions, lost epidermal integrity, lysed or disintegrated (Figure 3.8E). Given this extreme phenotype, we asked whether it was the result of a disruption in neoblast function, as the strongest planarian phenotypes are generally observed following knockdown of chromatin remodelers and other cell cycle markers (Solana et al., 2012). However, although the results are difficult to interpret due to overall tissue disintegration, it appears that neoblasts are still present in the decaying fragment (Figure 3.8H). Since our expression analysis shows that *SPILE-Z1* is expressed in *aquaporin*⁺ parenchymal cells, we hypothesise that the phenotype following knockdown of *SPILE-Z1* may stem from dysfunctional osmoregulation, leading to compromised epidermal integrity. In contrast, knockdown of the paralogue *SPILE-Z2* did not lead to a noticeable phenotype upon visual inspection (Figure 3.8I).

Lastly, we did not observe a phenotype following knockdown of *SPILE-E1* (Figure 3.8I), which is consistent with a developmental expression profile biased towards early embryonic stages (Figure 3.7A). Therefore, further experiments in planarian embryos may reveal functions for the SPILE-E gene family in planarians.

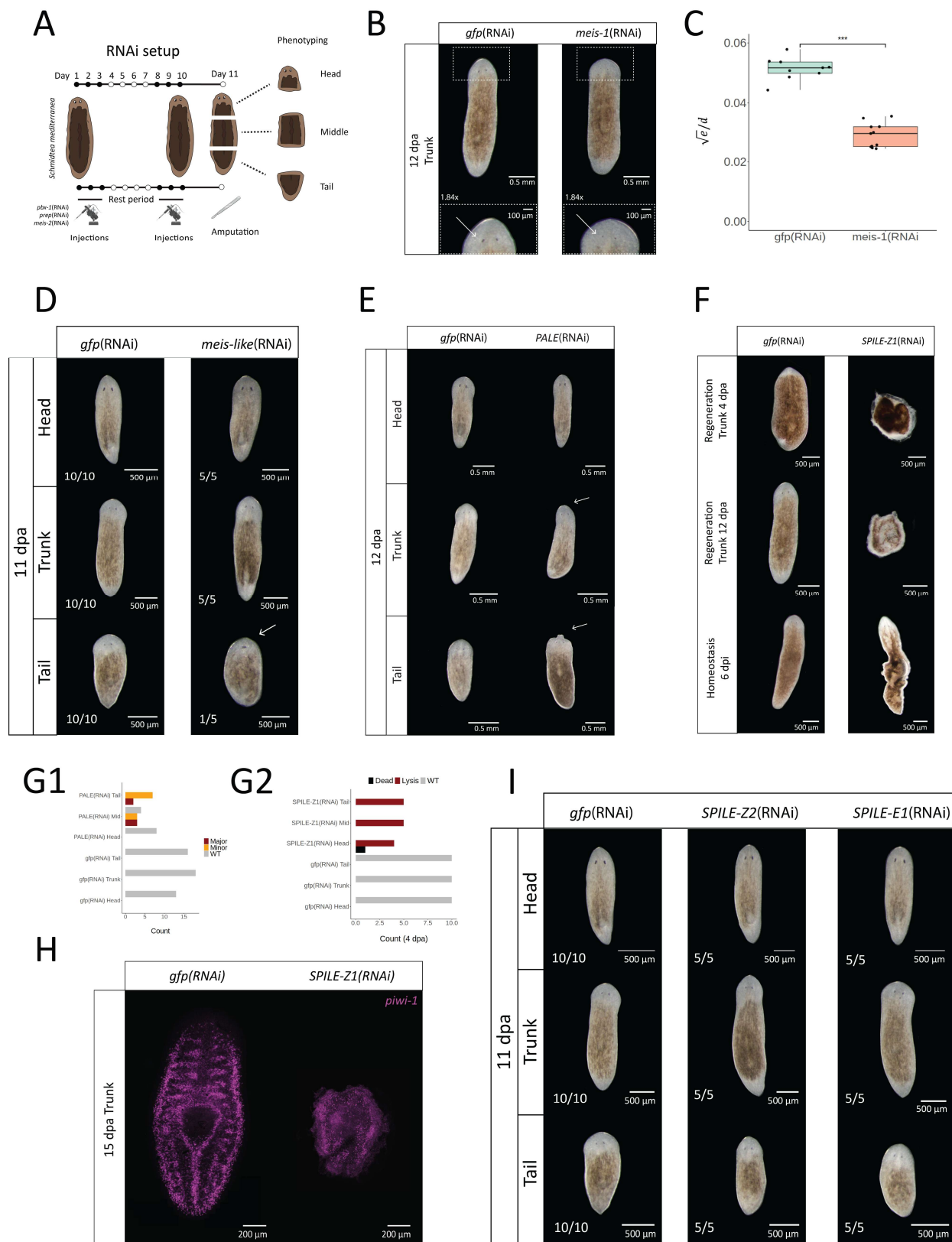


Figure 3.8 Functional characterisation of TALE genes in *S. mediterranea*. (A). Experimental setup for a standard ten-day RNAi schedule. Worms were injected at days 1, 2, 3, rested for four days, and injected again at days 8,9, and 10. At day 11, worms were amputated into head, trunk, and tail fragments. (B). Trunk fragments of regenerating *meis-1*(RNAi) worms at 12 days post amputation (12 dpa). The head region is magnified (1.84x) in the lower panel. (C). Eye size in proportion to interocular distance in control worms and after knockdown of *meis-1*. (D). Head, trunk, and tail fragments at 11 dpa after knockdown of *meis-like*. (E). Head, trunk, and tail fragments at 12 dpa after knockdown of *PALE*. (F). Knockdown phenotype of *SPILE-Z1*(RNAi). The upper panel represents trunk fragments at 4dpa, the middle panel represents trunk fragments at 12 dpa, and the lower panel the most severe phenotype of lysing worms six days after the start of injections. (G1). The number of the milder phenotype (widened head) and the stronger phenotype (“snouty” head) after two independent rounds of knocking down *PALE*. (G2). The number of degrading and dead worms at 4 dpa after knockdown of *SPILE-Z1*. (H). Fluorescence in situ hybridisation of showing neoblast distribution (*pimi-1*) after knockdown of *SPILE-Z1*. The images are maximum intensity projections. (I). Head, trunk, and tail fragments after knockdown of *SPILE-Z1* and *SPILE-E1*.

3.3 Discussion

In this study, we revised the metazoan phylogeny of TALE factors, identified spiralian-specific TALE factors in additional spiralian phyla, and uncovered both conserved and lineage-specific TALE factors in the genome of *S. mediterranea*. TALE homeobox genes represent an ancestral class of transcription factors predating metazoans, yet they have diversified extensively within the spiralian clade. We find support for a monophyletic spiralian-specific SPILE clade, which includes four genes in *S. mediterranea*: *SPILE-Z1*, *SPILE-Z2*, *SPILE-E1*, and *SPILE-E2*. Moreover, our phylogenetic analysis groups previously orphan spiralian TALE factors and establishes an additional spiralian-specific clade, LOPHILE, along with a platyhelminth-specific clade, PALE.

In earlier phylogenies, representatives of the SPILE-E clade have been found in multiple spiralian phyla (Paps et al., 2015; Morino et al., 2017; Barton-Owen et al., 2018), and our phylogenetic analysis similarly suggests that SPILE-E genes are present in molluscs, platyhelminths, brachiopods, and nemerteans. In molluscs, members of the SPILE-E clade have been found to be maternally deposited, expressed broadly in the early embryo until the 32-cell stage, and required for quartet specification (Morino et al., 2017; Phuangphong et al., 2021; Morino and Yoshikawa, 2023). In our expression analysis, we found that *SPILE-E1* and *SPILE-E2* are expressed during early development, and preliminary knockdown of *SPILE-E1* did not yield a visible phenotype in planarian regeneration. Therefore, the expression pattern suggests that the involvement of *SPILE-E* genes in early development is conserved in planarians. However, as early development in planarians does not follow the spiralian cleavage program but rather undergoes dispersed cleavage (Martín-Durán et al., 2012), this may indicate that planarian SPILE-E genes are not involved in blastomere fate segregation. On the other hand, as SPILE-E appears to be the only SPILE clade present in the planarian genome alongside the SPILE-Z clade, this suggests that the function of SPILE-E has been transformed from the spiralian cleavage program to meet the constraints of dispersed cleavage. Further studies on planarian early development would shed more light on this matter; however, early planarian development has been notoriously inaccessible for experimental manipulation (Martín-Durán et al., 2012).

Intriguingly, the planarian SPILE-Z clade appears to have diverged from roles of early development and attained a physiological role in adults. The closest studied orthologue, *SPILE-Z* in the annelid *S. kraussii*, is expressed in macromeres at the 8-cell stage, and is later restricted to a specific subset of blastomeres at the 16- and 32-cell stages (Morino et al., 2017). Our expression analysis indicates that the planarian *SPILE-Z2*

is expressed during early development and may play a role in blastomere specification, while *SPILE-Z1* is expressed later during organogenesis; adult worms rapidly lysed following knockdown of this gene. The diversified function could be a result of the acquisition of zinc finger motifs; further research on the role of SPILE-Z genes in other flatworms would reveal the evolution of the putative physiological function of *SPILE-Z1*. Furthermore, the platyhelminth-specific PALE clade displays a patterning phenotype during regeneration, although expression is biased towards early development, raising the possibility of additional pleiotropic developmental programs in the early planarian embryo regulated by this gene. Likewise, zinc fingers in the PALE sequence could modulate DNA binding activity, constituting a novel class of animal zinc finger containing-homeobox genes alongside e.g. *Zbx* genes (Nail et al., 2020).

All in all, this study has uncovered both conserved and dynamic molecular evolution of TALE genes in planarians. The diversification of TALE genes constitutes one of the largest cases of rapidly-evolving homeobox genes in metazoans (Holland et al., 2017; Morino et al., 2017). Spiralian make up almost half of metazoan phyla, yet our understanding of their development is based on a handful of species. Further investigation into the diversification of TALE genes in planarians and other spiralian will shed light on the evolution of animal body plans.

3.4 Methods

3.4.1 Phylogenetic tree construction

We gathered protein sequences from both available proteomes and transcriptome (Table S3.1). TALE sequences for well-characterised model species, including *H. sapiens*, *M. musculus*, *D. rerio*, *D. melanogaster*, *C. elegans* were obtained directly from Uniprot. Proteomes were directly obtained for *M. leidyi*, *T. castaneum*, *B. floridae*, *L. anatina*, *H. robusta*, *S. purpuratus*, *S. kowalevskii*, *A. queenslandica*, *N. vectensis*, *H. miamia*, *L. longissimus*, *P. laevis*, and *P. australis*, while the transcriptomes of *P. dumerilii* and *M. virgata* were translated by using Transdecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>). The proteome of *S. mediterranea* was generated in Chapter 2 (Neiro et al., 2022). Reciprocal BLAST using human TALE sequences as query sequences was employed to discover protein sequences for TALE factors in these species (Moreno-Hagelsieb and Latimer, 2008). TALE sequences for *P. fucata*, *C. gigas*, *L. gigantea*, *E. multilocularis*, *A. vava*, *N. fuscoviridis*, and *M. virgata* were obtained using information from previous phylogenies (Morino et al., 2017; Phuangphong et al., 2021). The presence of a homeodomain was validated by using InterProScan, and homeodomain-less sequences were filtered out (Jones et al., 2014). Lastly, Orthofinder (Emms and Kelly, 2019) was used to find orthogroups in the proteomes, and any genes belonging to orthogroups that included TALE sequences identified in the first reciprocal BLAST search but were missing from the set of TALE genes were extracted for a second reciprocal BLAST search. Sequences identified as TALE genes in the reciprocal BLAST search were added to the final set TALE sequences.

Protein sequences were aligned using MAFFT with the L-INS-I option (Katoh et al., 2019). Homeodomains were extracted using Bedtools (Quinlan et al., 2014) and a custom Python script. The full sequences or homeodomains were used as input for phylogenetic analysis using IQ-TREE (Minh et al., 2020). The phylogenetic tree was inferred by using ModelFinder, SH-aLRT test, and ultrafast bootstrap with 1,000 replicates (Kalyaanamoorthy et al., 2017; Hoang et al., 2018). R packages seqinr, ape, msa, treeio, and ggtree were used to extract sequences for clade-specific phylogenies (Charif and Lobry, 2007; Bodenhofer et al., 2015; Yu et al., 2017; Paradis and Schliep, 2019; Wang et al., 2020). The trees were visualised using the Interactive Tree of Life software (Letunic and Bork, 2024). The NCBI Multiple Sequence Alignment Viewer (MSA) (<https://www.ncbi.nlm.nih.gov/tools/msaviewer/>) was used to visualise protein alignments.

3.4.2 Functional domain analysis

Functional domains were queried using InterProScan (Jones et al., 2014) and domain structures were visualised using IBS 1.0.3 (Liu et al., 2015). The 3D structures of TALE genes were predicted using ColabFold (v1.5.2) (Jumper et al., 2021; Mirdita et al., 2022) and visualised with ChimeraX v1.8rc202405230136 (Pettersen et al., 2021).

3.4.3 Expression analysis

Bulk RNA-seq samples from different stages of planarian development (Davies et al., 2017) and different positions along the AP axis (Stückemann et al., 2017) were downloaded using Sra-tools (<https://github.com/ncbi/sra-tools>) and mapped to the *S. mediterranea* annotation (Neiro et al., 2022) using Kallisto (Bray et al., 2016). Expression analysis was performed using DESeq2 (Love et al., 2014).

3.4.4 Planarian culture

The asexual strain of *Schmidtea mediterranea* was used for experiments with *meis-1* and *SPILE-Z1*, while the sexual strain was used for knocking down *meis-like*, *PALÉ*, *SPILE-Z2*, and *SPILE-E1*. Worms were cultured in 0.5g/L Instant Ocean freshwater (0.5g Instant Ocean Aquarium sea salt mixture added to 1L of ultrapure water). The worms were kept in plastic boxes (20 cm x 20 cm x 15 cm) in the dark at 20 °C. The worms were fed calf liver twice a week, with the water being changed after each feeding. The worms were starved 7 days prior to RNAi experiments.

3.4.5 Cloning

3.4.5.1 Total RNA isolation

Total RNA was isolated by using the Trizol protocol. Two starved worms were collected into Eppendorf tubes on ice. Trizol (500 µl) was added, the sample was incubated at room temperature for 5 min, vortexed, and then frozen (-20°C) for 30 min. The sample was homogenised by pipetting up and down. Chloroform (100 µl) was added, the mixture was shaken for 15 seconds, and incubated at room temperature for 5 min. The tube was centrifuged (15 min, 12,000g, 4°C) and the aqueous phase was collected into another tube. Isopropanol (250 µl) was added, the tube vortexed, and incubated for 10 min at room temperature. The tube was centrifuged (10 min, 12,000g, 4°C), the supernatant was discarded, and the pellet was purified using 75% ethanol (500 µl). The pellet was centrifuged (5 min, 7500g at 4°C), the supernatant was discarded, and the pellet was air-dried before being resuspended in nuclease-free water. Nanodrop was used to measure the quality and quantity of RNA. Samples were stored at -80°C.

3.4.5.2 cDNA synthesis

The cDNA synthesis solution (total volume 12 µl) was prepared on ice: 1 µg total RNA, 1 µl Oligo(dT)₁₈ primers (Invitrogen), 1 µl 10mM dNTPs, and 9 µl nuclease-free water. The solution was incubated (65°C, 5 min) to denature RNA secondary structures, before being placed back on ice. Then, 4 µl First-strand buffer mix (5x Invitrogen), 2 µl DTT 0.1M, 1 µl RNaseOUT (1 U), and 1 µl SuperScript Reverse Transcriptase III (200 U) were added to the reaction mix. The reaction mix was incubated (37°C 50 min, 70°C 15 min) and then quenched on ice. The samples were stored at -20°C.

3.4.5.3 Cloning

The primers were designed by using PRIMER-BLAST (Ye et al., 2012) with default parameters, aiming for a target PCR product of 650-750 bp in length. The primers used to amplify genes from cDNA are listed in Table S3.2. The PCR sample (1 µl cDNA, 25 µl OneTaq 2X Master Mix (NEB), 1 µl forward primer, 1 µl reverse primer, 22 µl nuclease-free water) was cycled using the following parameter: denaturation 95°C 3 min; 35 cycles including denaturation 95°C 45 sec, annealing 60°C 30 sec, extension 72°C 1 min; final extension 72°C 10 min. Agarose gel electrophoresis and the Qiagen MinElute PCR purification kit were used to extract PCR products.

PPR-T4P (1 µg) vector was linearised with the restriction enzyme Sma1 (NEB) according to the instructions given in the kit. The digested plasmid was extracted from gels using the Qiagen MinElute PCR purification kit. The vector and the PCR product were T4P-treated separately. Both reactions included 300 ng DNA diluted in nuclease-free water up to a total volume of 8 µl, 1 µl 10x NEBuffer 2.1 (NEB), and 0.5 µl T4 DNA polymerase (NEB). 0.5 µl 50 mM dGTP was added to the vector reaction, while 0.5 µl 50 mM dCTP was added to the insert reaction. The reactions were incubated (22°C, 30 min) and heat inactivated (75°C, 20 min). The T4P-treated inserts and vectors were incubated (1h) at a ratio 5:1 at room temperature for ligation. The ligation mixture was added to 50 µl of competent DH5α cells (ThermoFisher). The transformation reaction was incubated (2 min) on ice, heat-shocked (42°C, 30 sec), and incubated (2 min) on ice. Luria-Bertani (LB, 250 µl) medium was added to the solution, and the reaction was cultivated (37°C, 1h, shaking incubator). The culture was spread on LB agar plates (kanamycin 50 µg/ml) and the plates were incubated (37°C overnight). The colonies were screened via colony PCR. Positive colonies were selected and incubated in LB broth (50 µg/ml kanamycin) overnight, and the plasmids were extracted with the QIAprep Spin Miniprep Kit.

3.4.5.4 Synthesis of dsRNA and riboprobes

M13 forward and M13 reverse primers were used for dsRNA synthesis, while AA18 and PPR224 primers were used for riboprobe synthesis. The PCR reaction (1 µl plasmid solution containing about 100 ng, 1 µl forward primer, 1 µl reverse primer, 25 µl OneTaq 2x Master Mix (NEB), 22 nuclease-free water) was cycled using the following parameters: denaturation 95°C 3 min; 40 cycles including denaturation 95°C 45 sec, annealing 55°C 30 sec, extension 72°C 1 min; final extension 72°C 10 min. The PCR product was purified from agarose gels with the Qiagen MinElute PCR purification kit and eluted in 30 µl of nuclease-free water.

For dsRNA synthesis, the transcription reaction (300-500 ng of purified PCR product template, 4 µl 10x RNAPol buffer NEB, 5 µl of each rNTP 10 mM, 4 µl T7 RNA polymerase (NEB), and nuclease-free water up to a final volume of 40 µl) was incubated at 37°C for 6 hours. TURBO DNase (2 µl, Invitrogen) and DNase buffer (4 µl) were added to the solution, incubated for an additional 30 min at 37°C. Nuclease-free water (54 µl) was added, and the tube was placed on ice. The RNA was precipitated by adding sodium acetate (10 µl 3M, pH 5.2) and ethanol (600 µl, -20°C), and the tube was vortexed mixed well before being incubated at -20°C overnight. This solution was centrifuged (13,000g, 20 min, 4°C), the supernatant was discarded, the pellet was purified with pre-chilled 70% ethanol (-20°C), and the sample was centrifuged again (13,000g, 20 min, 4°C). The pellet was air-dried and resuspended in nuclease-free water (30 µl). To anneal the RNA strands, the solution was incubated at 68°C for 15 min and then at 37°C for 45 min. The quality of dsRNA was assessed on an agarose gel and stored at -80°C.

For riboprobe synthesis, the transcription reaction (500 ng of purified PCR product template, 2 µl T7 RNA polymerase (NEB), 2 µl 10X RNAPol buffer (NEB), 2 µl 10XDIG RNA labelling mix (Roche) or 10X FITC RNA labelling mix (Roche), nuclease free water up to a total volume of 20 µl) was incubated at 37°C for 3 hours. Then TURBO DNase (1 µl) and DNase buffer (2 µl) was added to the solution, which was incubated for an additional 30 min at 37°C. Nuclease-free water (27 µl), lithium chloride (0.5 M), glycogen (1 µl), and ethanol (300 µl) were added to the solution that was incubated at -20°C overnight. This solution was centrifuged (13,000g, 20 min, 4°C), the supernatant was discarded, the pellet was purified with pre-chilled 70% ethanol (-20°C), and the sample was centrifuged again (13,000g, 20 min, 4°C). The solution was incubated at 68°C for 15 min, the quality of RNA was assessed on agarose gels, and the riboprobes were stored at -80°C.

3.4.6 RNA interference

dsRNA was microinjected into planarians to perform RNA interference. A PB-7 Microscope Micropipette Puller (Narishige) was used to pull injection needles from 3.5" Drummon Scientific glass capillaries.

Nanoject II microinjector (Drummond Scientific) was used to inject 3x69nl of dsRNA (1 µg/µl) each injection day. For a single RNAi experiment, each planarian was injected for 3 consecutive days, followed by a waiting period of 4 consecutive days, before being injected again for 3 consecutive days. The worms were amputated at day 11.

The phenotypes were monitored by using a Zeiss Stereo Discovery V8 microscope and a manually installed Canon EOS1200D digital camera. The worms were imaged using 1.6-2x magnification. A ruler was used to calibrate the scale bar. The size of eyes (area) and interocular distance were measured by using ImageJ (Schindelin et al., 2012).

3.4.7 In situ hybridisation

Fluorescence in situ hybridisation was performed according to the protocol devised by King et al. (2013).

3.5 Supplementary material

Table S3.1. Protein sequences used in the metazoan phylogeny. “Portal” refers to the resource provider, “ID” indicates the identifier of the proteome or transcriptome, and “Reference” is the original sequencing project, if the information is available in the database.

Species	Portal	ID	Reference
<i>Mnemiopsis leidyi</i>	Mnemiopsis Genome Project Portal	ML.2.2.aa.gz	(Moreland et al., 2020)
<i>Amphimedon queenslandica</i>	Genbank	GCF_000090795.2	(Srivastava et al., 2010)
<i>Nematostella vectensis</i>	Genbank	GCF_932526225.1	Wellcome Sanger Institute
<i>Hydra vulgaris</i>	Genbank	GCF_022113875.1	University of Vienna
<i>Hofstenia miamia</i>	Genbank	GCA_900660155.1	(Gehrke et al., 2019)
<i>Strongylocentrotus purpuratus</i>	Genbank	GCF_000002235.5	(Sodergren et al., 2006)
<i>Saccoglossus kowalevskii</i>	Genbank	GCF_000003605.2	(refer to Genbank)
<i>Branchiostoma floridae</i>	Genbank	GCF_000003815.2	(Simakov et al., 2020)
<i>Danio rerio</i>	Uniprot	UP000000437	(refer to Uniprot)
<i>Mus musculus</i>	Uniprot	UP000000589	(refer to Uniprot)
<i>Homo sapiens</i>	Uniprot	UP000005640	(refer to Uniprot)
<i>Caenorhabditis elegans</i>	Uniprot	UP000001940	(refer to Uniprot)
<i>Drosophila melanogaster</i>	Uniprot	UP000000803	(refer to Uniprot)
<i>Tribolium castaneum</i>	Genbank	GCF_031307605.1	United States Department of Agriculture
<i>Intoshia linei</i>	Genbank	GCA_001642005.1	(Mikhailov et al., 2016)
<i>Pomphorhynchus laevis</i>	Genbank	GCA_012934845.2	(Mauer et al., 2020)
<i>Adineta vaga</i>	Genoscope	Adineta_vaga.v2.pep.fa	(refer to Genoscope)
<i>Nipponacmea fuscoviridis</i>	-	(Sequences in publication)	(Morino et al., 2017)
<i>Lottia gigantea</i>	JGI Genome Portal	Lotgi1_GeneModels_AllModels_20070424_aa.fasta	(Simakov et al., 2013)
<i>Mytilisepta virgata</i>	-	8Sequences from publication)	(Phuangphong et al., 2021)
<i>Pinctada fucata</i>	Marine Genomics OISR Japan	pfu_aug2.0.AA.fasta	(Takeuchi et al., 2016)
<i>Crassostrea gigas</i>	GigaDB	oyster.v9.fa	(Fang et al., 2012)
<i>Lingula anatina</i>	Genbank	GCF_001039355.2	(Luo et al., 2015)
<i>Phoronis australis</i>	Marine Genomics OISR Japan	pau_genome_v2.0_prot.fa.gz	(Luo et al., 2018)
<i>Platynereis dumerilii</i>	PdumBase	Trinity_95_orf_v2.fasta	(Chou et al., 2018)
<i>Spirobranchus kraussii</i>	-	(Sequences in publication)	(Morino et al., 2017)
<i>Helobdella robusta</i>	JGI Genome Portal	Helobdella_robusta.Helro1.pep.all.fa	(Simakov et al., 2013)
<i>Lineus longissimus</i>	Genbank	GCF_910592395.1	(Kwiatkowski et al., 2021)
<i>Schmidtea mediterranea</i>	(Chapter 2)	(Chapter 2)	(Neiro et al., 2022)
<i>Echinococcus multilocularis</i>	WormBase	echinococcus_multilocularis.PRJEB122.WBPS18.protein.fa	(Tsai et al., 2013)

Table S3.2 Forward and reverse primer sequences to clone *SPILE-E1*, *meis-like*, *SPILE-Z2*, *meis-1*, and *PALE*.

Primer	Sequence
SPILE-E1 Forward	TGGAAATGCATTCAGAGAGTACGTT
SPILE-E1 Reverse	GTTGTCACTTCGGTTGCGCC
meis-like Forward	ATGGTTTTCTCGGCATTTTCGC
meis-like Reverse	CGGAAGATTTAAATTTCGAGTGAAGG
SPILE-Z2 Forward	TGGCCACGACCCAAGTCAA
SPILE-Z2 Reverse	AGTGAGAAGAGACGGTGGCGT
meis-1 Forward	TTTCAGCGCAGATCCTGACA
meis-1 Reverse	ACATTCTGGTGGTCAACGCT
SPILE-Z1 Forward	GGCGGACTGACATCTGACAA
SPILE-Z1 Reverse	TTGCATGCTGAAGGAGGAGG
PALE Forward	GCCGTCTTTGCACCAACACA
PALE Reverse	TTGACATAACGACGCCGAGCA

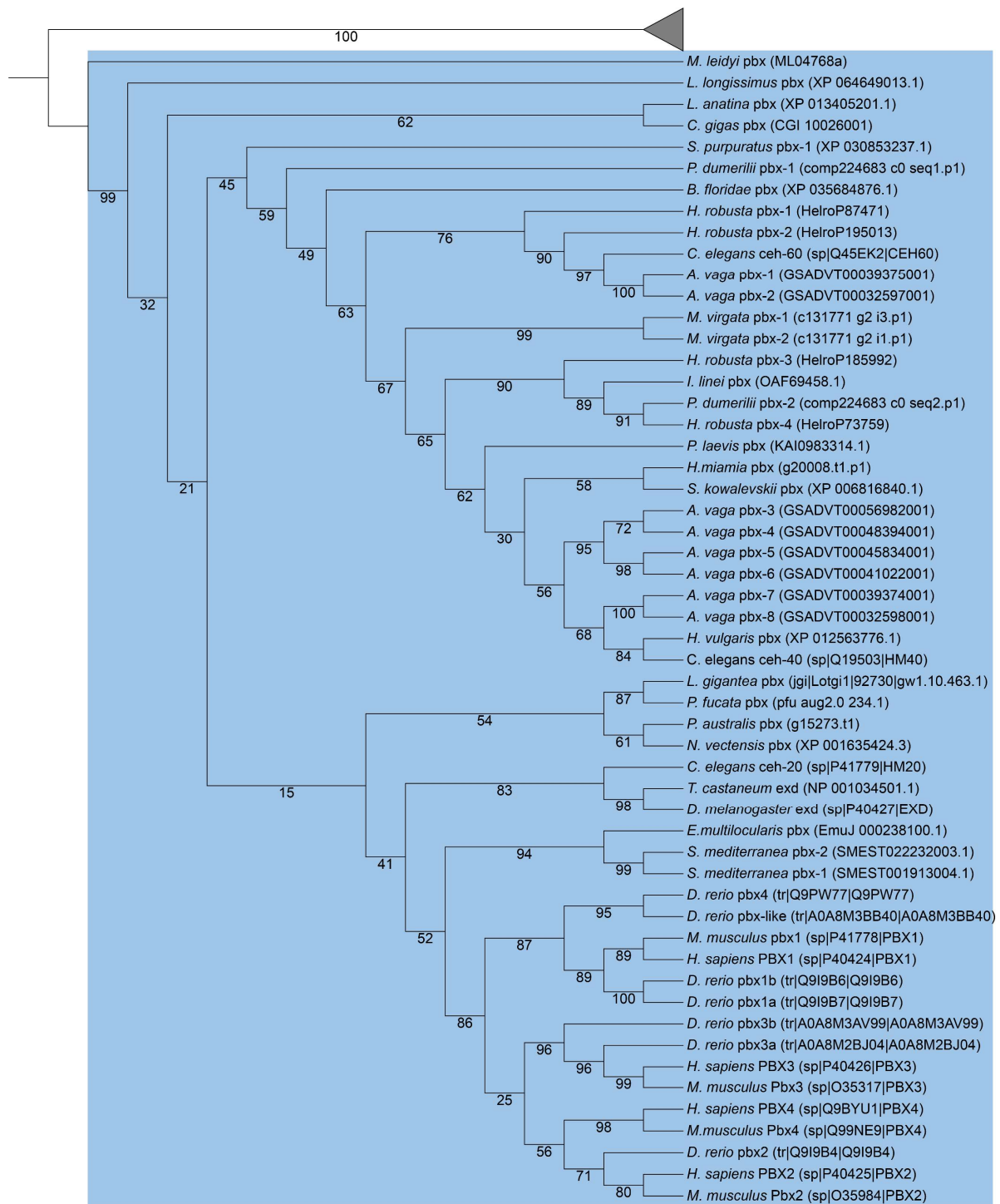


Figure S3.1. Maximum likelihood phylogeny of metazoan Pbx proteins. The tree was constructed by extracting Pbx genes in the general homeodomain-based phylogeny of metazoan TALE genes (Figure 3.2B) and constructing a separate phylogeny based solely on Pbx homeodomains. Bootstrap values are shown at branches (Hoang et al. 2018). Full species names are given in Figure 3.1A. Human Hox genes were used as an outgroup.

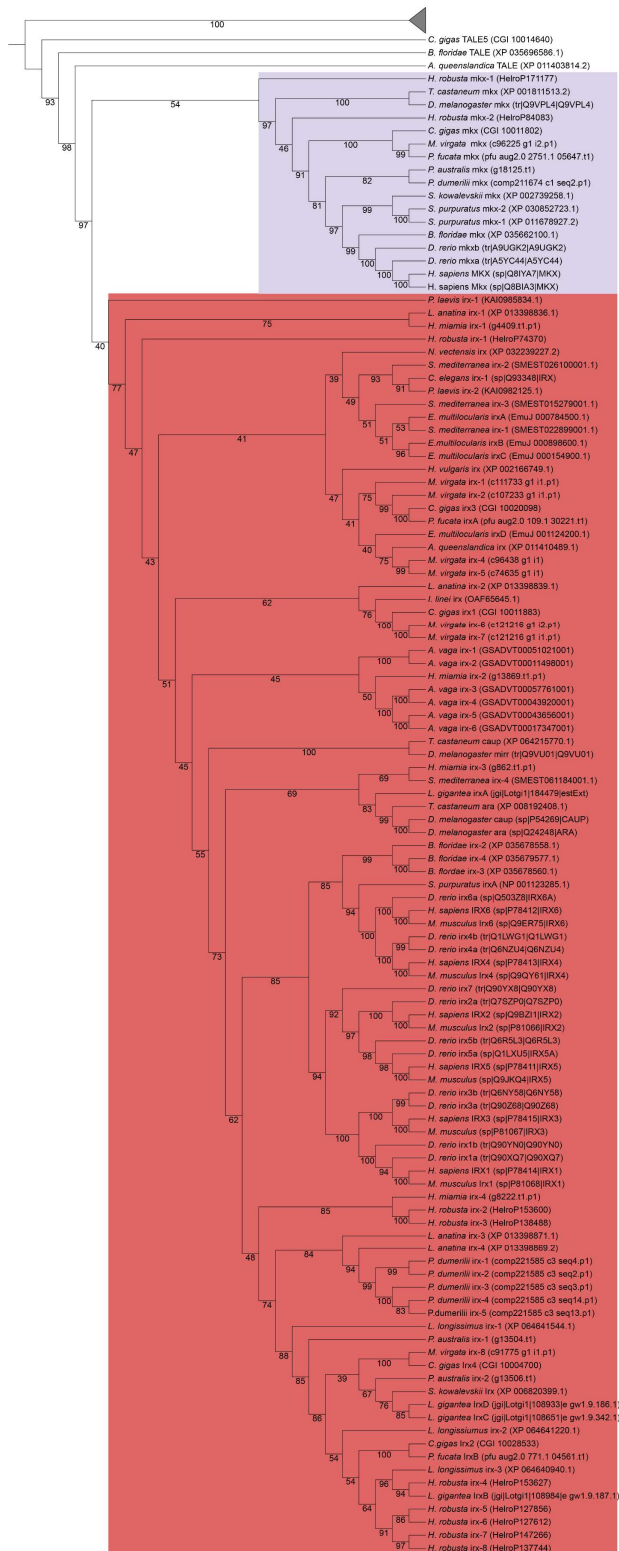


Figure S3.2. Maximum likelihood phylogeny of metazoan Irx and Mxk proteins. The tree was constructed by extracting Irx and Mxk genes in the general homeodomain-based phylogeny of metazoan TALE genes (Figure 3.1D2) and constructing a separate phylogeny based on full-length sequences. Bootstrap values are shown at branches (Hoang et al. 2018). Full species names are given in Figure 3.1A. Human Hox genes were used as an outgroup.

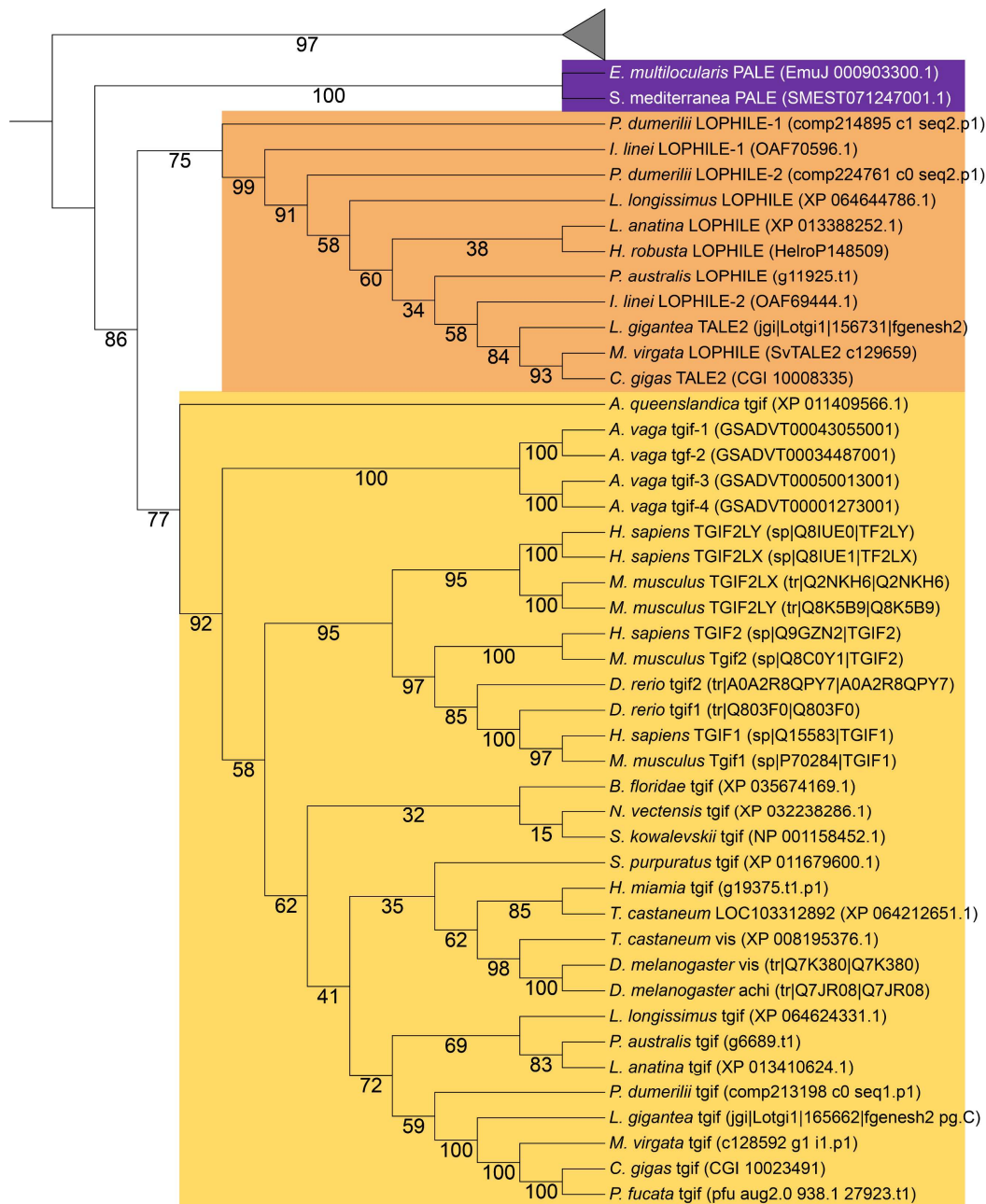


Figure S3.3 Maximum likelihood phylogeny of metazoan Tgif, LOPHILE and PALE proteins. The tree was constructed by extracting Tgif, LOPHILE, and PALE sequences in the general homeodomain-based phylogeny of metazoan TALE genes (Figure 3.1D2) and constructing a separate phylogeny based on full-length sequences. Bootstrap values are shown at branches (Hoang et al. 2018). Full species names are given in Figure 3.1A. Human Hox genes were used as an outgroup.

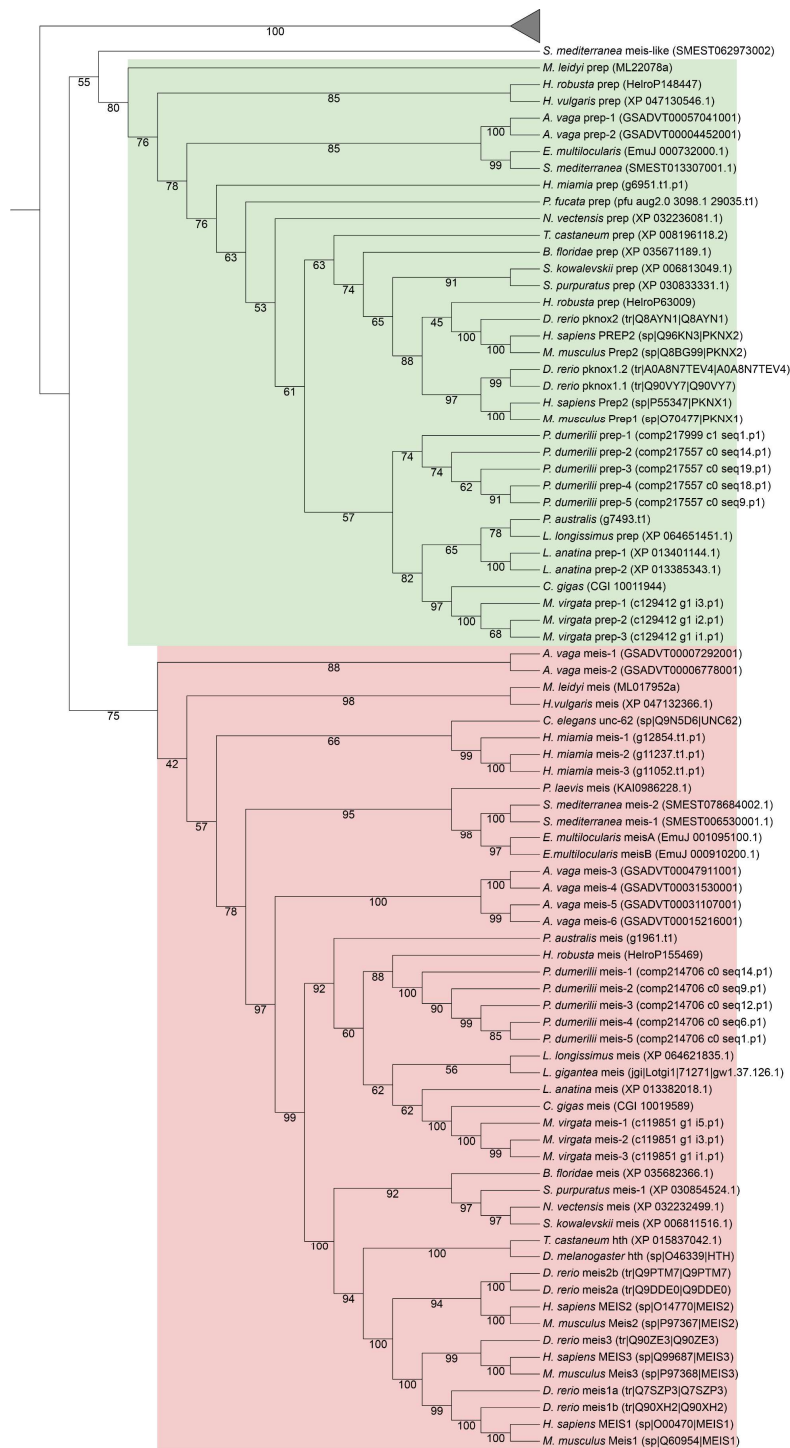


Figure S3.4. Maximum likelihood phylogeny of metazoan Meis and Prep proteins. The tree was constructed by extracting Meis and Prep sequences in the general homeodomain-based phylogeny of metazoan TALE genes (Figure 3.1D2) and constructing a separate phylogeny based on full-length sequences. Bootstrap values are shown at branches (Hoang et al. 2018). Full species names are given in Figure 3.1A. Human Hox genes were used as an outgroup.

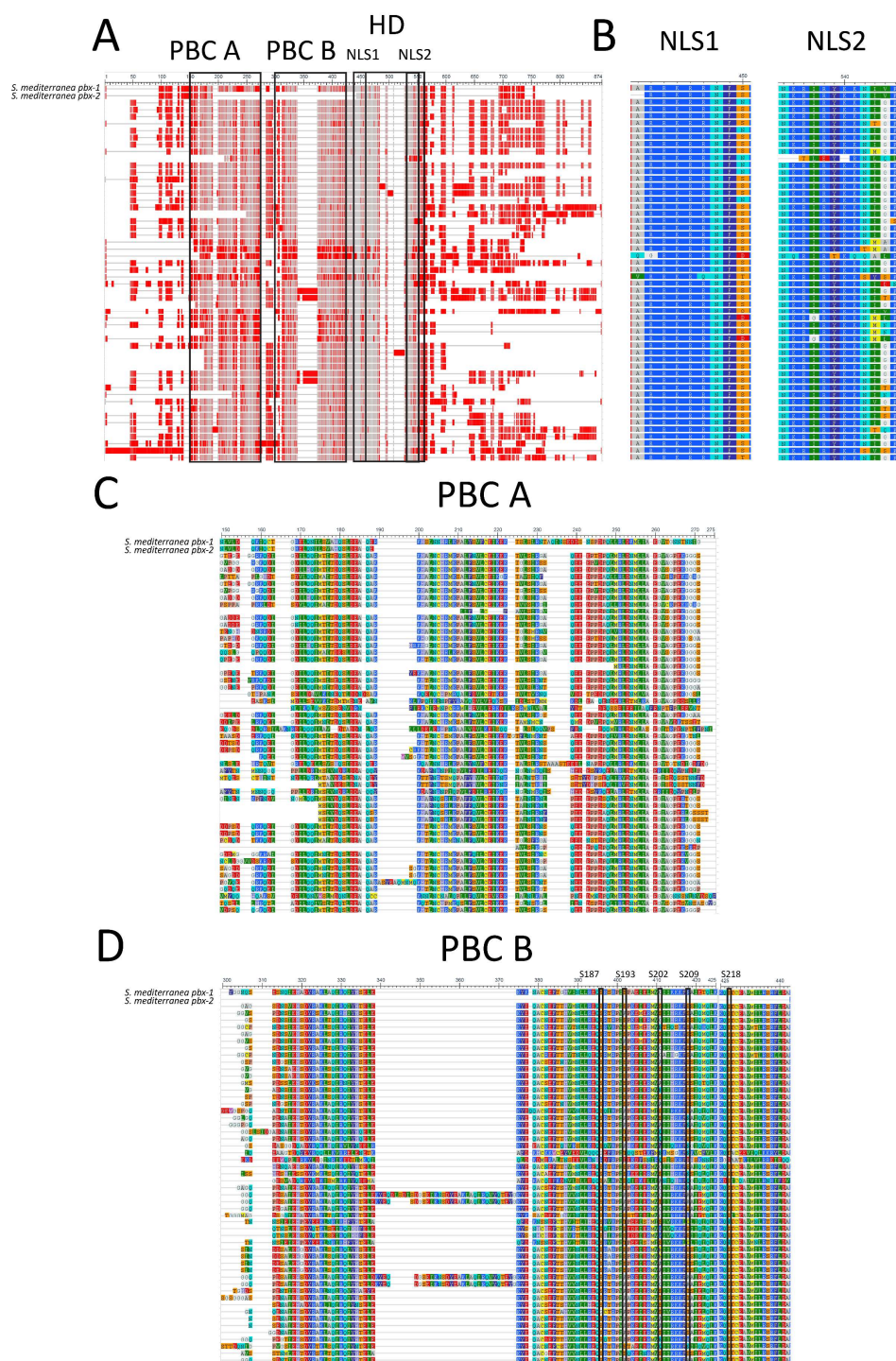


Figure S3.5 Multiple sequence alignment of metazoan Pbx proteins. **(A)** Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The conserved domains PBC-A, PBC-B, NLS1, and NLS2 are labelled in the alignment. **(B)** Magnified view of the NLS1 and NLS2 domains in the alignment. **(C)** Magnified view of the PBC-A domain in the alignment. **(D)** Magnified view of the PBC-B domain in the alignment.

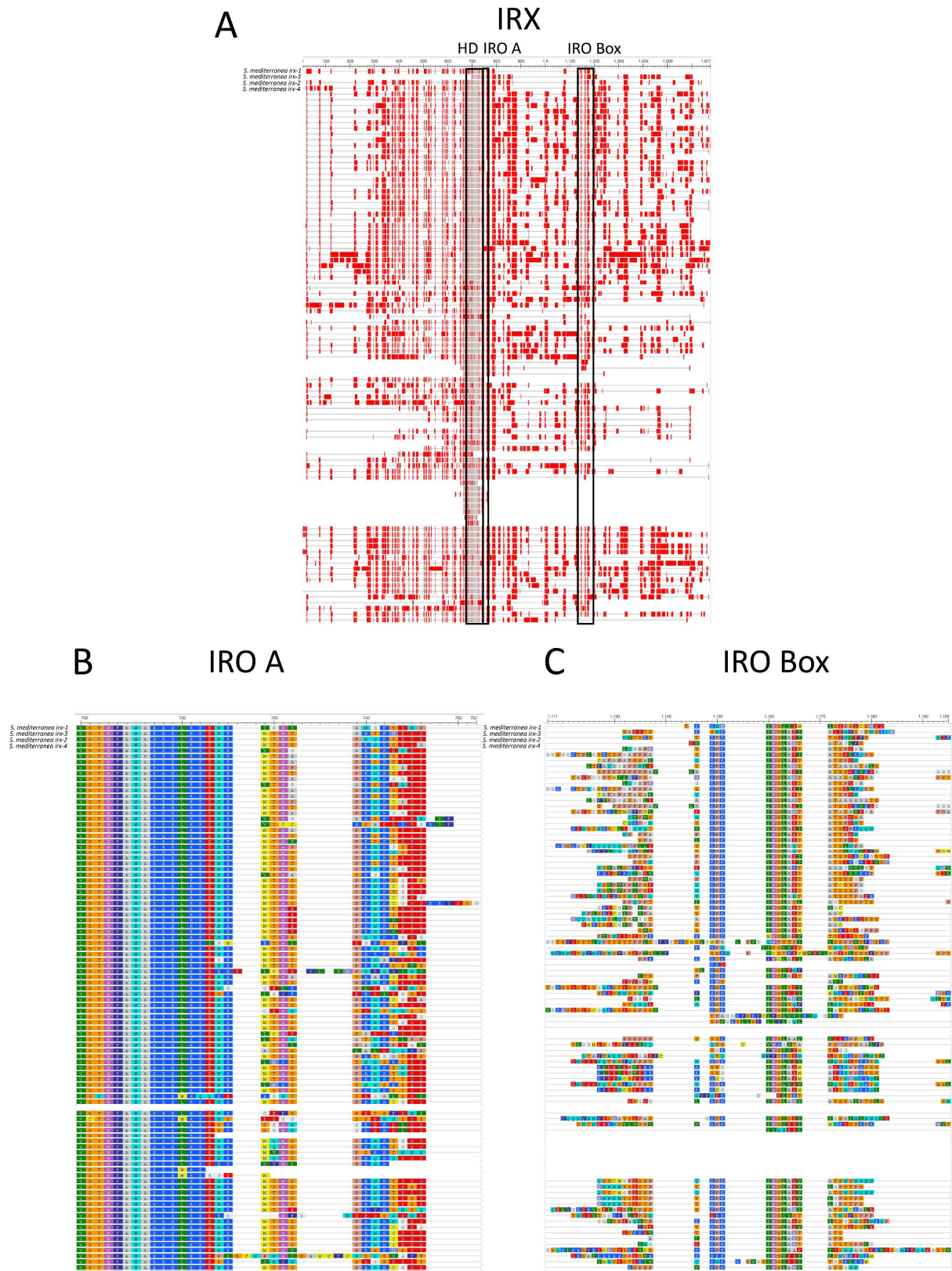


Figure S3.6 Multiple sequence alignment of metazoan Irx proteins. (A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain, the IRO A, and IRO box are labelled in the alignment. **(B)** Magnified view of the IRO A domain in the alignment. **(C)** Magnified view of the IRO Box domain in the alignment.

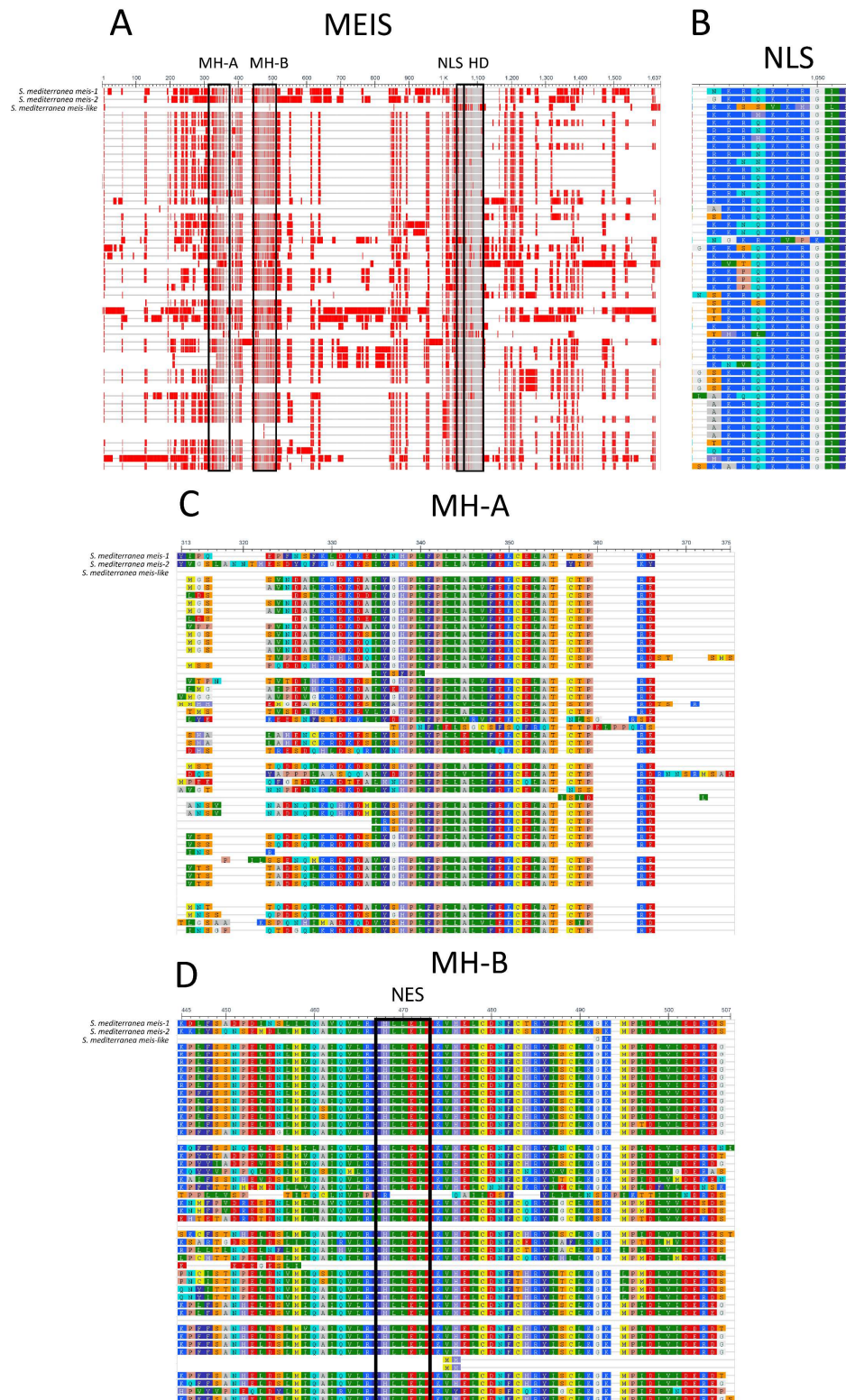


Figure S3.7 Multiple sequence alignment of metazoan Meis proteins. **(A)** Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain, the NLS domain, and the MH-A and MH-B domains are labelled. **(B)** Magnified view of the NLS domain in the alignment. **(C)** Magnified view of the MH-A domain in the alignment. **(D)** Magnified view of the MH-B domain in the alignment.

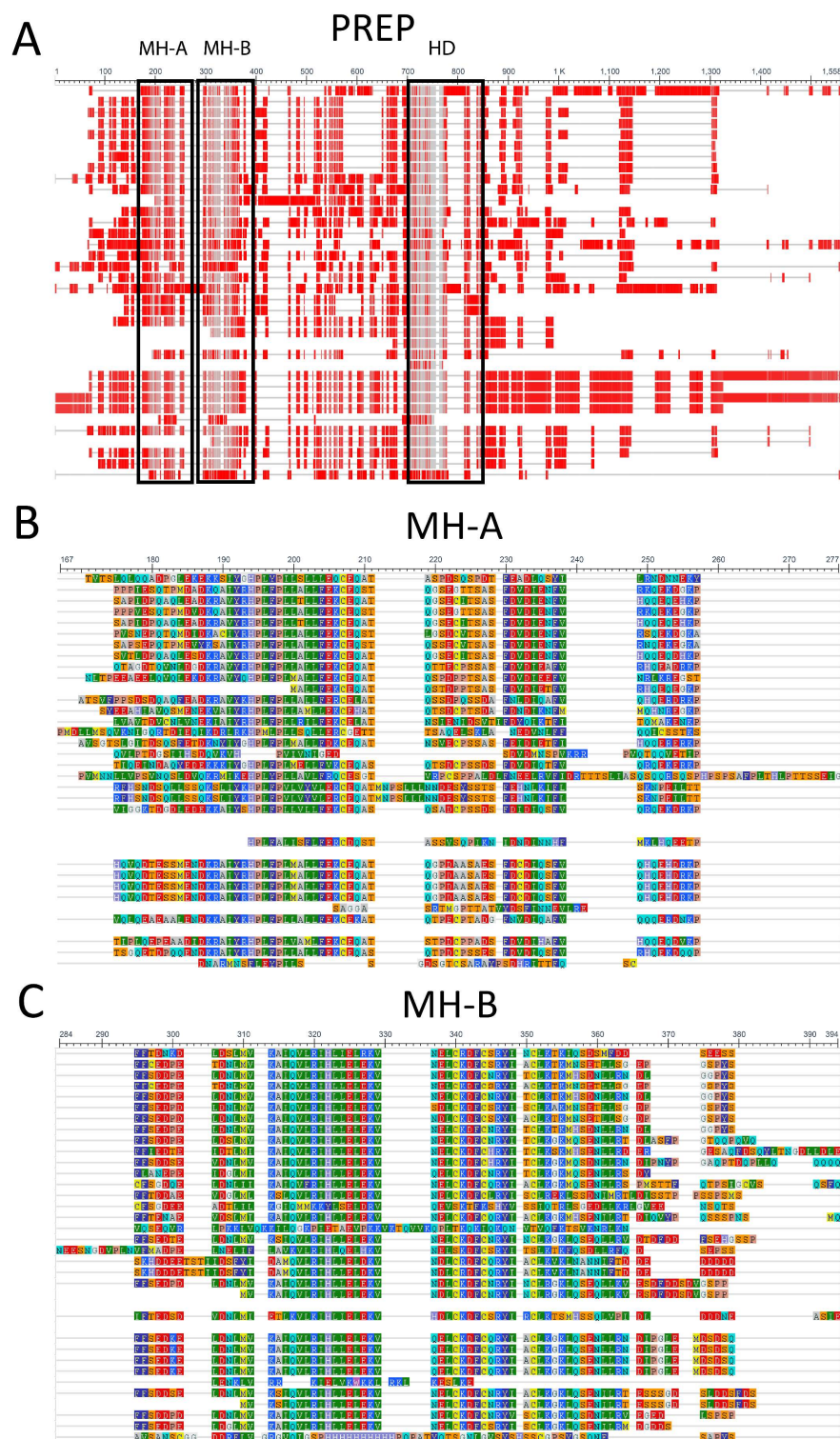


Figure S3.8 Multiple sequence alignment of metazoan Prep proteins. (A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain and the MH-A and MH-B domains are labelled. **(B)** Magnified view of the MH-A domain in the alignment. **(D)** Magnified view of the MH-B domain in the alignment.

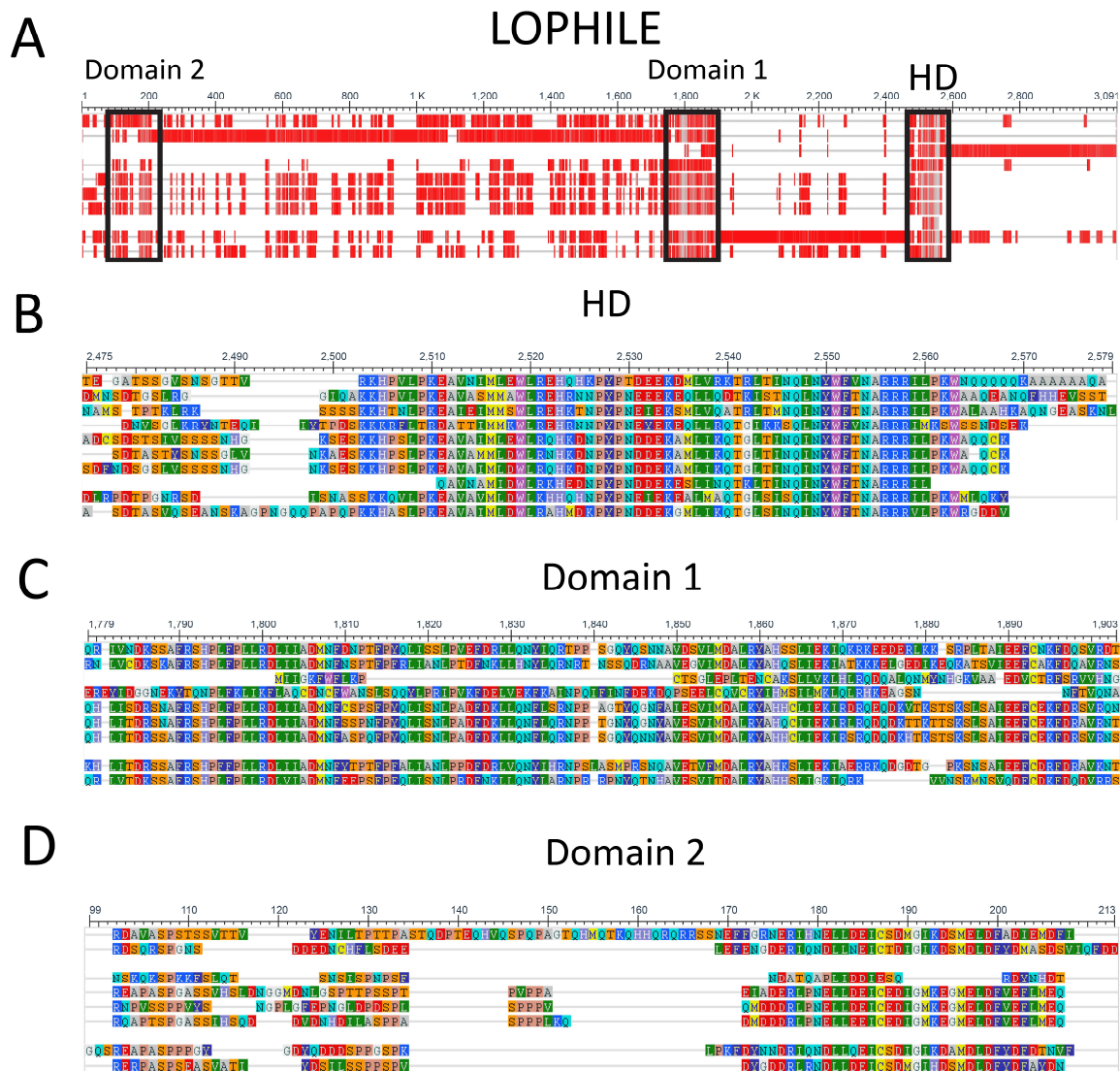
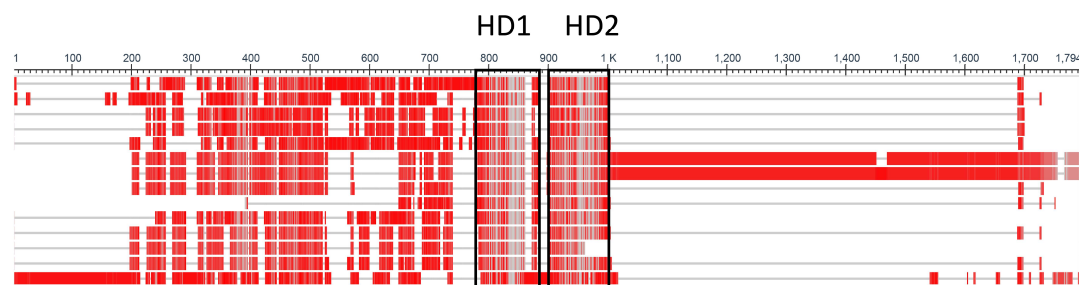


Figure S3.9 Multiple sequence alignment of spiralian-specific LOPHILE proteins. (A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain and two conserved regions are labelled in the alignment. **(B)** A magnified view of the homeodomain demonstrates the presence of a homeodomain with the TALE-specific three amino acids insertion (PYP). **(C)** Magnified view of the first conserved domain. **(D)** Magnified view of the second conserved domain.

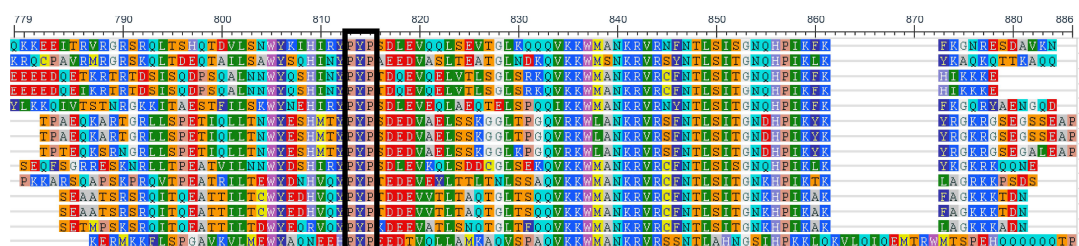
A

SPILE-B



B

HD1



C

HD2

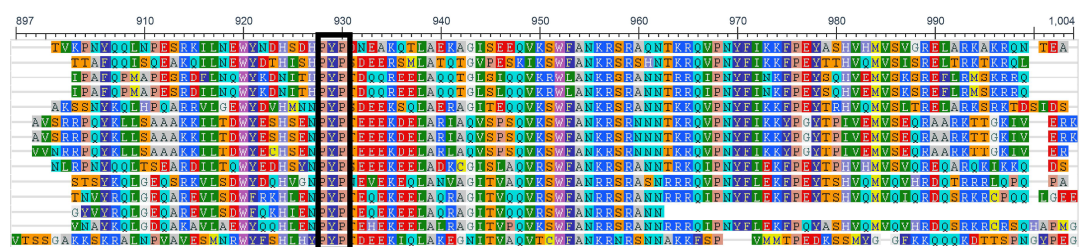


Figure S3.10 Multiple sequence alignment of SPILE-B proteins in the phylogeny of SPILE genes.

(A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The two homeodomains are labelled in the alignment. **(B)** A magnified view of the first homeodomain demonstrates the presence of the TALE-specific three amino acids insertion (PYP). **(C)** Magnified view of the second homeodomain also demonstrates the presence of the TALE-specific three amino acids insertion (PYP).

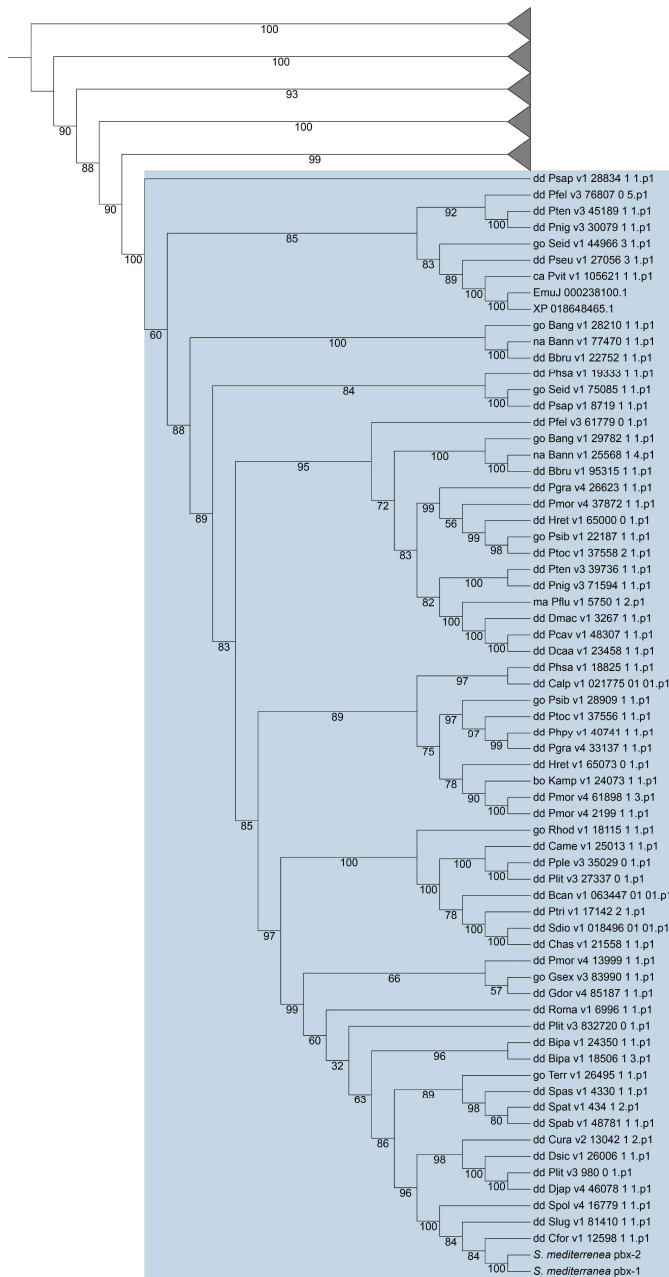


Figure S3.11 A detailed view of the Pbx clade in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviations: dd_Dtah_v3 = *Dugesia tabitiensis*, dd_Bbru_v1 = *Bdellocephala* cf. *brunnea*, dd_Bcan_v1 = *Bdelloura candida*, dd_Bipa_v1 = *Bipalium kewense*, dd_Calp_v1 = *Crenobia alpina*, dd_Came_v1 = *Camerata robusta*, dd_Cfor_v1 = *Cura foremanii*, dd_Chas_v1 = *Cercyra hastata*, dd_Cura_v2 = *Cura pinguis*, dd_Dcaa_v1 = *Dendrocoelid* sp. 1, dd_Djap_v4 = *Dugesia japonica*, dd_Dmac_v1 = *Dendrocoelum* cf. *maculatum*, dd_Dsic_v1 = *Dugesia sicula*, dd_Gdor_v4 = *Girardia dorotocephala*, dd_Hret_v1 = *Hymanella retenuova*, dd_Pcav_v1 = *Dendrocoelid* sp. 2, dd_Pfel_v3 = *Polycelis felina*, dd_Pgra_v4 = *Phagocata gracilis*, dd_Phpy_v1 = *Phagocata pyrenaica*, dd_Phsa_v1 = cf. *Atrioplanaria*, dd_Plit_v3 = *Procerodes littoralis*, dd_Pnig_v3 = *Polycelis nigra*, dd_Pple_v3 = *Procerodes plebeius*, dd_Psap_v1 = *Polycelis Sapporo*, dd_Pseu_v1 = *Cylindrostoma* sp., dd_Pten_v3 = *Polycelis tenuis*, dd_Ptri_v1 = *Paucumara* sp., dd_Roma_v1 = cf. *Romankenkins*, dd_Sdio_v1 = *Sabussowia dioica*, dd_Slug_v1 = *Schmidtea lugubris*, dd_Spab_v1 = cf. *Spathula* sp. 3, dd_Spas_v1 = cf. *Spathula* sp. 2, dd_Spat_v1 = cf. *Spathula* sp. 1, dd_Spol_v4 = *Schmidtea polychroa*, dd_Ptoc_v1 = *Planaria torva* - Croatian strain, go_Bang_v1 = *Bdellocephala angarensis*, go_Gsex_v3 = *Girardia tigrina* sexual strain, go_Psib_v1 = *Phagocata sibirica*, go_Rhod_v1 = *Rhodax* sp., go_Seid_v1 = cf. *Seidlia*, go_Terr_v1 = *Caenoplania decolorate*.

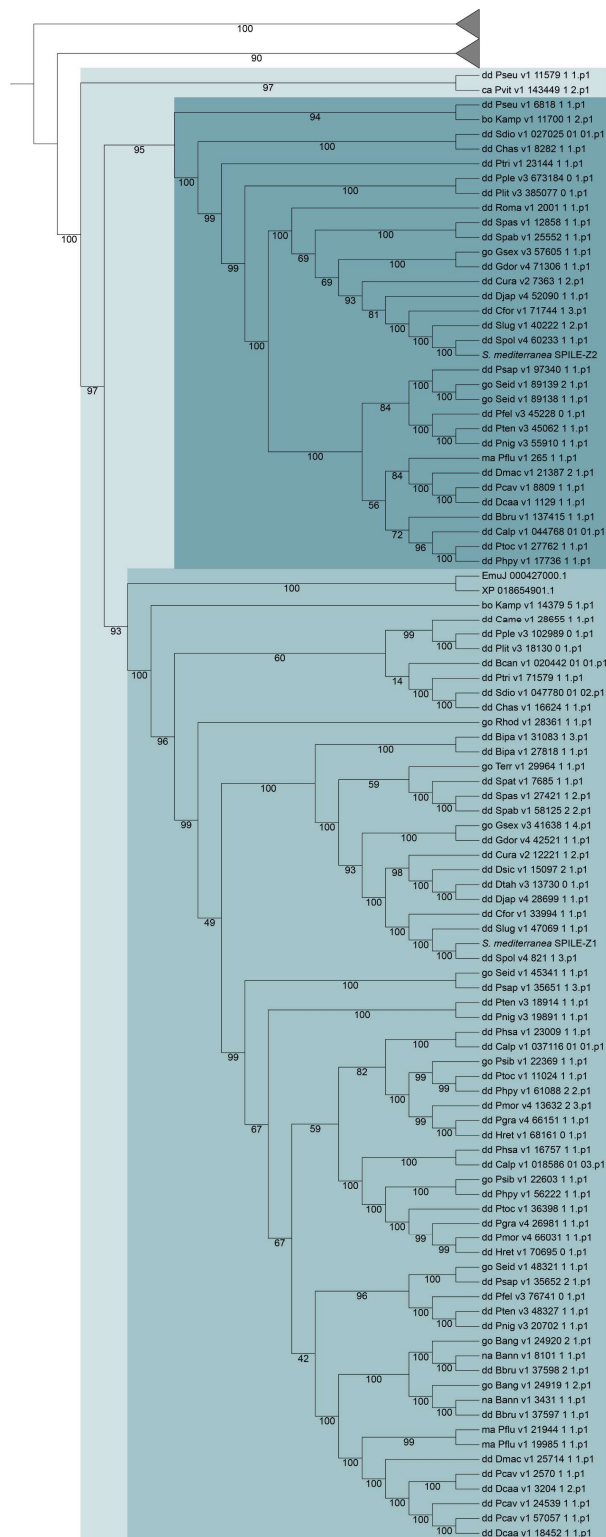


Figure S3.12 A detailed view of the SPILE-Z clade in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviation the same as in S.11.

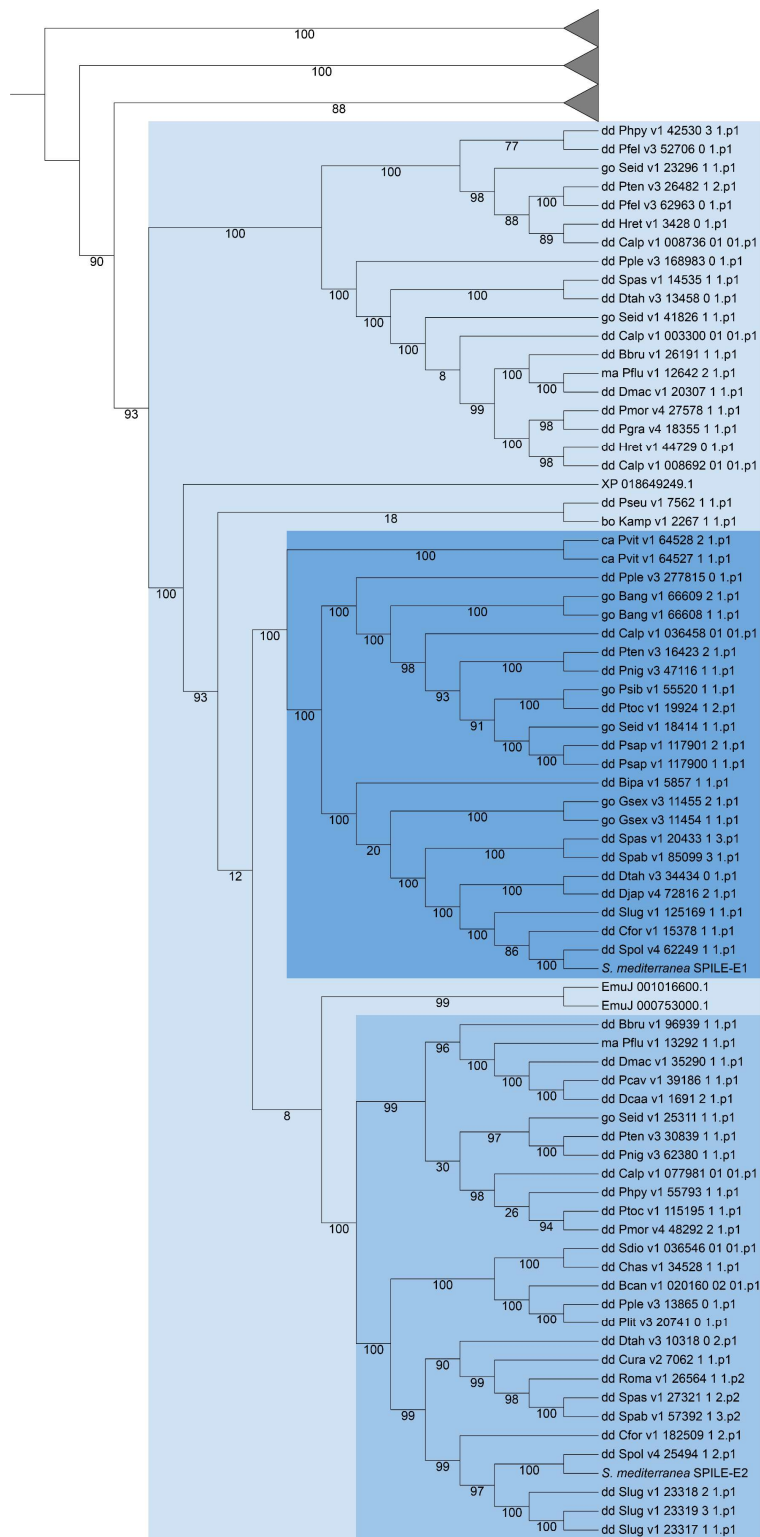


Figure S3.13 A detailed view of the SPILE-E clade in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviation the same as in S.11.

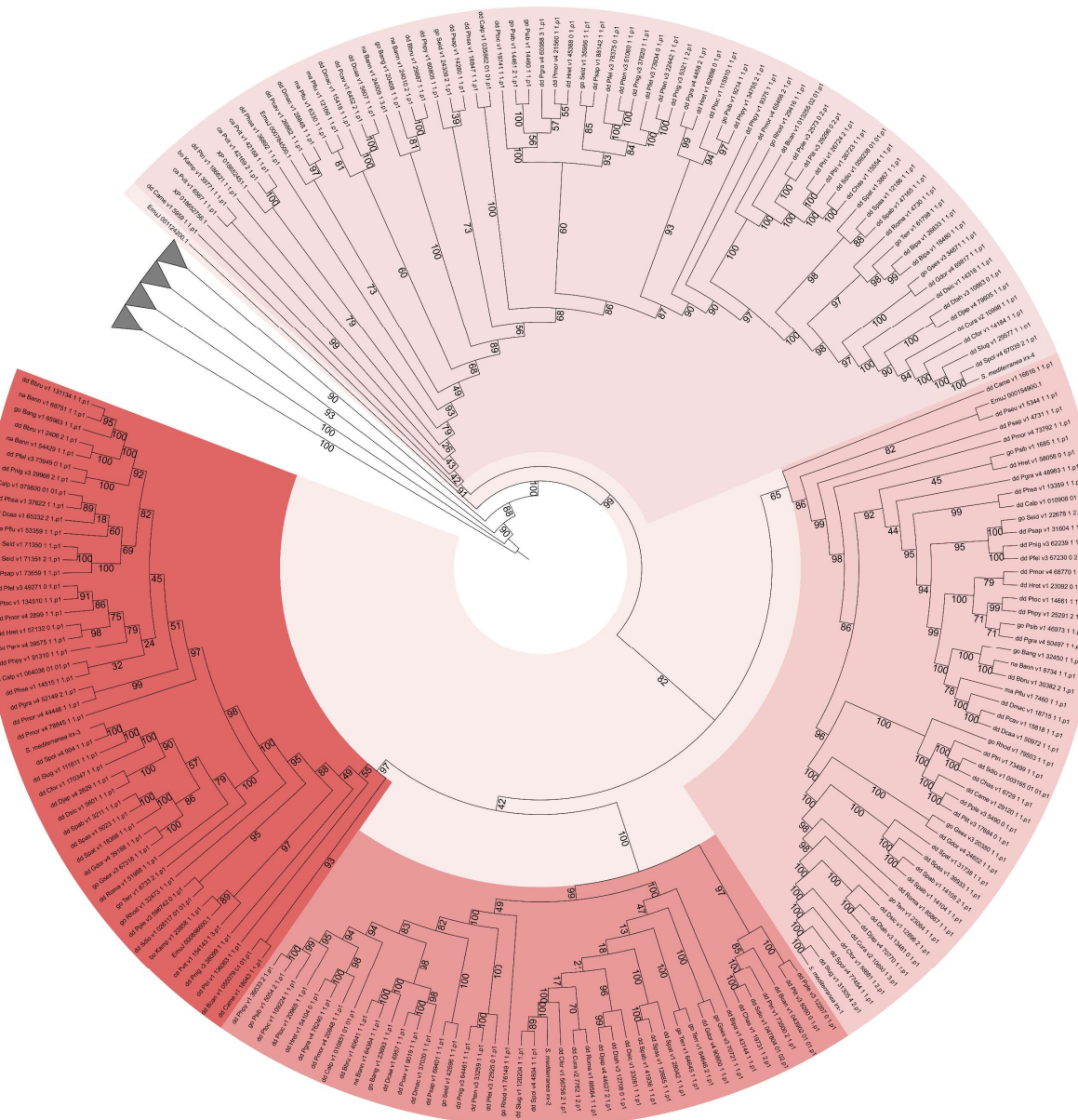


Figure S3.14 A detailed view of the *Irx* clade in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviation the same as in S.11.

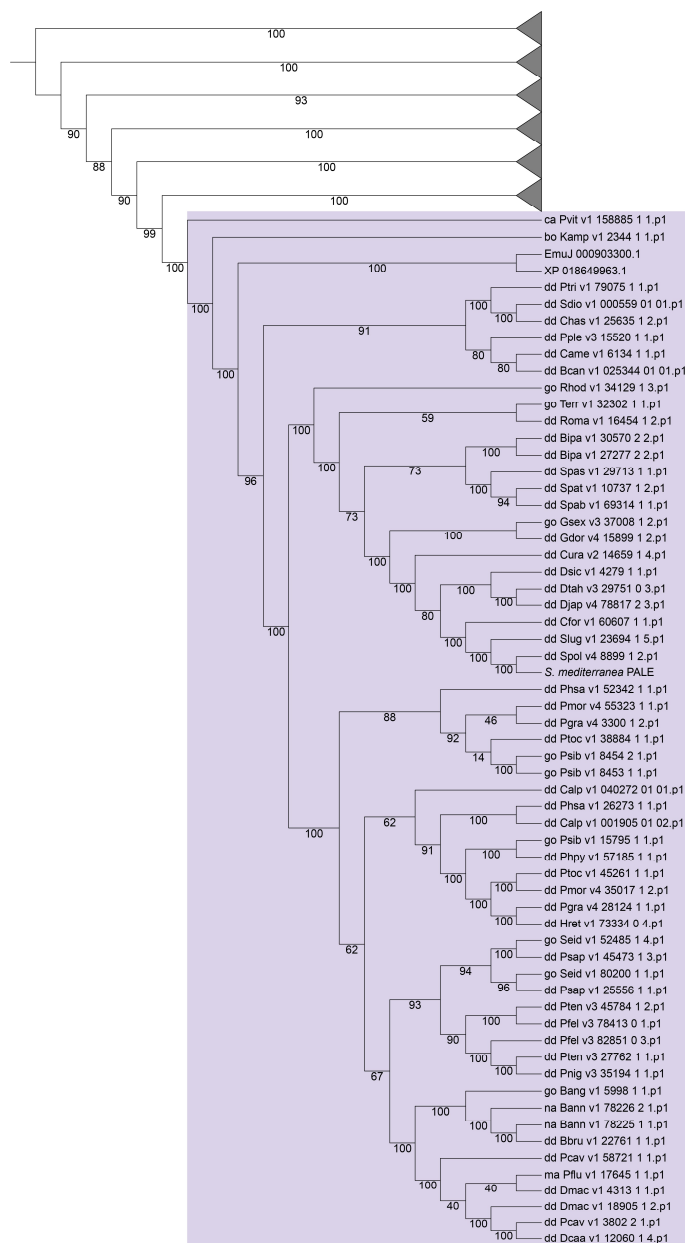


Figure S3.15 A detailed view of the PALE clade in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviation the same as in S.11.

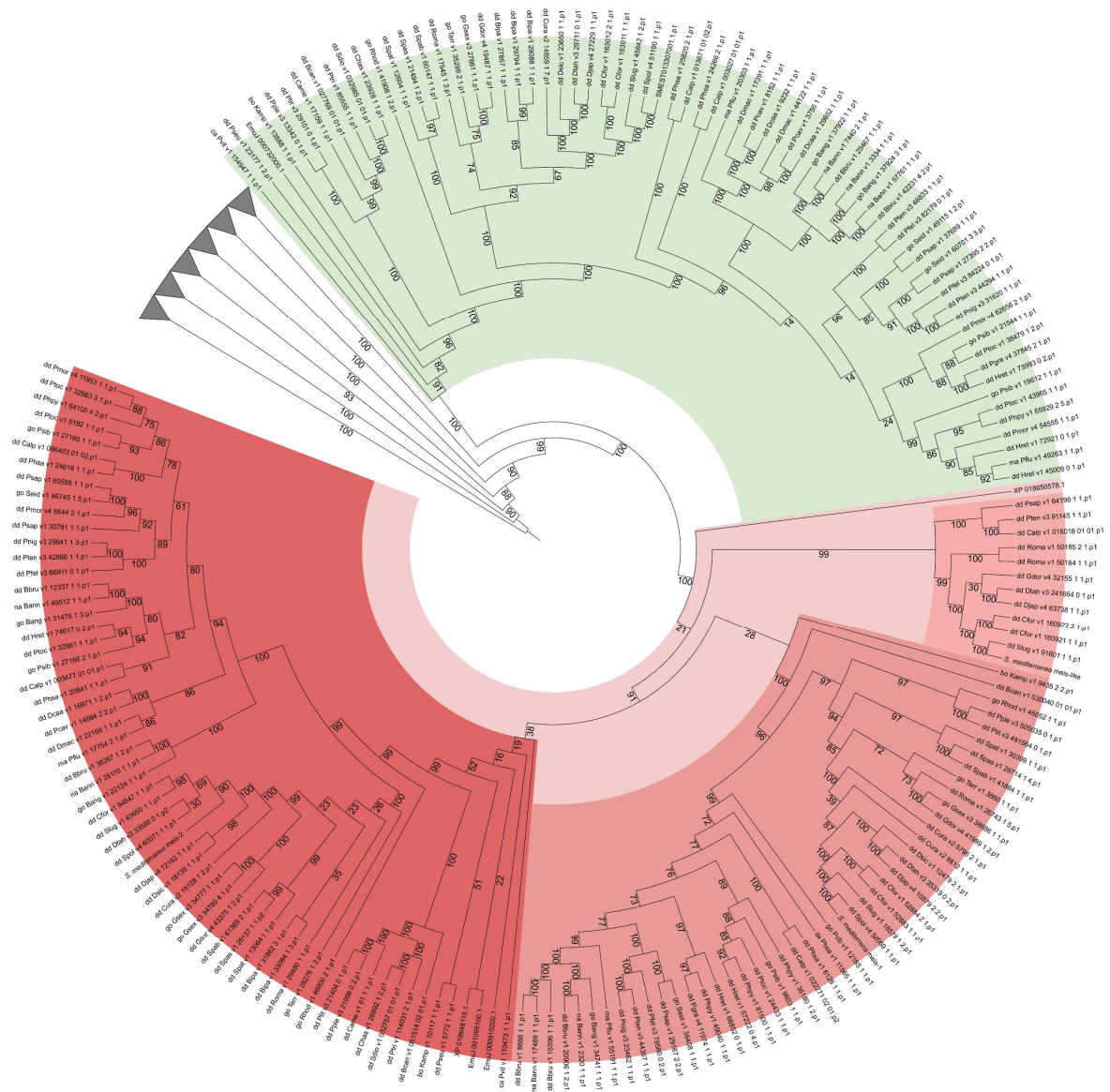


Figure S3.16 A detailed view of the MEINOX clade including Prep and Meis in the planarian phylogeny. Bootstrap values are shown next to branches. Abbreviation the same as in S.11.

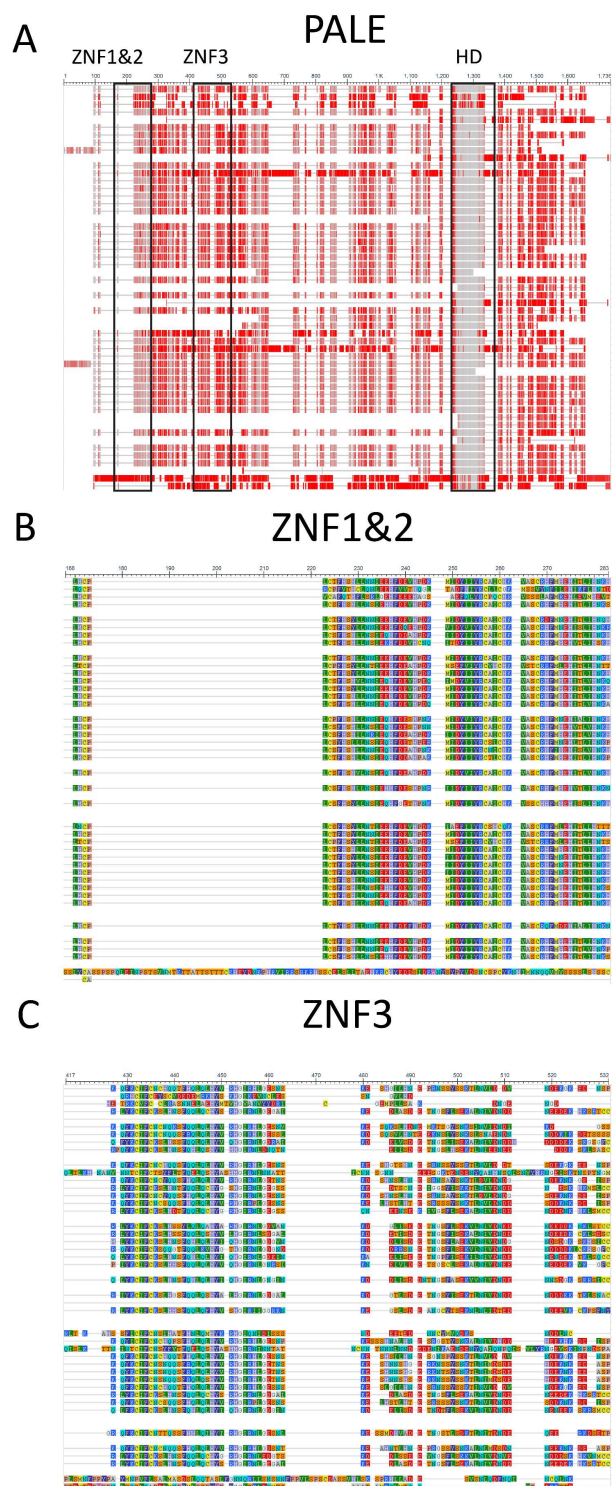


Figure S3.17 Multiple sequence alignment of PALE proteins in the planarian phylogeny. (A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain and two zinc finger motifs are labelled. **(B)** A magnified view of the first zinc finger. **(C)** A magnified view of the second zinc finger.

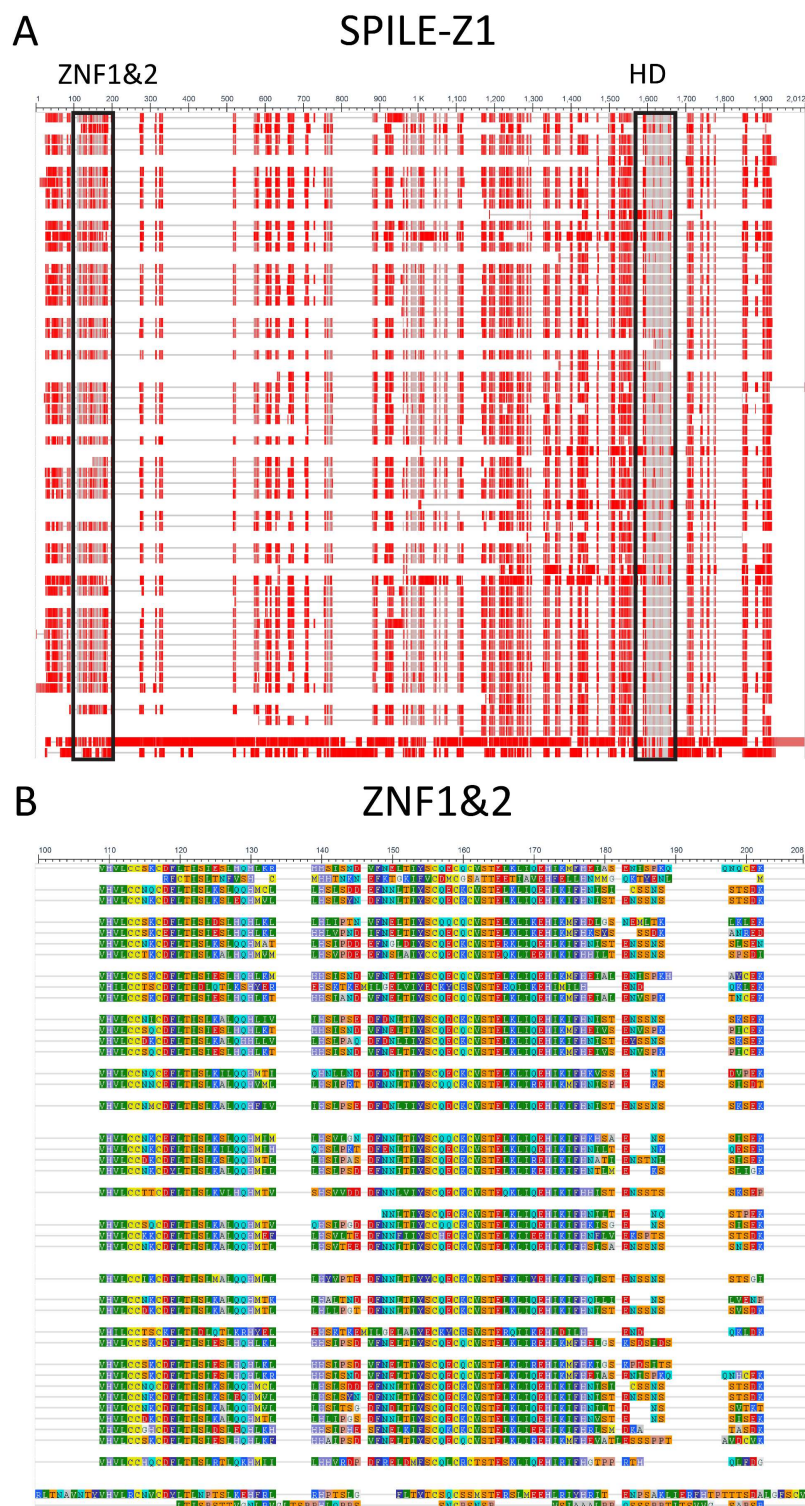


Figure S3.18 Multiple sequence alignment of SPILE-Z1 proteins in the planarian phylogeny. (A) Alignment of the entire sequence. Conserved regions are displayed in grey, variable regions in red. The homeodomain and two zinc finger motifs are labelled. **(B)** A magnified view of the first and second zinc fingers.

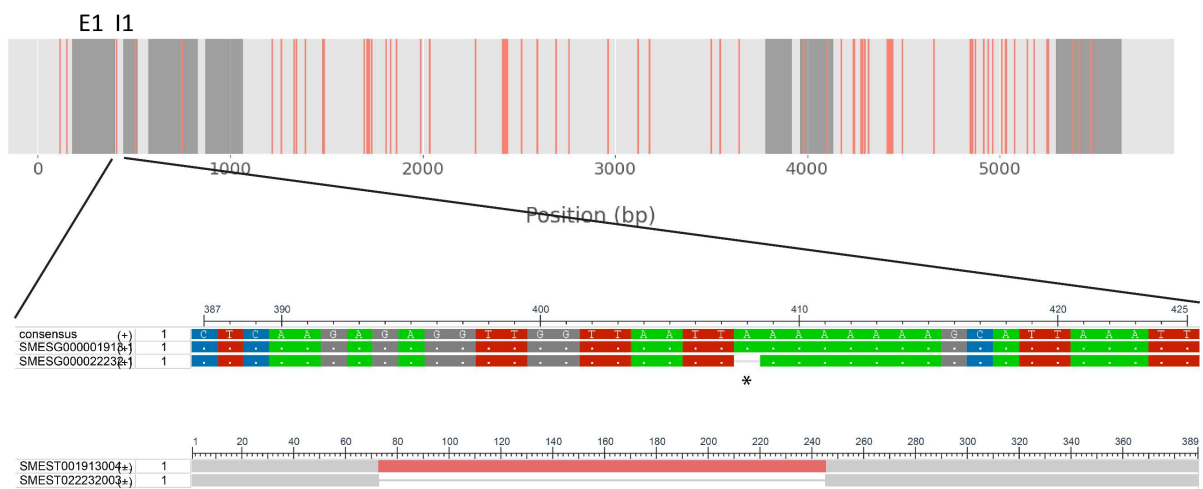
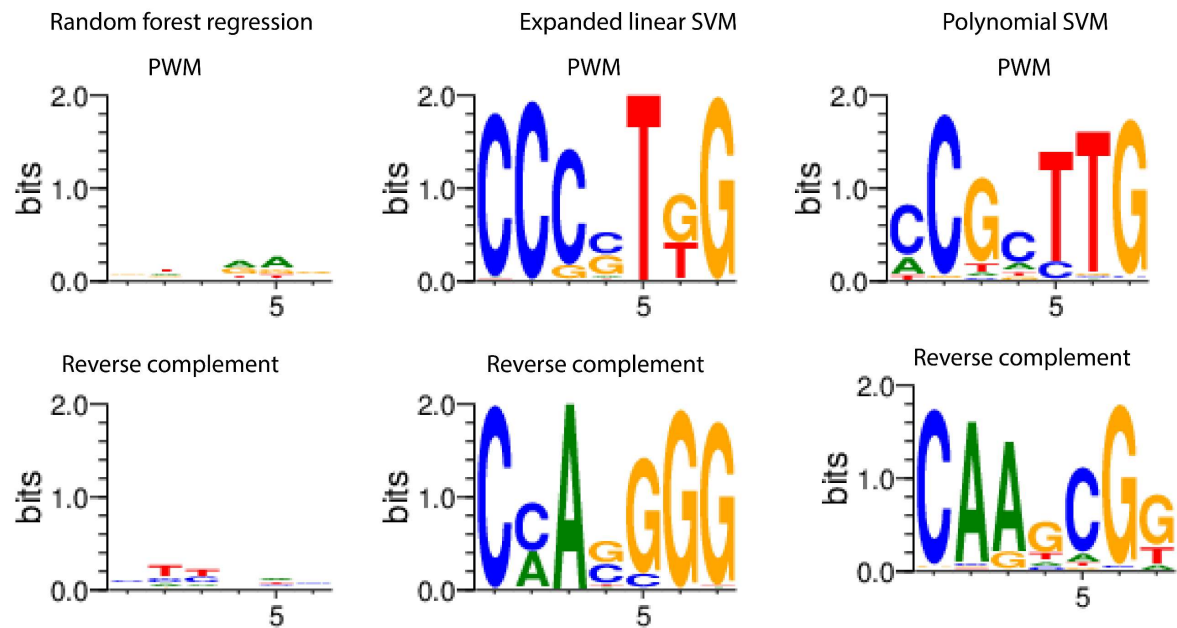


Figure S3.19 Pairwise sequence alignment of *pbx-1* and *pbx-2*. Coding regions are displayed in dark grey, intronic sequences in light grey, and variable sites in orange. A splice-site mutation in the first intron is predicted to cause a truncated form of *pbx-1*.

A SPILE-Z1



B PALE

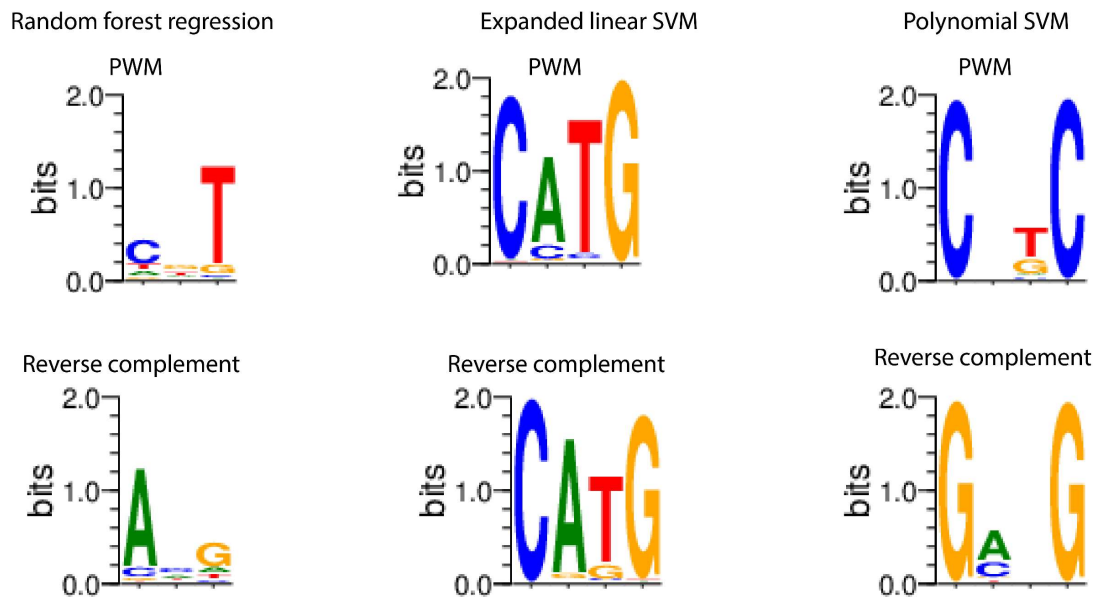


Figure S3.20 Prediction of DNA binding sites for zinc finger motifs in SPILE-Z1 and PALE in *S. mediterranea*. (A) Predicted binding sites for zinc finger motifs in SPILE-Z1 using random forest regression, expanded linear support vector machine (SVM), and polynomial SVM implemented in ZFPrinceton (Persikov et al., 2009; Persikov et al., 2014). (B) Predicted binding sites for zinc finger motifs in PALE using ZFPrinceton.

Chapter 4. A functional genomic approach infers TALE-dependent and TALE-independent regulatory mechanisms in planarian axial regeneration

Regeneration requires regulatory mechanisms that provide positional information to the regrowing structure. In planarians, regeneration is enabled by the coordinated interplay between wound-induced signalling, the resetting of the positional information system, and the proliferation and differentiation of neoblasts. After amputation, rescaled Wnt signalling levels provide positional identity to the anterior and posterior blastemas. Within the blastemas, a number of transcription factors are required for the formation of organising centres that direct morphogenesis. Among these transcription factors, the TALE homeobox genes *prep* and *pbx-1* are both required for the formation of the anterior pole, whereas *pbx-1* is also necessary for the specification of the posterior pole. Given the conserved role for another TALE subclass Meis in developmental programs, we asked whether members of the Meis gene family are required for axial regeneration, and discovered that *meis-2* is essential for restoring posterior poles. To understand the genes regulated by these TALE factors, we analysed the transcriptome and chromatin accessibility in response to TALE knockdowns to infer direct target genes. Our analysis suggests that *pbx-1* and *meis-2* are upstream regulators of *wnt1* at the posterior pole, while *pbx-1* and *prep* directly regulate anterior pole markers and FGFRL family members in anterior blastema. Furthermore, we explored our chromatin accessibility data in combination with established epistatic genetic interactions to infer high-confidence regulatory interactions between planarian patterning genes. Altogether, our findings expand our knowledge of the regulatory mechanisms that specify position during planarian regeneration.

4.1 Introduction

Successful regeneration relies on the specification of missing tissue in accordance with the polarity and dimensions of the remaining tissue fragment. Highly regenerative planarians have emerged as a leading model system to understand the cellular and molecular mechanisms that govern the identity and patterning of regenerating tissues. Planarians are able to regenerate any missing body part following injuries, thanks to a population of adult pluripotent stem cells and a dynamic positional information system that directs regenerative outgrowth. This system encompasses a molecular coordinate system of signalling molecules that direct the differentiation of neoblasts and their progeny along the anterior-posterior (AP), medial-lateral (ML), and dorsal-ventral (DV) axes. After injury, early planarian regeneration proceeds through three main phases: the generic phase involving a generic wound response (0 to 18 hpa or hours post amputation), a second phase of axis reestablishment (18 to 48 hpa), and the third phase of fate-specific cell differentiation and morphogenesis in the regenerating blastema (48 hpa onwards). Asymmetric expression of the Wnt inhibitor *notum* at anterior-facing wounds reduces Wnt signalling levels, and the level of Wnt signalling specifies the identity of a signalling centre known as blastemal poles that form at the tip of the regenerating blastema (Blassberg et al., 2013; Chen et al., 2013; Scimone et al., 2014a; Vasquez-Doorman and Petersen, 2014; Vogg et al., 2014; Li et al., 2019). In the anterior, the transcription factors *foxD*, *zic-1*, *prep*, and *pbx-1* are required for pole formation, while the transcription factors *pbx-1*, *pitx* and *islet1* are required for the formation of the posterior pole.

Among these transcription factors, *prep* and *pbx-1* belong to the subclass of TALE homeobox genes that are known for their roles in development and cell differentiation across the animal kingdom. In intact

planarians, the expression of *prep* is restricted to anterior areas, while *pbx-1* is expressed ubiquitously, including neoblasts, neoblast progeny, and the cephalic ganglia (Felix and Aboobaker, 2010; Blassberg et al., 2013). After injury, *prep* is expressed in both anterior and posterior blastemas, but it exerts its function in the anterior, as knockdown of *prep* results in cyclopic or eyeless worms but does not perturb tail regeneration (Felix and Aboobaker, 2010). Meanwhile, *pbx-1* is required for the specification of both anterior and posterior fates and for pharynx regeneration (Blassberg et al., 2013; Chen et al., 2013). In addition, *pbx-1* likely regulates some patterning genes along the DV axis, as knockdown of *pbx-1* results in the reduction of the dorsal marker *bmp4* and the ventral marker *admp*, but not *nlg-8*, a noggin-like gene involved in BMP signalling pathway (Chen et al., 2013). Given the critical role of *pbx-1* and *prep* in axial regeneration in planarians, this raises the question of whether other TALE factors also function as patterning factors, particularly the family of Meis genes, which are conserved in several developmental contexts across metazoans (Schulte and Geerts, 2019).

Although the identity of many genes required for axial regeneration has been uncovered, we know little about the underlying regulatory interactions between these genes, let alone the system-wide gene regulatory network. As in other model systems, the transcriptomic response following knockdowns has been used to identify downstream effectors of transcription factors and Wnt signalling components (Currie and Pearson, 2013; Cowles et al., 2014; Vogg et al., 2014; Reuter et al., 2015; Scimone et al., 2017; Tewari et al., 2019; Arnold et al., 2021; Pascual-Carreras et al., 2023). As for TALE factors, an exploratory transcriptomic analysis of anterior blastemas at 24 hpa in *prep(RNAi)* worms also identified Prep-dependent transcripts. However, as transcriptomics does not discriminate between direct and indirect regulatory interactions, studies have begun to incorporate epigenomic data to infer direct regulatory targets (Li et al., 2021; Neiro et al., 2022; Pascual-Carreras et al., 2023; Ivanković et al., 2024).

In this study, we uncovered a role for an additional TALE factor, *meis-2*, in posterior specification during regeneration. We employed bulk RNA-seq and ATAC-seq in wildtype and knockdown worms to extract the most likely direct targets. We find support for *pbx-1* and *meis-2* being direct regulators of the posterior pole marker *wnt1*, while *pbx-1* and *prep* are putative direct regulators of pole markers and the FGFR family member *ndk* in the anterior blastema. Additionally, we leverage our blastemal accessibility data to infer regulatory interactions between planarian patterning genes to establish a model for a gene regulatory network providing positional information in planarian regeneration.

4.2 Results

4.2.1 *meis-2* is required for posterior regeneration

Given that both *prep* and *pbx-1* are necessary for planarian axial regeneration (Figure 4.1A) and that our bioinformatic screening and phylogenetic analysis had identified additional TALE homeobox genes in the genome of *S. mediterranea* (Chapter 2, Chapter 3), we first sought to understand whether other TALE factors are required for either anterior or posterior regeneration. In particular, we were interested in the subfamily of Meis factors, known to dimerise with Pbx factors and to be involved in pattern formation across metazoans. Previously, the *meis-1* factor had been shown to be essential for eye regeneration in *S. mediterranea*, prompting us to investigate its paralogs.

Knockdown of *meis-2* resulted in the characteristic tailless phenotype, similar to the posterior defect observed after knockdown of *pbx-1* (Figure 4.1A). In planarians, the tailless phenotype is distinguished by the lack of the posterior pole, and the fusion of posterior nerve cords and digestive branches behind the pharynx (Suredda-Gómez et al., 2015). In the posterior end of *meis-2(RNAi)* regenerates, the ventral nerve cords did not extend to the posterior tip as seen in controls but were fused and rounded (Figure 4.1B), resembling phenotypes observed after knockdown of *pbx-1* (Blassberg et al., 2013) and other factors required for posterior pole specification (Adell et al., 2009; Gurley et al., 2010; Suredda-Gómez et al., 2015; Pascual-Carreras et al., 2021; Pascual-Carreras et al., 2023). Similarly, the branches of the gastrointestinal

Figure 4.1 *meis-2* is required for posterior regeneration. (A) Live images of *control(RNAi)*, *prep(RNAi)*, *pbx-1(RNAi)*, and *meis-2(RNAi)* worms ten days after amputation (10 dpa). Pre- and postpharyngeal transverse amputations divide the worm into regenerating head fragments (upper row), trunk or middle fragments (middle row), or tail fragments (lower row). After knockdown of *prep*, the worms regenerate as cyclops or without eyes altogether. After knockdown of *pbx-1*, regeneration is perturbed both at anterior and posterior wounds. After knockdown of *meis-2*, regeneration is perturbed in the posterior, while eyes do not regenerate in the anterior. **(B)** Immunohistochemistry of regenerating trunk fragments (15 dpa) after knockdown of *meis-2* indicate that ventral nerve cords fuse and do not extend to the posterior end (tailless phenotype). Whole-mount in situ hybridisation of the gut-specific marker *porcupine* revealed that the posterior gut branches also extended towards the midline after knockdown of *meis-2*. WISH of *wnt1* in *meis-2(RNAi)* worms indicate that the posterior pole does not form properly in the regenerating adult worms (trunk fragments 14 dpa). **(C)** The expression domain of the anterior marker *sFRP-1* varies, and the observed difference in *meis-2(RNAi)* worms as compared to control worms was interpreted to not be significant. WISH of the brain marker *PC2* demonstrate that cephalic ganglia regenerate after knockdown of *meis-2*. **(D)** The pharynx-specific marker *laminin* is expressed in control regenerated tail fragments but mostly absent in *meis-2(RNAi)* regenerates (12 dpa, WISH).

system extended towards the midline after *meis-2(RNAi)*, consistent with earlier reports of the tailless phenotype (Figure 4.1B) (Gurley et al., 2010). Lastly, the *wnt1*-expressing posterior pole was absent in the adult regenerates following knockdown of *meis-2* (Figure 4.1B).

In the anterior, silencing of *meis-2* resulted in eyeless worms, but the shape of the head was not affected to the same extent as after knockdown of either *prep* or *pbx-1* (Figure 4.1A). The expression of the anterior marker *sFRP-1* was unaffected in adult regenerates (Figure 4.1C), and no significant abnormalities were observed in the overall morphology of the cephalic ganglia (Figure 4.1C), suggesting that knockdown of *meis-2* does not impact general anterior patterning in the same way as *prep* and *pbx-1* (Felix and Aboobaker, 2010; Blassberg et al., 2013; Chen et al., 2013). Furthermore, tail fragments of *meis-2(RNAi)* worms failed to regenerate the pharynx, indicating an additional role in pharynx specification (Figure 4.1D).

Although the replicate numbers for individual stainings are low, we interpret the totality of the data to suggest that *meis-2* is required for posterior specification during regeneration and the specification of eyes. Since *meis-2(RNAi)* worms were able to regenerate anterior cell types, the observed phenotype is not a result of a general defect in neoblast maintenance, proliferation, or differentiation. Altogether, our characterisation of *meis-2* expands the repertoire of TALE factors required for axial regeneration in planarians, revealing two pairs of co-acting TALE factors: *prep* and *pbx-1* in the anterior, and *meis-2* and *pbx-1* in the posterior.

4.2.2 Anterior pole markers and FGFR genes are downregulated after knockdown of *pbx-1* and *prep*

Since knockdown phenotypes indicate that *prep* and *pbx-1* are necessary for anterior patterning and *meis-2* appears to be involved specifically in eye regeneration, we sequenced the transcriptomes of anterior blastemas and compared gene expression in control and knockdown worms to begin to understand the genes directly regulated by these TALE factors (Figure 4.2A). Although the knockdown phenotypes of *prep* and *pbx-1* suggest that these factors are involved in the specification of the anterior pole in late-stage blastemas, we performed bulk RNA-seq at both early (18 hpa, hours post amputation) and late phases (60 hpa) to identify acutely misregulated genes. We followed a standard two-week microinjection schedule for

RNA interference, amputated the worms into head, trunk, and tail fragments, dissected the anterior blastema from tail fragments at either 18 or 60 hpa, and extracted RNA from tissues following standard protocols (Figure 4.2A). If the silenced trunk fragment replicated the patterning phenotype, we deduced that knockdown had been successful also in the tail fragment and sequenced RNA samples (Figure 4.2A). Altogether, our experimental setup generated a set of six conditions for differential gene expression analysis in the anterior blastema (*control(RNAi)* vs. *pbx-1(RNAi)* at 18 hpa and 60 hpa, *control(RNAi)* vs. *prep(RNAi)* at 18 and 60 hpa, *control(RNAi)* vs. *meis-2(RNAi)* at 18 and 60 hpa) (Figure 4.2D1, D2, E1, E2, F1, F2).

In the early blastema, we observed that most genes were downregulated after knockdown of *prep* (Figure 4.2E1: 163 DEGs, 84 down and 79 up), followed by *pbx-1* (Figure 4.2D1: 55 DEGs, 34 down and 21 up) and *meis-2* (Figure 4.2F1: 48 DEGs, 41 down and 7 up). Surprisingly, significantly more genes were misregulated in the late blastema after silencing *meis-2* (Figure 4.2F2: 1964 DEGs, 1328 down and 636 up) as compared to *pbx-1* (Figure 4.2D2: 54 DEGs, 33 down and 21 up) and *prep* (Figure 4.2E2: 18 DEGs, 14 down and 4 up). Although Pbx proteins are known to dimerise with Prep and Meis and coregulate target genes, we did not detect many shared DEGs between *pbx-1(RNAi)* and *prep(RNAi)*, or between *pbx-1(RNAi)* and *meis-2(RNAi)* (Figure 4.2B). Therefore, the quantity of differentially expressed genes does not correlate with the severity of the observed knockdown phenotype, and we focussed on well-known patterning genes and transcription factors to infer genes explanatory for the anterior defect present in *prep(RNAi)* and *pbx-1(RNAi)* worms. For this purpose, we used the list of transcription factors generated in our previous bioinformatic screening (Chapter 2), and compiled a similar list of planarian patterning genes. We reviewed the literature and assigned the patterning genes to genes in our genome annotation (Chapter 2 Neuro et al. 2022) by using BLAST.

As in situ hybridization experiments had produced contrasting results regarding whether early wound-induced *notum* expression is affected following knockdown of *pbx-1* (Blassberg et al., 2013; Chen et al., 2013), we examined the response of *notum* separately (Figure 4.2C). In the early anterior blastema, knockdown of *pbx-1* slightly reduced the expression of *notum* ($\log_2\text{fc} = 1$) but not in a statistically significant manner due to high variance, while *notum* was downregulated in the early anterior blastema of *meis-2(RNAi)* worms (Figure 4.2C). As RNA-seq has indicated that *notum* is upregulated by at least $\log_2\text{fc} = 3$ in anterior blastemas compared to the posterior, our findings suggest that although knockdown of *pbx-1* and *meis-2* downregulate *notum* in the anterior to some degree, this reduction is not sufficient to reverse polarity.

As suggested by earlier experiments, anterior pole markers (*notum* and *sFRP-1*) were misregulated in the late blastemas after silencing *prep* and *pbx-1*, while knockdown of *pbx-1* additionally downregulated the pole-specific transcription factors *foxD* and *zic-1* (Figure 4.2D1, D2, E1, E2). Interestingly, *notum* was also downregulated in the late-stage blastema after knockdown of *meis-2*, but other anterior pole markers were not dysregulated (Figure 4.2F1, F2). Furthermore, knockdown of *prep* and *pbx-1* downregulate the anterior patterning genes *ndk* and *ndl-4* of the FGFR-like family in both the early- and late-stage blastemas (Figure 4.2D1, D2, E1, E2). Notably, these FGFR-like genes were not downregulated by inhibition of *meis-2* in neither early- nor late-stage blastemas (Figure 4.2F1, F2). Previous ISH experiments had revealed that expression of the prepharyngeal marker *wnt2* is ectopically present in the anterior blastema of *pbx-1(RNAi)* regenerates at 6 dpa (Chen et al., 2013). In contrast, we found that *wnt2* is downregulated after knockdown of both *pbx-1* and *meis-2* (Figure 4.2D2, F2), putatively suggesting *wnt2* is first downregulated in the blastema/postblastema boundary at 3 dpa before being ectopically expressed in the anterior of *pbx-1(RNAi)* worms.

Concurrently, some genes of the posterior specification program (including *axinB*, *sp5*, *post2c*, *box4b*, *dd13065*, *dd4427*) (Tewari et al., 2019; Pascual-Carreras et al., 2023) were upregulated after inhibiting *pbx-1*

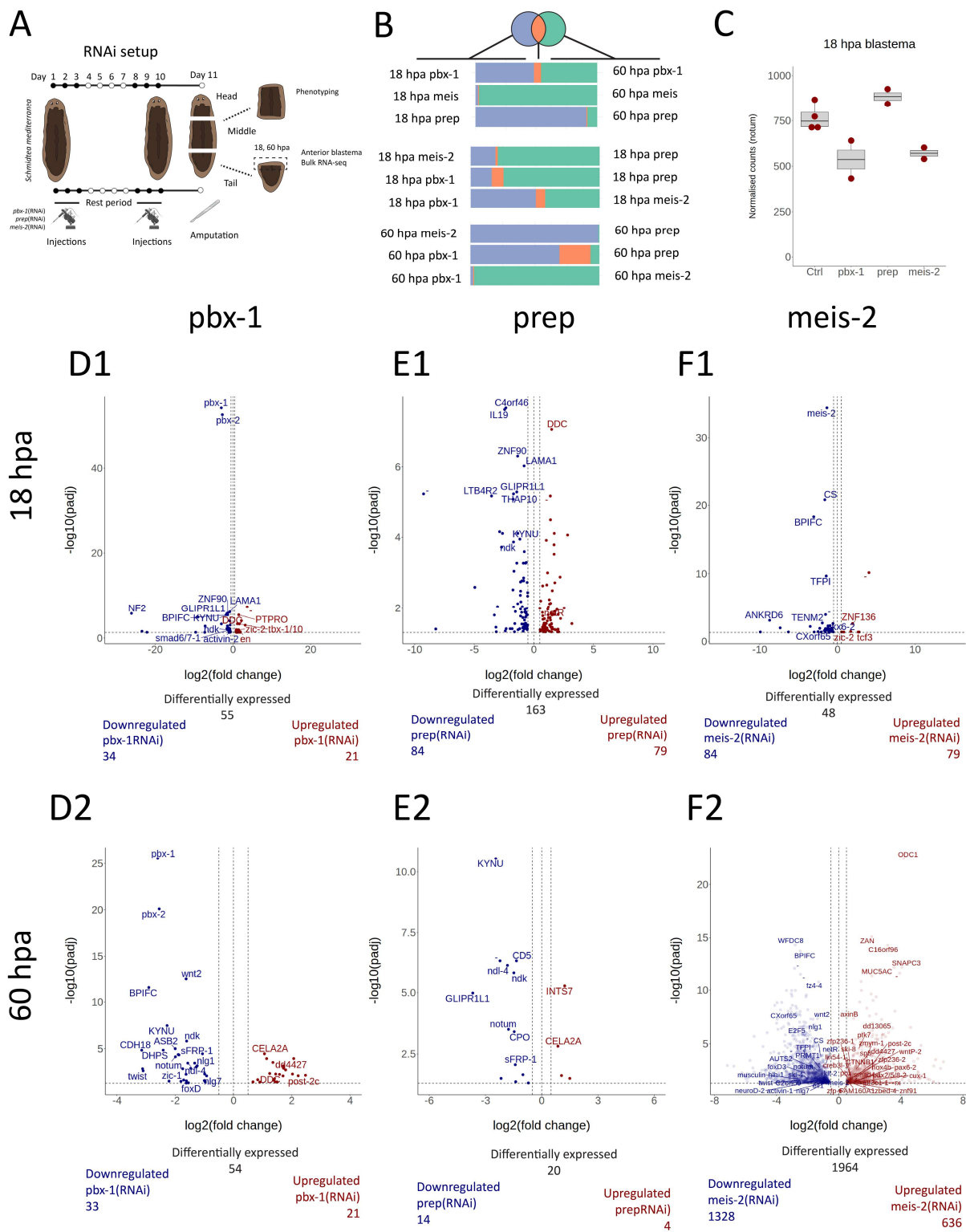


Figure 4.2 Differential gene expression analyses of *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)* in the anterior blastema. (A) Experimental setup. RNA sequencing was performed on dissected anterior blastemas after a standard ten-day RNAi schedule. Worms were injected at days 1,2,3, rested for four days, and injected again at days 8,9,10. The trunk fragment was used to confirm that knockdown had been successful before sequencing samples. **(B)** Linearised Venn diagrams displaying the proportional overlap of DEGs between conditions (orange) and DEGs specific to each condition. **(C).** The expression level of the gene *notum* in different knockdown conditions **(D, E, F)**. Volcano plots showing differentially expressed genes in the anterior blastema after knockdown of *pbx-1* at the early stage (18 hpa) **(D1)**, after knockdown of *pbx-1* at the late stage (60 hpa) **(D2)**, after knockdown of *prep* at the early stage (18 hpa) **(E1)**, after knockdown of *prep* at the late stage (60 hpa) **(E2)**, after knockdown of *meis-2* at the early stage (18 hpa) **(F1)**, and after knockdown of *meis-2* at the late stage (60 hpa) **(F2)**. In the volcano plots, the x-axis shows log₂ fold change (log₂FC) of expression levels between the knockdown and control worms. Positive log₂FC values indicate that transcripts were expressed at higher levels in the knockdown worms than in the controls (upregulated), while negative log₂FC values indicate that the transcripts were expressed at lower levels in knockdown worms (downregulated). The plots display names of planarian patterning genes, transcription factors, the ten most significantly misregulated genes, and genes shared between early and late timepoints for each knocked down TALE gene. Planarian patterning genes and transcription factors are written in lowercase, while the name of the best BLAST match to the human genome is shown in uppercase for other genes. If BLAST was not able to align the planarian transcript to a human gene, this is indicated with a dash “-“. Only differentially regulated genes are displayed in the plots (cutoffs log₂FC > 0.5, padj < 0.05). To concisely display all DEGs, the axes have been truncated in *meis-2(RNAi)* 18 hpa **(F1)** and 60 hpa **(F2)**, leaving some unnamed DEGs outside the scope of the plot.

and *meis-2*. This could indicate that knockdown of *pbx-1* and *meis-2* either posteriorise the regenerating blastema to a certain degree, or homeostatically modifies the pre-existing postblastemal region during the RNAi protocol (Figure 4.2E1, E2). With respect to signalling molecules of the DV axis, we did not detect dysregulation of *bmp4* and *admp* (Chen et al., 2013), but we observed that *nlg1* and *nlg7* were downregulated in both *pbx-1(RNAi)* and *meis-2(RNAi)* worms. In the anterior blastema, *nlg1* is expressed in two clusters around the pole at 72 hpa (Molina et al., 2009), while *nlg7* is expressed in the lateral regions of the regenerating blastema (Molina et al., 2009). Nonetheless, knockdown of neither *nlg1* nor *nlg7* generates a phenotype (Molina et al., 2011). As for Activin signalling, *activin-2* is downregulated in early blastemas after knockdown of *pbx-1*, while *activin-1* is downregulated in the late blastema following knockdown of *meis-2*. Knockdown of *activin-2* generates two-headed worms as a result of pole splitting (Cloutier et al., 2021), while knockdown of *activin-1* increases the number of *ovo+* progenitors in the anterior blastema. As knockdown data indicate that *pbx-1* is required for anterior pole formation and *meis-2* is required for eye formation (*ovo+* cells), this suggests that *activin-2* and *activin-1* are acting downstream of these events.

Interestingly, we did not find any previously undocumented transcription factors in the late-stage blastemas after knockdown of *pbx-1* and *prep* (Figure 4.2D2, E2). In the late blastema, knockdown of *pbx-1* downregulated *twist*, which is expressed in the pharynx and in a subset of neoblasts, and is presumed to be required for pharynx regeneration (Scimone et al., 2014b) (Figure 4.2D2). Thus, the loss of a pharynx in *pbx-1(RNAi)* worms could be mediated through the dysregulation of *twist* (Scimone et al., 2014b). In the early blastema, knockdown of *pbx-1* downregulates neoblast-expressed *smad6/7-1* (van Wolfswinkel et al., 2014), the anteriorly expressed *engrailed (en)* (Currie et al., 2016), and a previously undescribed transcription factor *tbx-1/10* (Figure 4.2D1). In the early blastema, silencing of *meis-2* downregulates the muscle-specific gene *nkx6-2* that is expressed in longitudinal muscle fibres (Scimone et al., 2017; Scimone et al., 2018), raising the possibility that *meis-2* might regulate some aspect of longitudinal fibre biology.

In contrast, multiple transcription factors were differentially expressed in late-stage blastemas after knockdown of *meis-2*. This set included neuronally associated factors, including *pax6-2* (Wenemoser et al., 2012; Scimone et al., 2014b), *musculin* (Cowles et al., 2013), *blb-1* (Cowles et al., 2013; Roberts-Galbraith et al., 2016), and *neuroD-2* (Cowles et al., 2013). Among the uncharacterised homeobox genes is a putative homolog of *rx* or *rax* (retinal homeobox protein or retina and anterior neural fold homeobox gene), which is required for photoreceptor differentiation in vertebrates (Nelson et al., 2009). This raises the possibility that the loss of eyes following knockdown of *meis-2* could be mediated by this downstream factor. Lastly, knockdown of *meis-2* downregulated a suite of uncharacterised zinc fingers (*zfp236-1*, *zfp236-2*, *zfp-6*, *zbed-4*, and *zmf91*), which could be additional regulators of eye progenitors or other neuronal subpopulations.

Altogether, although various patterning genes and transcription factors were downregulated following knockdown of *prep*, *pbx-1* and *meis-2*, we infer that the DGE pattern indicates the difference in the phenotypes of *prep* and *pbx-1*, as compared to *meis-2*, lies in the differential regulation of anterior pole markers and genes of the FGFR circuit in the anterior.

4.2.3 The posterior specification program is downregulated after knockdown of *pbx-1* and *meis-2*

Similarly, to gain further insight into the targets of TALE factors in tail regeneration, we performed bulk RNA-seq on posterior blastemas. The knockdown phenotypes indicate that *pbx-1* and *meis-2* are involved in posterior regeneration, but we also sequenced the transcriptome of *prep(RNAi)* worms to provide a point of comparison for analysis. As silencing of *pbx-1* and *meis-2* does not reverse polarity, they likely exert the main part of their function at the late blastemal stage. Nonetheless, we again sequenced both early-stage blastemas (18 hpa) and late-stage blastemas (60 hpa) to more effectively identify direct targets, as in the anterior. Again, we followed the standard RNA interference protocol, extracted posterior blastemas from head fragments, and isolated RNA (Figure 4.3A). If the left-aside trunk fragments exhibited the desired phenotype implying successful knockdown, we processed the samples for bulk RNA sequencing (Figure 4.3A). Again, the experimental setup included six conditions for differential gene expression analysis (Figure 4.3D1, D2, E1, E2, F1, F2).

In the early-stage blastema, we observed that most genes were downregulated after knockdown of *prep* (Figure 4.3E1: 103 DEGs, 61 up and 42 down), followed by *meis-2* (Figure 4.3F1: 34 DEGs, 20 down and 5 up) and *pbx-1* (Figure 4.3D1: 27 DEGs, 18 down and 9 up). In the late blastema, a much greater number of transcripts were downregulated after inhibition of *meis-2* (Figure 4.3F2: 883 DEGs, 448 down and 385 up) than after *pbx-1* (Figure 4.3D2: 22 DEGs, 16 down and 6 up) or *prep* (Figure 4.3E2: 19 DEGs 7 down and 12 up). Since knocking down *meis-2* in the posterior misregulates many more genes than *pbx-1* despite a similar severity in the posterior phenotype, we again focussed our analysis on patterning genes and transcription factors as in the anterior. As suggested by the knockdown phenotype, silencing of *prep* did not result in the misregulation of patterning genes or transcription factors (Figure 4.3 E1 and E2).

First, we looked at wound-induced *wnt1* expression in the posterior blastema, as earlier ISH stainings have indicated that early *wnt1* expression is not affected by knockdown of *pbx-1* (Blassberg et al., 2013; Chen et al., 2013). Our RNA-seq data reproduces this finding, as the expression level (normalised counts) of *wnt1* is not affected by knockdown of *pbx-1* (Figure 4.3C). On the contrary, *wnt1* expression is significantly reduced after knockdown of *meis-2* already in the early-stage blastema, but as reversal of polarity was never observed in *meis-2* worms after knockdown (Fig 1 A), this reduction in expression levels appears to not be sufficient to reverse polarity.

In the late blastema, knockdown of both *pbx-1* and *meis-2* downregulate *wnt1* (expressed specifically in the posterior pole), and other genes of the posterior specification program (including *wnt11-1*, *wnt11-2*, *fzd4-1*, *dd13065*, *fzd4-1*, *wntP-2*, *fzd4-2*, *dd4407*, *sp5*, *lox5a*) (Figure 4.3D2, F2) that are also downregulated by *beta-catenin-1(RNAi)* and *wnt1(RNAi)* (Tewari et al., 2019; Pascual-Carreras et al., 2023). Previous ISH

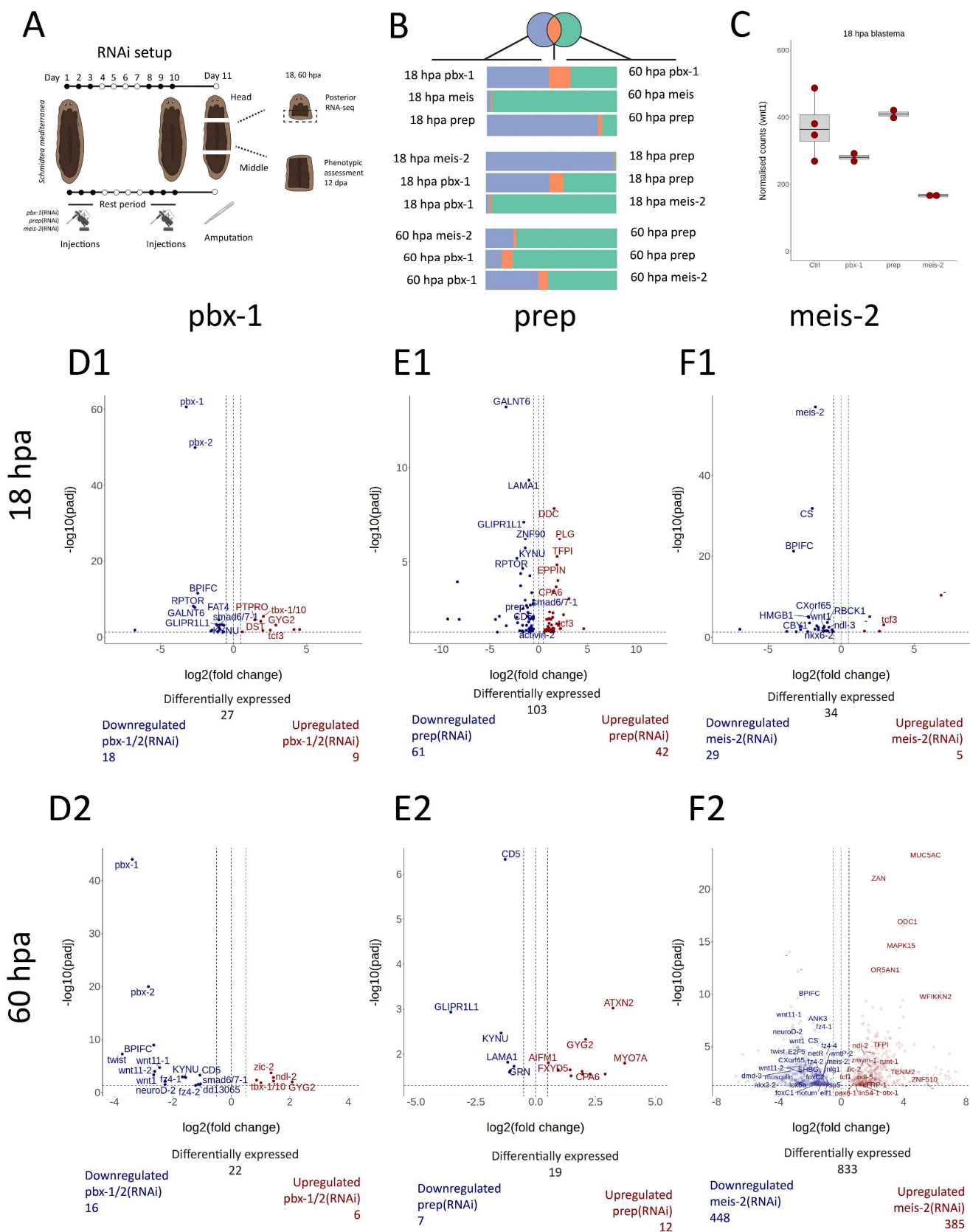


Figure 4.3 Differential gene expression analyses of *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)* in the posterior blastema. (A) Experimental setup. RNA sequencing was performed on dissected posterior blastemas from head fragments after the standard ten-day RNAi schedule. (B) Linearised Venn diagrams displaying the proportional overlap of DEGs between conditions (orange) and DEGs specific to each condition. (C). The expression level of the gene *wnt1* in different knockdown conditions (D, E, F). Volcano plots showing differentially expressed genes in the posterior blastema after knockdown of *pbx-1* at the early stage (18 hpa) (D1), after knockdown of *pbx-1* at the late stage (60 hpa) (D2), after knockdown of *prep* at the early stage (18 hpa) (E1), after knockdown of *prep* at the late stage (60 hpa) (E2), after knockdown of *meis-2* at the early stage (18 hpa) (F1), and after knockdown of *meis-2* at the late stage (60 hpa) (F2). In the volcano plots, the x-axis shows log2 fold change (log2FC) of expression levels between the knockdown and control worms. Positive log2FC values indicate that transcripts were expressed at higher levels in the knockdown worms than in the controls (upregulated), while negative log2FC values indicate that the transcripts were expressed at lower levels in knockdown worms (downregulated). The plots display names of planarian patterning genes, transcription factors, the ten most significantly misregulated genes, and genes shared between early and late timepoints for each knocked down TALE gene. Planarian patterning genes and transcription factors are written in lowercase, while the name of the best BLAST match to the human genome is shown in uppercase for other genes. If BLAST was not able to align the planarian transcript to a human gene, this is indicated with a dash “-“. Only differentially regulated genes are displayed in the plots (cutoffs log2FC > 0.5, padj < 0.05). To concisely display all DEGs, the axes have been truncated in *meis-2(RNAi)* 18 hpa (F1), leaving some unnamed DEGs outside the scope of the plot.

experiments had revealed that *pbx-1(RNAi)* abolishes *wnt1* expression in 2 dpa and 4 dpa posterior blastemas (Blassberg et al., 2013; Chen et al., 2013). Interestingly, *notum* was also downregulated in the late blastema following knockdown of *meis-2(RNAi)*, but to a lower degree than *wnt1* (Figure 4.3F2). Furthermore, knockdown of *pbx-1* and *meis-2* appears to anteriorise the blastema or pre-existing postblastemal tissues during the RNAi protocol, as some anterior PCGs are upregulated (*pbx-1(RNAi)*: *zic-2*, *ndl-2*; *meis-2(RNAi)*: *ndl-2*, *zic-1*, *ndl-5*, *sFRP-1*). Additionally, *meis-2(RNAi)* upregulates the transcription factor *otx-1* that has been implicated in photoreceptor specification (Figure 4.3F2) (Lapan and Reddien, 2011; Lapan and Reddien, 2012; Wenemoser et al., 2012; Zhu et al., 2015; Roberts-Galbraith et al., 2016). Silencing of *meis-2* also downregulates the prepharyngeal marker *ndl-2*.

Regarding transcription factors, the pharynx-associated transcription factor *twist* is downregulated in late-stage blastemas after knockdown of both *pbx-1(RNAi)* and *meis-2(RNAi)* as in anterior blastemas (Figure 4.3D2, F2). As in the early-stage anterior blastema, *pbx-1(RNAi)* upregulated *tbx-1/10* both in early- and late-stage posterior blastemas (Figure 4.3D2). Moreover, knockdown of *meis-2* also misregulated a suite of additional transcription factors, including the muscle-associated *nkx6-2* as in the anterior, neuronally associated transcription factors *musculin* (Cowles et al., 2013), *neuroD-2* (Cowles et al., 2013), *pax6-1* (Wenemoser et al., 2012; Scimone et al., 2014b; Fincher et al., 2018), two Fox genes (*foxC1* and *foxC2*) shown to display mild regeneration defects (Pascual-Carreras et al., 2021), and the previously uncharacterised *nkx3-2*.

Overall, we interpret the transcriptomes of *pbx-1(RNAi)* and *meis-2(RNAi)* worms to indicate that the tailless phenotype is explained by dysfunctional regulation of the posterior pole marker *wnt1* and other genes of the downstream posterior specification program.

4.2.4 Chromatin accessibility and footprinting analysis predicts direct targets of planarian TALE genes

To explore and reduce our list of candidate downstream targets, we optimised and performed bulk ATAC-seq to infer direct regulatory interactions. Accessible regions in chromatin can be used to identify cis-regulatory genomic regions, these regions can be linked to target genes, and the presence of transcription factor binding sites within these regions can be used to infer a regulatory interaction between the transcription factor and the target gene. In the past few years, ATAC-seq has been performed on planarian early-stage blastemas (0, 6, and 12 hpa) (Pascual-Carreras et al., 2023), stem cells (Li et al., 2021; Neuro et al., 2022; Poulet et al., 2023), tissue samples from the brain, epidermis, and intestine (Li et al., 2021; Poulet et al., 2023), and whole adult worms (Ivanković et al., 2024). We optimised and modified the Omni ATAC-seq protocol and performed ATAC-seq in anterior and posterior blastemas of controls and *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)* worms (Figure 4.4A). We performed ATAC-seq in late stage blastemas (60 hpa), as the knockdown phenotypes and DGE analysis indicate that the TALE factors are required for pole formation at these stages, and because ATAC-seq data was being independently generated for early-stage blastemas (Pascual-Carreras et al., 2023).

To first rigorously assess the quality of our data, we analysed publicly available datasets from planarians and compared established quality metrics for ATAC-seq data (Gehrke et al., 2019; Li et al., 2021; Neuro et al., 2022; Pascual-Carreras et al., 2023). As these samples have been generated in different tissue types, a strict comparison of experimental protocols is not possible, but nevertheless informs on sample quality. The insert size distribution of our ATAC-seq data displays a decreasing and periodical pattern, with two clear peaks corresponding to nucleosome-free and mono-nucleosome bound fragments (Figure 4.4B). The insert size distribution is comparable to those of publicly available datasets, although our samples do not capture di-nucleosomal reads as efficiently as some other samples (Figure 4.4C1, C2, C3, C4). Although the alignment rates of our ATAC-seq samples vary, they are within the range of publicly available datasets (Figure 4.4 D). Similarly, the percentage of deduplicated reads varies, and slightly more duplicate reads are on average present in our samples (Figure 4.4D). In contrast, the FriP score (fraction in peaks) is slightly higher in our samples, indicating the presence of more distinct and defined peaks, i.e. accessible regions, in our ATAC-seq samples (Figure 4.4D).

To infer direct target genes, we called peaks in the merged anterior and posterior wildtype ATAC-seq samples (Figure 4.4E, F). Then, we assigned these accessible regions to the nearest gene and selected regions linked to genes that were differentially regulated following *prep(RNAi)*, *pbx-1(RNAi)*, and *meis-2(RNAi)*. In these selected regions, we leveraged the high sequencing depth of our ATAC-seq samples and performed footprinting analysis to identify binding activity of TALE factors (Bentsen et al., 2020). In computational ATAC-seq analysis, footprinting is an approach that discovers small gaps in the ATAC-seq signal indicative of bound transcription factors (Bentsen et al., 2020). Since no motifs have been experimentally validated for planarian TALE factors, we used the Pbx, Prep, and Meis motifs available in the Jasp database (Rauluseviciute et al., 2024). If footprinting analysis indicated that a TALE motif was present and putatively bound by a transcription factor in the accessible region, we could construct putative regulatory interactions between the regulating TALE factor and its putative downstream target. To further rank these regulatory interactions, we designed an average percentile rank score taking into account four percentile ranks: (1) the log2 fold change in the putative target gene following knockdown of the TALE factor, (2) the adjusted p-value from differential gene expression analysis, (3) the absolute binding strength of the TALE factor at the motif in the target tissue, (4) the differential binding strength between the target and non-target tissues (anterior vs. posterior) (Methods) (Figure 4.4E).

Our footprinting analysis revealed that *ndk* and *ndl-4* are putative direct targets of both *pbx-1* and *prep* in the

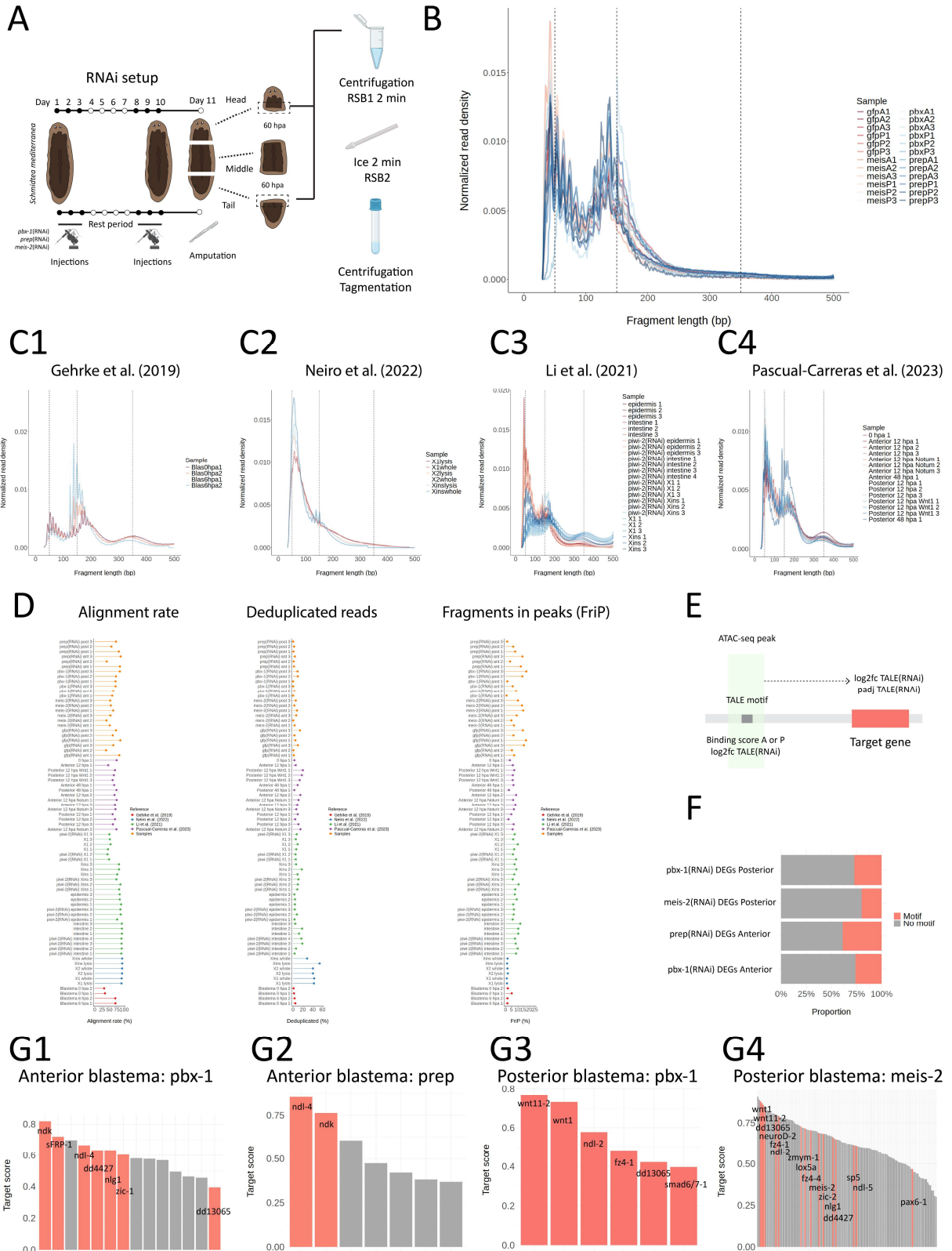


Figure 4.4 ATAC-seq analysis and prediction of direct targets of *prep*, *pbx-1*, and *meis-2*. (A) Schematic illustration of the experimental workflow to generate ATAC-seq data in knockdown worms. We generated ATAC-seq data from anterior and posterior late-stage blastemas (60 hpa, 3 replicates per condition) in *control(RNAi)*, *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)* worms. The blastemas were centrifuged, incubated in resuspension buffer 1, homogenised mechanically with a pestle, incubated on ice, RSB2 buffer was added to homogenate, which was filtered through a cell strainer. The filtrate was centrifuged and processed according to standard protocols (Buenrostro et al., 2013; Corces et al., 2017). (B) Insert size distribution of blastemal ATAC-seq samples. Enrichment of reads at 50 and 150 bp reflect nucleosome-free and mono-nucleosome-bound fragments. (C1) Insert size distributions of publicly available 0 hpa and 6 hpa blastemal samples (mixture of anterior and posterior blastemas) (Gehrke et al., 2019). (C2) The insert size distributions of X1, X2, and Xins samples (Neiro et al., 2022). (C3) Insert size distributions of ATAC-seq samples from epidermis, intestine, X1 cells, and Xins cells in wildtype worms and after knockdown of *pimi-2* (Li et al., 2021). (C4) Insert size distributions of blastemal samples (0 hpa blastemas, 12 hpa anterior blastemas, 12 hpa anterior *notum(RNAi)* blastemas, 12 hpa posterior blastemas, 12 hpa posterior *wnt1(RNAi)* blastemas, and 48 hpa anterior and posterior blastema (Pascual-Carreras et al., 2023). (D) Alignment rate (D1), proportion of reads remaining after deduplication, i.e. deduplicated reads (D2), and FriP score (fragments in peaks) (D3) in ATAC-seq samples as compared to the publicly available datasets in (C). (E) Schematic illustration of the computational pipeline to predict direct targets for *pbx-1*, *prep*, and *meis-2*. Wildtype ATAC-seq data was used to call peaks in the genome, and the peaks were assigned to the nearest gene (measured by genomic distance). Percentile ranks were calculated for the log2 fold change of the target gene after knockdown of the TALE gene of interest, the adjusted p value of differential expression after knockdown, the footprint binding score in the wildtype tissue of interest (anterior or posterior), and the differential binding score after knockdown of the TALE of interest. The target score is an average of these percentile ranks and used for ranking targets in (G). (F) Proportion of DEGs that were predicted to be direct targets (“Motif”) as compared to non-targets (“No motif”). (G) Ranking of direct targets of *pbx-1* in the anterior blastema (G1), of *prep* in the anterior blastema (G2), of *pbx-1* in the posterior blastema (G3), and of *meis-2* in the posterior blastema (G4). Planarian patterning genes are labelled and their rank is marked with orange.

anterior blastema, while the pole markers *sFRP-1* and *zic-1* are also direct targets of *pbx-1* (Figure 4.4G1, G2). In the posterior, *pbx-1* and *meis-2* share as predicted targets the posterior pole marker *wnt1*, the more broadly expressed *wnt11-2*, the receptor *fr4-1*, and the transcript *dd13065* (Figure 4.4G3, G4). Additional components of the posterior specification program are also predicted to be targets of *meis-2*, including *lox5a*, *sp5*, and *dd4427* (Figure 4.4).

4.2.5 Functional characterisation and RNAi screening informs on the regulatory mechanisms of TALE factors

Our analysis suggests that both *pbx-1* and *meis-2* are potential direct regulators of *wnt1* at the posterior pole during regeneration. Since loss of the posterior pole leads to tailless phenotype as demonstrated by knockdown of *wnt1*, *wnt11-2*, *pitx*, and *islet* (Currie and Pearson, 2013; März et al., 2013; Sureda-Gómez et al., 2015), the tailless knockdown observed in *pbx-1(RNAi)* and *meis-2(RNAi)* worms is consistent with *pbx-1* and *meis-2* being direct regulators of *wnt1*. Additionally, previous knockdown experiments have shown that *pbx-1* is essential for the formation of the posterior pole (2 and 4 dpa), providing functional evidence that *pbx-1* could act as a direct target of *wnt1* (Blassberg et al., 2013; Chen et al., 2013). Since *pbx-1* and *meis-2* are able to dimerise and coregulate genes (Longobardi et al., 2014), this further suggests that *pbx-1* and *meis-2* could function together in regulating the expression of *wnt1* at the posterior pole.

To characterise the putative regulatory interaction between *pbx-1/meis-2* and *wnt1*, we localised the footprints to the first intron of *wnt1* (Figure 4.5A). Consistent with previous observations, we identified two accessible regions in the first intron of *wnt1* (Figure 4.5A) (Neiro et al., 2022; Pascual-Carreras et al., 2023; Ivanković et al., 2024). These regions are the best characterised putative cis-regulatory elements in the genome of *S. mediterranea*. Their cis-regulatory status is supported by conservation of accessibility between planarian species and the presence of a foxG motif, which has been shown to regulate the expression of *wnt1* in the posterior blastema (Pascual-Carreras et al., 2023; Ivanković et al., 2024).

To further investigate the cis-regulatory dynamics of the *wnt1* gene, we analysed all footprints in the surrounding regions and ranked their potential as direct regulators of *wnt1*. Using all available metazoan transcription factors from the Jasp database (Rauluseviciute et al., 2024), we calculated percentile ranks based on their absolute binding strength (footprint score) in wildtype posterior blastemas and their differential binding strength compared to the anterior blastema. After averaging the percentile ranks (target score), we observed that motifs associated with Meis, Pbx, and foxG factors were among the highest-ranked (Figure 4.5E). Therefore, our analysis provides additional support for foxG being a direct regulator of *wnt1*, as not only is a foxG motif present in the first intron of *wnt1*, but this motif is also among the most strongly bound footprints, indicating direct binding of the foxG factor (Figure 4.5E). Additionally, as the high-ranked motif (MA1640.1, the “OCTA” motif) has been demonstrated to be bound by a MEIS:PBX complex (Penkov et al., 2013), this further supports *pbx-1* and *meis-2* being direct regulators of the *wnt1* gene.

As *pbx-1* has been shown to be necessary for the formation of the posterior pole, we sought to experimentally validate the role of *meis-2*. Although the number of replicates is low, ISH experiments suggest that knockdown of *meis-2* abolishes *wnt1* expression in the posterior pole at 60 hpa (Fig. 4.5C). To provide evidence for coexpression, we analysed publicly available scRNA-seq data from 72 hpa trunk fragments (including both anterior and posterior blastemas) (Cui et al., 2023a), and observed coexpression in muscle cells, further supporting that *meis-2* acts as a direct regulator of *wnt1* (Figure 4.5E).

In the anterior blastema, *ndk* was identified as a candidate direct target of both *prep* and *pbx-1*. The motifs classified as Prep or Pbx in the Jasp database correspond to the “DECA” motif, which has been found to be bound by monomeric Prep or a dimeric Prep:Pbx complex (Penkov et al., 2013). This suggests that *pbx-1* and *prep* could co-regulate *ndk* in planarians (Fig. 4.5B). Previous double knockdown experiments demonstrate that *prep;ndk(RNAi)* double knockdown worms have fused cephalic ganglia at anterior tips but does not rescue the posterior expansion of brain tissues observed in *ndk(RNAi)* worms, suggesting that *ndk* may be downstream of *prep* (Felix and Aboobaker, 2010). Similarly, *pbx-1;ndk(RNAi)* double knockdown worms also display posterior expansion of brain tissues, suggesting that *ndk* acts downstream of *pbx-1* (Blassberg et al., 2013). In the genomic region of the *ndk* gene, we found that the strongest bound Prep and Pbx motifs were localised to the promoter region. Additionally, we found that footprints for Prep (PKNOX1_MA0782.2) and Pbx (PBX3_MA1114.1) were among the highest ranked among all metazoan motifs, and scRNA-seq data from 72 dpa trunks displayed co-expressing cells (*prep* and *ndk*) in the muscle and neuronal clusters (Figure 4.5E). Although the number of replicates is low, ISH experiments suggest that *ndk* expression is reduced following *prep(RNAi)* (Figure 4.5D). Altogether, this evidence supports that *prep* acts as a direct regulator of *ndk*, possibly in a dimeric complex with *pbx-1*.

In our DGE analysis, we observed a handful of uncharacterised misregulated genes. We found that the transcript *dd13065* is a putative direct target of both *pbx-1* and *meis-2*. Although *dd13065* has been shown to exhibit a posteriorly biased expression pattern in adult worms and to be downregulated following reduced Wnt signalling levels, no functional characterisation appears to have been reported (Reuter et al., 2015; Scimone et al., 2016; Stückemann et al., 2017). We queried the *dd13065* transcript using BLAST and Interproscan and observed that it contains a DEATH domain and is a putative tumour necrosis factor

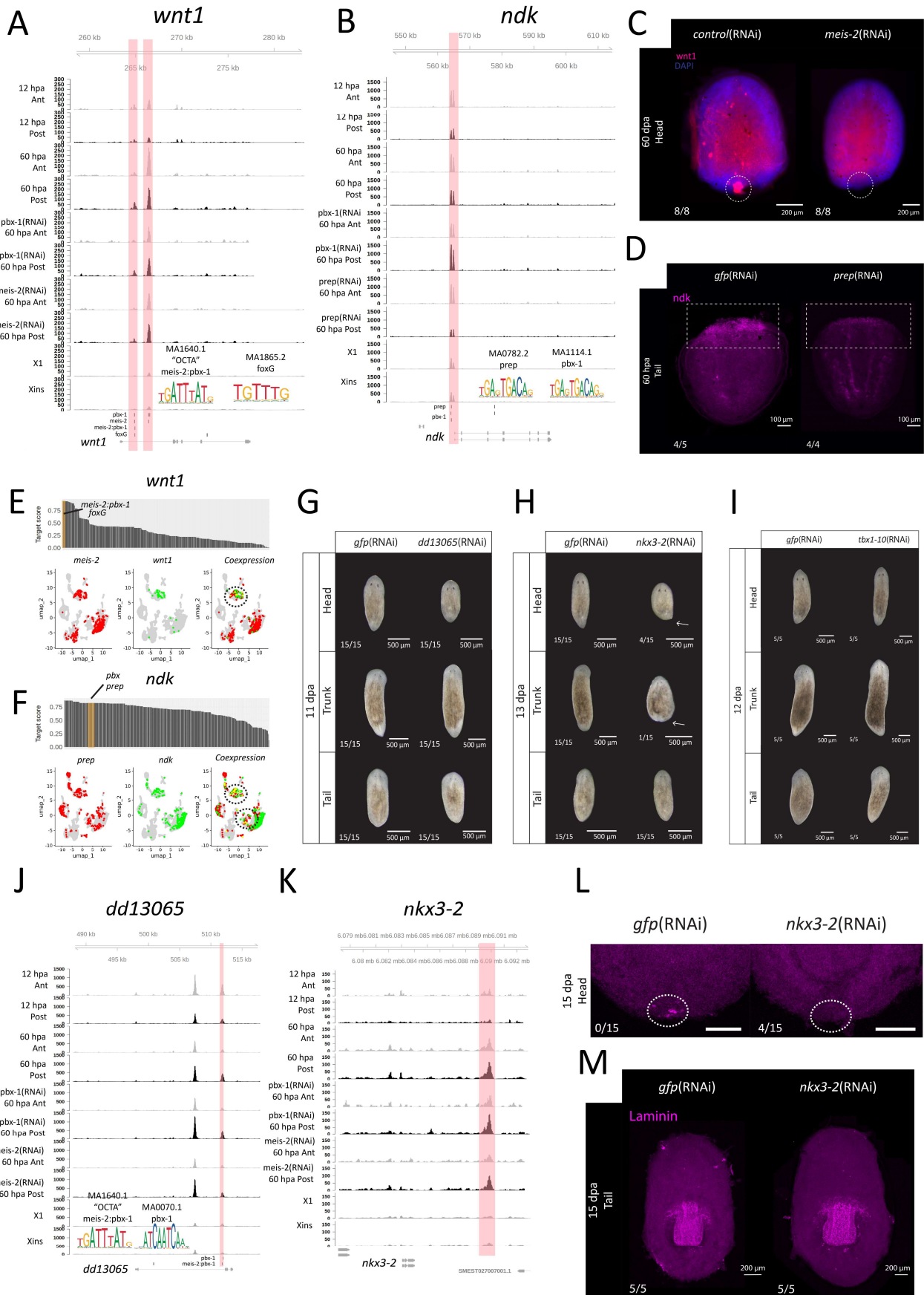


Figure 4.5 Functional characterisation of TALE targets. (A) The genomic region of *wnt1*. The first, second, ninth, and tenth tracks are from publicly available datasets (Li et al., 2021; Pascual-Carreras et al., 2023). From top to bottom, the tracks depict ATAC-seq coverage (RPKM) in early-stage anterior blastemas (12 hpa), early-stage posterior blastemas, late-stage anterior blastemas (60 hpa), late-stage posterior blastemas (60 hpa), late-stage anterior blastemas after knockdown of *pbx-1* (60 hpa), late-stage posterior blastemas after knockdown of *pbx-1* (60 hpa), late-stage anterior blastemas after knockdown of *meis-2* (60 hpa), late-stage posterior blastemas after knockdown of *meis-2* (60 hpa), X1 cells, and Xins cells (insensitive to irradiation, differentiated cells). Two highly accessible regions in the first intron of *wnt1* are marked with orange. Bound motifs are displayed under ATAC-seq tracks and above the gene model of *wnt1*. The PWMs of the Meis and FoxG motifs are displayed above the transcript. **(B)** The genomic region of *ndk*. The six first tracks along with the ninth and tenth tracks are the same as in (A), but the seventh track depicts ATAC-seq coverage in late-stage anterior blastemas after knockdown of *prep*, and the eighth track depicts ATAC-seq coverage in late-stage posterior blastemas after knockdown of *prep*. The footprints of *prep* and *pbx* are marked under the tracks. **(C)** Fluorescence in situ hybridisation (FISH) of the posterior pole showing expression of *wnt1* in late-stage posterior blastemas (60 hpa) of head fragments in control *gfp(RNAi)* and *meis-2(RNAi)* worms. Counts for the phenotype represented in the image are indicated in the lower left corner. **(D)** FISH of *ndk* in late-stage anterior blastemas (60 hpa) of tail fragments in control *gfp(RNAi)* and *prep(RNAi)* worms. **(E)** The upper figure displays the ranking of footprints linked to the *wnt1* gene. The rank of the Meis and foxG motifs are indicated with orange. The lower UMAP plots display the expression and coexpression of *meis-2* and *wnt1* in 72 hpa fragments with both anterior and posterior blastemas (Cui et al., 2023a). **(F)** The upper figure displays the ranking for footprints linked to the *ndk* gene, and the rank of *pbx* and *prep* are marked with orange. The lower UMAP plot displays expression and coexpression of *prep* and *ndk* in 72 hpa worms. **(G)** Knockdown phenotype of *dd13065(RNAi)* worms, no visible abnormalities are observed. **(H)** Knockdown phenotype of *nkx3-2(RNAi)* worms with potential mild defect. Counts for the pole-lacking phenotype are indicated to the lower left of the worms (e.g. 4 out of 15 *nkx3-2(RNAi)* head fragments displayed the posterior defect in the image). **(I)** Knockdown phenotype of *tbx1-10(RNAi)* worms. No visible abnormalities are observed under brightfield. **(J)** The genomic region of *dd13065*. The ATAC-seq tracks are the same as in (A). Footprints of *meis* are shown under the tracks and putative intronic cis-regulatory regions are marked in orange. **(K)** The genomic region of *nkx3-2*. A putative cis-regulatory region is marked downstream of the gene. **(L)** FISH of *wnt1* in posterior pole cells in *control(RNAi)* and *nkx3-2(RNAi)* worms. **(M)** FISH of the pharynx (*laminin+*) in control and *nkx3-2(RNAi)* worms.

receptor. Although we found the “OCTA” Motif in a putative intronic cis-regulatory region (Figure 3.5J), we did not observe a visible phenotype after knockdown of *dd13065* (Figure 3.5G).

Furthermore, we sought to understand whether additional uncharacterised transcription factors are involved in posterior regeneration, and selected as candidates *nkx3-2* (downregulated in *meis-2(RNAi)* posterior blastemas) and *tbx-1/10* (upregulated in *pbx-1(RNAi)* blastemas). Other members of the Nkx family have been implicated in planarian regeneration and cell specification, while the Tbx family members remain to be functionally characterised in general (Scimone et al., 2017; Tewari et al., 2019). Preliminary knockdown results (3 rounds of RNAi, 4 worms out of 15 in a single round display defect) suggest that *nkx3-2* could also be required for posterior regeneration and formation of the posterior pole (Figure 3.5H, L), but likely not as a direct target of *meis-2*, as no bound footprint was linked to the *nkx3-2* locus (Figure 3.5K). Unlike *meis-2(RNAi)* worms, *nkx3-2(RNAi)* worms regenerate the pharynx, indicating that *nkx3-2* is not involved in this regulatory cascade. Lastly, knockdown of *tbx-1/10* did not result in a visible phenotype under brightfield conditions (Figure 4.5I).

4.2.6 Footprinting analysis infers direct regulatory interactions in the anterior blastema

In our previous work, we were able to distinguish enhancer-like elements associated with patterning genes, even though our analysis was based on neoblasts rather than differentiated cells where patterning genes are mainly expressed (Chapter 2, Neiro et al., 2022). Therefore, we took advantage of our blastemal ATAC-seq data and footprint ranking method to profile chromatin accessibility in the vicinity of crucial patterning genes, identify putative cis-regulatory regions and footprints, interpret the biology of bound motifs, and integrate transcriptomic and functional data to infer TALE-dependent and TALE-independent regulatory interactions.

In our analysis of direct *pbx-1* targets, the anterior pole marker *zic-1* was among the candidates (Figure 4.4 G1). We found a Pbx footprint localised to the promoter of *zic-1* (Figure 4.4A). Previously, the expression of *pbx-1* has been shown to be unaffected by the knockdown of *zic-1*, while *pbx-1(RNAi)* reduces the expression of *zic-1* in 24 hpa blastemas (Vogg et al., 2014; Vásquez-Doorman et al., 2014). Although coexpression of *zic-1* and *pbx-1* has not been demonstrated in the anterior pole, *pbx-1* has been shown to be coexpressed with the other pole marker *sFRP-1* (Chen et al., 2013), and *zic-1* is expressed in the anterior pole with other pole markers (Vogg et al., 2014; Vásquez-Doorman et al., 2014). Additionally, although anterior pole cells are present in low numbers, we found coexpressing cells in the scRNA-seq dataset from 72 hpa trunk fragments (Figure 4.6A), and *pbx-1* was highly expressed in the specific transcriptomic dataset collected from anterior poles (Figure S4.2) (Li et al., 2019). In our ranking of footprints linked to *zic-1*, we observed that Pbx motifs were in the upper end of ranked motifs (Figure 4.6A). In contrast, we found that Zic motifs were among the highest ranked footprints, possibly suggesting that *zic-1* expression in the anterior pole is mainly maintained by an autoregulatory loop of *zic-1* binding to a putative cis-regulatory region in its intron (Figure 4.6A).

Although the other anterior pole marker *foxD* was not downregulated in *prep(RNAi)* worms, we found that a Prep motif was among the highest ranked footprints linked to the *foxD* locus, localised to an accessible region downstream of the *foxD* gene (Figure 4.6B). In support of this regulatory interaction, earlier FISH experiments have shown that *prep* and *foxD* are coexpressed in the same cells (Scimone et al., 2014a), and both *prep* and *foxD* are highly expressed in the transcriptomic dataset from anterior pole cells (Figure S4.2). Moreover, *prep(RNAi)* abolishes the *foxD*-expressing anterior pole in 72 hpa anterior blastemas (Scimone et al., 2014a). Our footprint ranking also indicate that a Smad2 motif is among the highest ranked footprints linked to *foxD*, localised to another accessible region downstream of the *foxD* locus (Figure 4.6B). Knockdown of *activin-2* bifurcates the primary axis and splits the anterior pole, raising the possibility that activin signalling transduced via *smad2/3* controls *foxD* expression and the proper assembly of the anterior pole. Co-expressing cells are present in low numbers in the 72 hpa scRNA-seq dataset (Figure 4.6B), but *smad2/3* is expressed at higher levels than *foxD* in the bulk RNA-seq dataset from the anterior pole, providing support for the regulatory interaction (Figure S4.2).

The pole marker *sFRP-1* was also predicted to be a direct target of *pbx-1*, and we localised the footprint to an accessible region upstream of the *sFRP-1* locus (Figure 4.6C). Earlier ISH experiments have demonstrated that *pbx-1* and *sFRP-1* are coexpressed in the anterior pole (Chen et al., 2013), and we also found coexpression in the scRNA-seq dataset and in the anterior pole sample (Figure 4.6C, S4.2). Moreover, we found that Zic motifs were among the highest ranked footprints linked to the *sFRP-1* locus and were localised to another accessible region upstream of *sFRP-1*. Therefore, epigenomic data suggest that *zic-1* could be a direct regulator of *sFRP-1* in the anterior pole and could contribute to maintaining its expression.

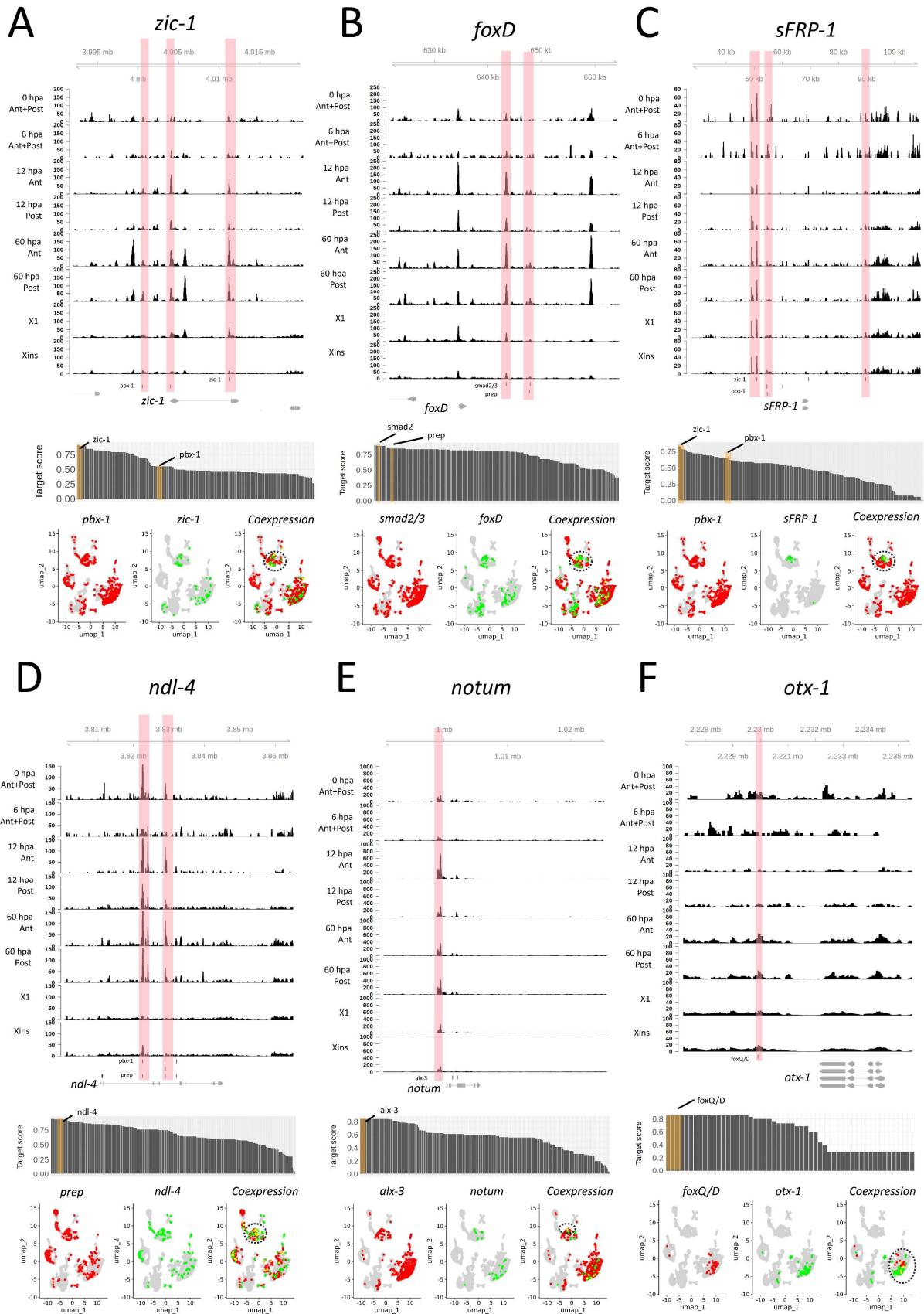


Figure 4.6 Prediction of direct regulators for anterior patterning genes. (A, B, C, D, E, F) The first, second, third, fourth, ninth, and tenth tracks are from publicly available datasets (Gehrke et al., 2019; Li et al., 2021; Pascual-Carreras et al., 2023). From top to bottom, the tracks depict ATAC-seq coverage (RPKM) in 0 hpa blastemas (mixture of anterior and posterior, Gehrke et al., 2019), 6 hpa blastemas (mixture of anterior and posterior, Gehrke et al., 2019), 12 hpa anterior blastemas (Pascual-Carreras et al., 2023), 12 hpa posterior blastemas (Pascual-Carreras et al., 2023), 60 hpa anterior blastemas, 60 hpa posterior blastemas, X1 cells (neoblasts, Li et al., 2021), and Xins cells (differentiated cells). **(A)** The genomic region of *zic-1*. Accessible regions with footprints of ZIC and PBX are coloured. The middle plot shows the footprint ranking of metazoan footprints linked to the *zic-1* locus, and the rank of ZIC and PBX motifs are indicated in orange. The lower UMAP plot displays expression of *pbx-1* (red), *zic-1* (green), and coexpressing cells (yellow) in a 72 hpa trunk fragment scRNA-seq atlas (Cui et al., 2023a). **(B)** The genomic region of *foxD*. Accessible regions with footprints of SMAD2 and PREP are marked. The middle plot displays the footprint ranking with the rank of SMAD2 and PREP indicated, and the lower plot shows expression of *smad2/3* (red), *foxD* (green), and coexpressing cells (yellow) in 72 hpa trunks. **(C)** The genomic region of *sFRP-1*. Accessible regions including ZIC and PBX motifs are marked, the middle plot displays the footprint ranking with the rank of ZIC and PBX marked, and the lower plot displays expression of *pbx-1* (red), *sFRP-1* (green), and their coexpression (yellow) in 72 hpa trunk fragments. **(D)** The genomic region of *ndl-4*. Accessible regions with footprints of PBX and PREP are labelled, the footprint ranking plot displays the rank of *pbx-1* and *prep*, and the lower plot displays the expression of *prep* (red), *ndl-4* (green), and coexpression (yellow) in 72 hpa trunk fragments. **(E)** The genomic region of *notum*. The promoter region containing an *alx-3* footprint is labelled, the top position of ALX3 is shown in the footprint ranking plot, and the lower plot displays the expression of *alx-3* (red), *notum* (green), and coexpressing cells (yellow) in 72 hpa trunk fragments. **(F)** The genomic region of *otx-1*. An accessible region containing FoxQ and FoxD footprints is labelled, the rank of foxQ and D motifs is labelled in the footprint ranking plot, and lower plot displays the expression of *foxQ/D* (red) and *otx-1* (green) along with the coexpression (yellow) in 72 hpa trunk fragments.

Besides *ndk*, the other FGFR family member *ndl-4* was also predicted to be a direct target of both *pbx-1* and *prep* (Figure 4.4G1, G2). Furthermore, footprints of Prep and Pbx were among the highest ranked footprints linked to the *ndl-4* locus, including the DECA motif (MA0782.2) known to be bound by a Pbx:Prep dimer in mouse (Figure 4.6D) (Penkov et al., 2013). These footprints were localised to the second and fourth introns of the *ndl-4* gene (Figure 4.6D). *Ndl-4* is expressed anteriorly in intact worms, and *pbx-1(RNAi)* abolishes *ndl-4* expression in blastemas at 6 dpa (Chen et al., 2013; Scimone et al., 2016). In the trunk scRNA-seq dataset, we found numerous cells coexpressing *prep* and *ndl-4* (Figure 4.6D). Therefore, multiple pieces of indirect evidence suggest a regulatory interaction between *prep/pbx-1* and *ndl-4*.

Intriguingly, we found that an Alx3 motif was the highest-ranked footprint linked to the anterior pole marker *notum*, localised to the highly accessible promoter (Figure 4.6E). Recently, the transcription factor *alx-3* has been found to be required for *notum* expression from 6 hpa onwards, and *alx-3(RNAi)* worms do not regenerate the anterior pole at 72 hpa (Akheralie et al., 2023). Furthermore, FISH experiments have demonstrated that *alx-3* and *notum* are expressed in the anterior pole (Akheralie et al., 2023), and our analysis of scRNA-seq data and bulk RNA-seq provides support for coexpression (Figure 4.6E, S4.2). Therefore, multiple independent sources of evidence indicate that *alx-3* is a direct regulator of *notum*.

As eye regeneration is specific to the anterior blastema and is driven by eye progenitors in the blastema, we used our anterior accessibility data to infer regulatory interactions in photoreceptor specification. Both *otx-1* and *foxQ/D* are required for eye differentiation and *otx-1* is expressed in a subset of neoblasts (Lapan and

Reddien, 2012; Scimone et al., 2014b), and we found that foxQ and foxD footprints were the highest ranked footprints linked to the *otx-1* locus (Figure 4.6F). Coexpression in the neuronal cluster further supports that *foxQ/D* could act as a direct regulator of *otx-1* in specification of photoreceptor cells (Figure 4.6F).

4.2.7 Footprinting analysis infers elements of the posterior gene regulatory network

In a similar fashion, we capitalised on our blastema ATAC-seq data and footprint ranking to identify putative cis-regulatory regions and footprints linked to posterior patterning genes, inferring TALE-dependent and TALE-independent regulatory interactions in the regenerating tail. Transcriptomics along the planarian AP axis and the transcriptional response to knockdown of Wnt signalling components have revealed a posterior specification program of coregulated genes, including *wnt1* expressed in the posterior pole, the other Wnt ligands *wnt11-1* and *wnt11-2*, the receptor *frz4-1*, the Hox factors *post-2c*, *lox5a*, and *box4*, and the transcription factor *sp5* (Reuter et al., 2015; Scimone et al., 2016; Tewari et al., 2019; Pascual-Carreras et al., 2023). Moreover, a subset of these genes is misregulated following knockdowns in the acoel *Hofstenia miamia*, suggesting that some downstream effectors of the Wnt signalling pathway are conserved among metazoans.

Among the genes downregulated by reduced Wnt signalling is *sp5*, which is also downregulated in *H. miamia* after *beta-catenin-1(RNAi)* (Reuter et al., 2015; Stückemann et al., 2017; Tewari et al., 2019; Pascual-Carreras et al., 2023). Interestingly, *sp5* has been found to be a direct target of Wnt signalling in both vertebrates and hydra (Weidinger et al., 2005; Fujimura et al., 2007). In our ranking of metazoan footprints bound in the posterior blastema linked to *sp5*, we observe that tcf7/tcf7l1/tcf7l2 footprints localised to the promoter are among the highest ranked footprints (Figure 4.7A), suggesting that Wnt signalling directly regulates *sp5* in the planarian posterior blastema. Moreover, we found cells coexpressing *tcf4* (Brown et al., 2018) and *sp5* in 72 hpa trunk fragments (Figure 4.7A). Altogether, our epigenomic data from posterior blastemas further support *sp5* as a direct target of Wnt signalling in planarians.

The Wnt ligand *wnt11-2* is another predicted direct target of both *pbx-1* and *meis-2*. Knockdown of *wnt11-2* produces tailless worms, similar to those observed in *pbx-1(RNAi)* and *meis-2(RNAi)* following amputation, suggesting that *pbx-1*, *meis-2*, and *wnt11-2* act at the same level in the regulatory hierarchy of posterior specification (Adell et al., 2009; Gurley et al., 2010; Sureda-Gómez et al., 2015). In our footprint ranking, we found that the OCTA motif (MA1640.1) is among the highest-ranked footprints and localised to the promoter, providing further evidence that *meis-2* and *pbx-1* are direct regulators of *wnt11-2* (Figure 4.7B). Moreover, the scRNA-seq atlas from 72 hpa trunk fragments uncovers coexpressing cells (Figure 4.7B). In contrast, the other Wnt ligand, *wnt11-1*, was not among the predicted targets of TALE factors, while tcf7/tcf7l1/tcf7l2 footprints localised to the third intron was among the highest ranked footprints and *tcf4* displays coexpression with *wnt11-1* in 72 hpa trunk fragments (Figure 4.7C). Therefore, analysis of chromatin accessibility infers that *wnt11-2* is a direct target of *pbx-1* and *meis-2*, while *wnt11-1* is putatively activated by Wnt signalling.

Moreover, the Wnt receptor *frz4-1* is predicted to be a target of *pbx-1* and *meis-2*. *frz4-1* is expressed in the posterior pole of late-stage posterior blastemas (72 hpa), displays a tailless phenotype after knockdown, and is considered to be the receptor for *wnt11-2* (Pascual-Carreras et al., 2021). In our footprint ranking, we found that the OCTA motif is among the highest ranked motifs, supporting that a dimeric complex of *meis-2* and *pbx-1* could be regulating the expression of *frz4-1* by binding to the promoter (Figure 4.7D). The scRNA-seq atlas indicates that *meis-2* and *frz4-1* are coexpressed in muscle cells, providing support for this regulatory relationship (Figure 4.7D). Additionally, we observe that an Isl footprint localised to a downstream accessible region is among the highest-ranked footprints, suggesting that *isl-1*, expressed in the posterior pole, could also be a direct regulator of *frz4-1* (Figure 4.7D).

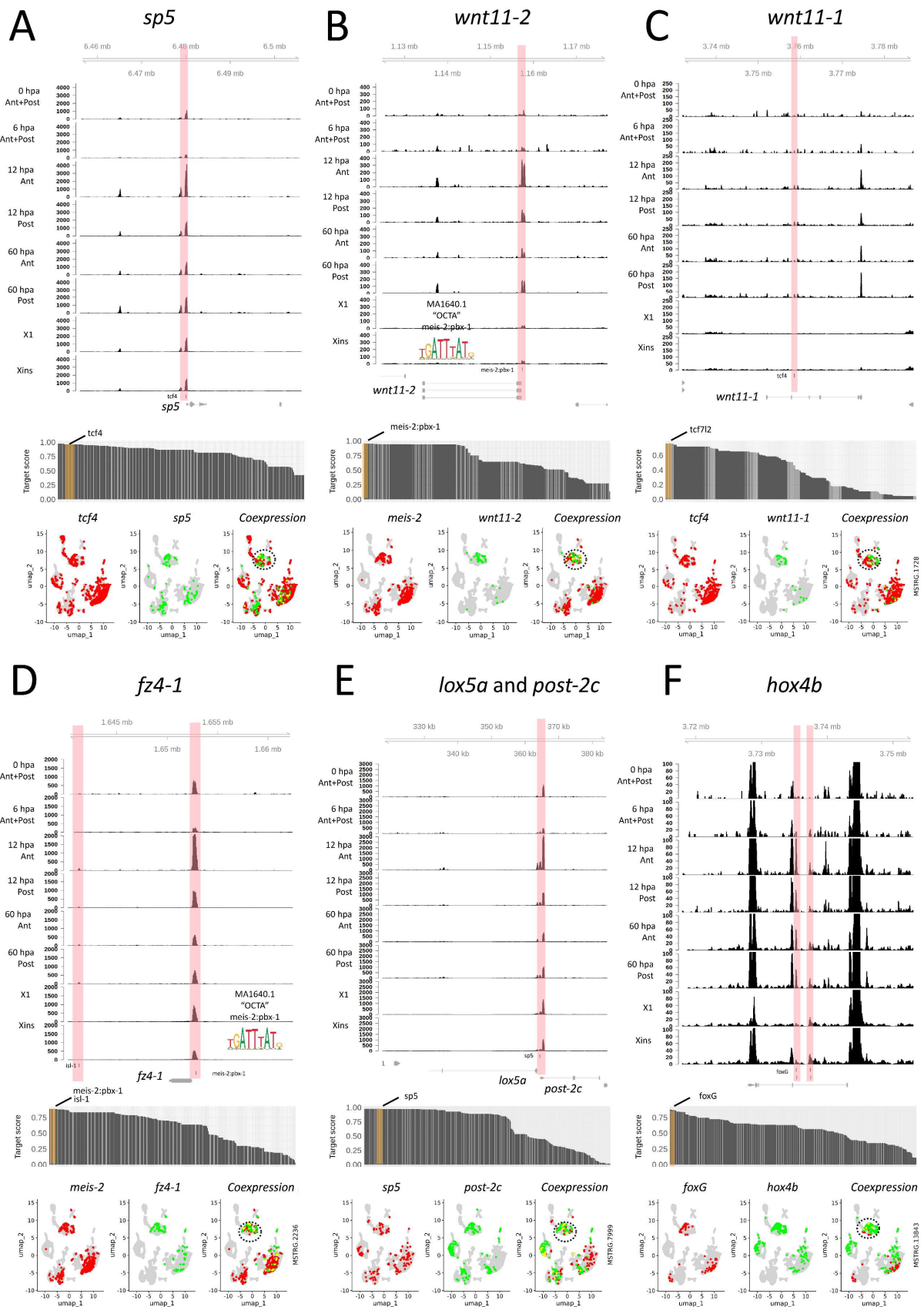


Figure 4.7 Prediction of direct regulators for posterior patterning genes. (A, B, C, D, E, F). The ATAC-seq tracks are the same as in Fig. 6. **(A)** The genomic region of *sp5*. The promoter with a TCF7L2 footprint is labelled. The middle plot shows the footprint ranking of metazoan footprints linked to the *sp5* locus, and the rank of TCFL72 footprints are indicated in orange. The lower UMAP plot displays expression of *tcf4* (red), *sp5* (green), and coexpressing cells (yellow) in a 72 hpa trunk fragment scRNA-seq atlas (Cui et al., 2023a). **(B)** The genomic region of *wnt11-2*. The promoter containing a Meis footprint is marked, the rank of Meis footprints is labelled in the footprint ranking, and the lower UMAP plots display expression of *meis-2*, *wnt11-2*, and coexpressing cells (yellow) in 72 hpa trunks. **(C)** The genomic region of *wnt11-1*. An accessible region containing a TCF7L2 motifs is labelled, the rank of the footprint is labelled in orange in the footprint ranking, and the lower plot displays expression of *tcf4* (red), *wnt11-1* (green), and their coexpression (yellow) in 72 hpa trunk fragments. **(D)** The genomic region of *fox4-1*. Accessible regions with footprints of PBX, MEIS, and ISL are labelled, the footprint ranking plot displays the rank of Pbx, Prep, and ISL motifs, and the lower plot displays the expression of *meis-2* (red), *fox4-1* (green), and coexpression (yellow) in 72 hpa trunk fragments. **(E)** The genomic region of *lox5a* and *post-2c*. The intervening region between the genes containing a SP5 footprint is labelled, the rank of the SP5 footprint is indicated in the footprint ranking, and the lower plot displays the expression of *sp5* (red), *post-2c* (green), and coexpressing cells (yellow) in 72 hpa trunk fragments. **(F)** The genomic region of *box4b*. Accessible regions with foxG footprints are labelled, the rank of the foxG motif is labelled in the footprint ranking plot, and lower plot displays the expression of *foxG* (red) and *box4b* (green) along with the coexpression (yellow) in 72 hpa trunk fragments.

In planarians, *Hox* genes are not clustered as in many metazoans (Durstion et al., 2011; Currie et al., 2016). Nonetheless, the *Hox* genes *lox5a* and *post-2c* in *S. mediterranea* are adjacent to each other and form a two-gene cluster (Figure 4.7E). Interestingly, although four *Hox* genes (*box4b*, *lox5a*, *post-2c*, *post-2d*) display posteriorly biased expression in *S. mediterranea* and are misregulated following knockdown of Wnt signalling, knockdown of *Hox* genes does not produce major regenerative patterning defects in the posterior (Currie et al., 2016; Scimone et al., 2016; Tewari et al., 2019; Pascual-Carreras et al., 2023). Nonetheless, *lox5a* displays indented tail blastemas at a low frequency (16 %) (Scimone et al., 2016). In our previous work, we found a *sp5* footprint localised to intermediate region between *lox5a* and *post-2c*, and our footprint ranking on blastemal data further indicate that this footprint is among the highest ranked footprints (Figure 4.7E). Moreover, scRNA-seq from 72 hpa trunk fragments uncovers cells coexpressing *post-2c* and *sp5* (Figure 4.7E). Altogether, blastemal accessibility data further supports *sp5* as a direct regulator of either *lox5a* or *post-2c*.

Lastly, we queried the posteriorly expressed *box4b* gene and found foxG among the highest ranked footprints, being localised to accessible regions in the last intron (Figure 4.7F). As *foxG* is expressed in a scattered manner in the posterior blastema from 24 hpa onwards (Pascual-Carreras et al., 2023) and *box4b* expression starts at 24 hpa (Iglesias et al., 2008; Tewari et al., 2019), *box4b* could another direct target of *foxG* in the posterior blastema, together with *wnt1*. Additionally, cells coexpressing *foxG* and *box4b* were present in 72 hpa trunk fragments, supporting this regulatory interaction (Figure 4.7F).

4.3 Discussion

In this study, we aimed to understand the regulatory logic of TALE factors in planarian axial regeneration and patterning. Previously, *pbx-1* and *prep* had been shown to be necessary for the formation of the anterior pole, while *pbx-1* is also required for pole specification in the posterior blastema. We expanded the repertoire of TALE factors involved in axial regeneration in *S. mediterranea*, identifying the Meis family member *meis-2* as essential for posterior regeneration. To infer the genes regulated by TALE factors, we analysed the transcriptomic and epigenomic response following knockdown of these factors. In the anterior blastema, we discovered that anterior pole markers and FGFR family members are putative targets of *pbx-1* and *prep*, while the pole marker *wnt1* and other genes of the posterior specification program were inferred to be direct targets of *pbx-1* and *meis-2* in the posterior blastema. Furthermore, we mined the chromatin accessibility data and integrated it with existing planarian literature to infer TALE-independent regulatory interactions in axial regeneration.

In our analysis, we rely on bulk RNA-seq and ATAC-seq data, and perform preliminary functional experiments to validate the *meis-2-wnt1* and *prep-ndk* regulatory interactions. Continued functional characterisations with larger replicate numbers, along with FISH experiments demonstrating coexpression of *meis-2-wnt1* and *prep-ndk*, would further corroborate these findings. Nonetheless, some inferred regulatory interactions, such as *pbx-1-sFRP-1*, are already supported by knockdown experiments and coexpression patterns (Chen et al., 2013). Since transgenic constructs and reporters are not yet available in the planarian model system, the field has to rely on an arsenal of indirect methods to discover cis-regulatory regions. So far, approaches for identifying cis-regulatory elements have utilised chromatin accessibility and histone modifications indicative of enhancers (Gehrke et al., 2019; Li et al., 2021; Neiro et al., 2022; Pascual-Carreras et al., 2023; Wiggins et al., 2023; Ivanković et al., 2024). Furthermore, a spatiotemporal expression atlas over the regeneration time course has been used to infer regulatory modules, while a recent effort has begun the integration of scRNA-seq and scATAC-seq in the intact adult worm to infer gene regulatory networks regulating the differentiation of cell lineages (Cui et al., 2023b; Pérez-Posada et al., 2024). For regenerative patterning, a time-course of multiome data from anterior and posterior blastemas would inform on the axial regulatory mechanisms further, along with techniques such as targeted ChIP-seq against transcription factors and the transcriptional cofactor p300, and chromosome conformation capture for enhancer-promoter interactions (Johnson et al., 2007; Visel et al., 2009; Schoenfelder et al. 2015; Preissl et al., 2023).

Interestingly, our analysis suggests that although the transcriptomic response to *pbx-1(RNAi)* and *meis-2(RNAi)* differ in overall gene composition, a key set of patterning genes are downregulated by both factors. This suggests that the evolutionarily conserved Pbx-1:Meis-2 dimer could be regulating the posterior pole marker *wnt1*, *wnt11-2*, and its putative receptor *fzd4-1*. Intriguingly, TALE factors have been associated with Wnt signalling in axial regeneration in vertebrates, raising the possibility of a conserved regulatory module (Kakebeen et al., 2020; Mariani et al., 2021). Similarly, *pbx-1* and *prep* share as targets the FGFR family members *ndk* and *ndl-4* in the anterior, suggesting that a Pbx-1:Prep dimer could be a direct regulator of these genes. Moreover, at least *pbx-1* appears to be a direct regulator of anterior pole markers.

Taken together, this study has inferred both TALE-dependent and TALE-independent regulatory interactions in the regenerating planarian, providing a framework and testable hypotheses for studying planarian axial regeneration. Understanding the gene regulatory networks of the versatile positional information system that enables whole-body regeneration in planarians provides broader insights into the fundamental principles of regenerative abilities.

4.4 Methods

4.4.1 Planarian culture, cloning, and RNA interference

Planarians were cultured as in **3.4.4 Planarian culture**. Worms of the asexual strain were used for collecting RNA-seq and ATAC-seq data following knockdown of *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)*. Asexual worms were also used for staining with *wnt1*, *sFRP-1*, and Synapsin in adult *meis-2(RNAi)* worms and *wnt1* in 60 hpa *meis-2(RNAi)* fragments. Sexual worms were used for staining *porcupine*, *PC2*, and *laminin* in *meis-2(RNAi)* worms, as well as *ndk* in *prep(RNAi)* 60 hpa fragments. Sexual worms were used for knocking down *dd13065*, *nkx3-2*, and *tbx-1/10*.

Genes were cloned and dsRNA synthesised as outlined in **3.4.5 Cloning**, dsRNA was prepared as in **3.4.5.4 Synthesis of dsRNA and riboprobes**, and RNA interference was performed as in **3.4.6 RNA interference**.

4.4.2 RNA sequencing

The RNAi injection schedule was setup involving controls, *pbx-1(RNAi)*, *prep(RNAi)*, and *meis-2(RNAi)* worms. The worms were amputated into head, trunk, and tail fragments at day 11. Around ten blastemas from head or tail fragments (posterior or anterior blastemas) were dissected for each RNA-seq sample. After dissection, the blastemas were transferred to an Eppendorf tube containing planarian water. The tube was kept on ice until the required number of blastemas was reached. Planarian water was removed, Trizol added, and total RNA isolated as outlined in 3.4.5.1 Total RNA isolation, with the addition that 1 µl glycogen was added together with isopropanol to facilitate pelleting. The trunk fragments were cultivated for 11 to 20 days, and the phenotype was assessed under a brightfield microscope. RNA samples were sequenced using poly A enrichment (Novogene, Illumina Sequencing PE 150, mRNA library preparation (poly A enrichment)). Sequenced samples are listed in Table S4.2 and Table S4.3.

4.4.3 ATAC-seq

4.4.3.1 Tagmentation reaction

The ATAC-seq protocol was optimised based on standard protocols (Buenrostro et al., 2013; Corces et al., 2017). For each sample, 20 blastemas were dissected under the microscope and transferred to an Eppendorf tube containing planarian water. The tube was kept on ice until all blastemas had been dissected. The tube was briefly centrifuged to collect all blastemas at the bottom of the tube. Planarian water was removed, and 100 µl resuspension detergent buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40, 0.1% Tween-20, 0.01% Digitonin) was added to the mixture. The tube was incubated on ice for 2 min. The blastemas were mechanically disassociated using a microcentrifuge pestle until almost a homogenous solution was obtained (less than 3 min disassociation). The tube was incubated on ice for an additional 2 min. Then 1000 µl resuspension wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20) was added to the tube. The resuspension was mixed by pipetting up and down. The suspension was filtered through a cell strainer. The filtered solution was transferred to an Eppendorf tube, and the suspension was centrifuged (500g, 10 min, 4°C). The supernatant was discarded and the pellet was resuspended in 50 µl of tagmentation solution (25 µl 2x TD buffer, 2.5 µl Transposase (100 mM), 16.5 µl PBS (Phosphate-buffered saline), 0.5 µl 1 % digitonin, 0.5 µl 10 % Tween-20, 5 µl nuclease-free water). The tagmentation solution was incubated in a Thermomixer (37°C, 30 min, 800 rpm). The tagmentation solution was purified by using a Qiagen MiniElute Reaction Cleanup kit. The transposed DNA was eluted into 20 µl nuclease-free water and stored at -20°C.

4.4.3.2 Amplification

The PCR reaction (20 µl of purified transposed DNA, 2.5 µl of universal Ad1 index primer 25 µM, 2.5 µl specific Ad2 index primer, and 25 µl NEBNext High-Fidelity 2x PCR Master) was cycled for an initial PCR

reaction according to the parameters in Table 4.1. The index primers 24 Ad2.1 to Ad2.24 as provided by Buenrostro et al. (2013) were used for amplification.

Table 4.1 Initial PCR program for amplifying ATAC-seq samples.

Step	Temperature	Time	Cycles
Initial elongation	72°C	5 min	1
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	5
Annealing	63°C	30 sec	5
Elongation	72°C	1 min	5
Hold	4°C	∞	1

Following initial PCR, the sample tubes were incubated on ice during the qPCR reaction step. A qPCR reaction (10 µl KAPA SYBR FAST qPCR Master Mix 2x, 1 µl Ad1 primer 25 µM, 1 µl Ad2 index primer 25 µM, 4 µl nuclease-free water, 4 µl of DNA from the initial PCR reaction) were prepared on a qPCR plate according to the program in Table 4.2. From the qPCR plot, the number of cycles required to reach the third of maximum intensity was established, and the original PCR reaction was amplified for this number of additional cycles (i.e. 5 initial cycles followed by these additional cycles).

Table 4.2 qPCR program for determining optimal number of additional PCR cycles.

Step	Temperature	Time	Cycles
Initial denaturation	98°C	1 min	1
Denaturation	98°C	10 sec	25
Annealing	63°C	30 sec	25
Elongation	72°C	1 min	25
Hold	4°C	∞	1

4.4.3.3 Size selection

Double-sided size selection with AMPure beads was used to size select the samples for sequencing. The sample was transferred to a microcentrifuge tube, AMPure XP beads (25 µl, corresponding to 0.5x of the original volume 50 µl) were added to the mixture, and the reaction was mixed by pipetting up and down multiple times. The reaction was incubated at room temperature (10 min), and the beads were collected at the bottom (5 min) using a magnetic rack. The supernatant was transferred to a new tube after the beads had settled and the solution had become clear. AMPure beads (65 µl, corresponding to 1.3x of the original volume 50 µl) was added, and the solution was mixed by pipetting. The reaction was incubated (room temperature, 10 min) and beads were collected (5-10 min) on the magnetic rack. The supernatant was discarded when all beads had accumulated at the bottom of the tube. The solution was purified with freshly prepared 80 % ethanol (200 µl) by carefully pipetting to the sides of the microcentrifuge tube and not directly on the beads. Ethanol was carefully discarded without perturbing the beads, and the beads were air-dried (37°C) until the beads were dry. The beads were resuspended in nuclease-free water (20 µl), the solution was mixed well, and the solution was incubated (10 min). The beads were collected using the magnetic rack (5 min), and the supernatant was transferred to a new microcentrifuge tube. The DNA concentration was measured with Qubit (1 µl of the sample, 1 µl dye and 198 µl of buffer).

4.4.3.4 Quality control and sequencing

The fragment size distribution was assessed with the TapeStation system. The samples were sequenced with Novaseq 150 PE (Novogene) to a depth of 50M reads. Sequenced samples are listed in and Table S4.4.

4.4.4 RNA seq analysis

The reads were trimmed and quality filtered using fastp (Chen et al., 2018) and mapped to the *S. mediterranea* annotation (Neiro et al., 2022) using Kallisto (Bray et al., 2016). Differential gene expression analysis was performed using DEseq2 (Love et al., 2014). The pipeline of Cui et al. (2023a) was used for scRNA-seq analysis.

4.4.5 ATAC-seq analysis

4.4.5.1 Quality assessment

The reads were trimmed and quality filtered using fastp (Chen et al., 2018) and aligned using Bowtie2 (Langmead and Salzberg, 2012). Samtools was used to retain unique reads with a quality score higher than 10 (Li et al., 2009). Insert size distributions were generated using Picard (<https://github.com/broadinstitute/picard>). MACS3 (--nomodel --extsize 200 --shift -100) was used to call peaks (<https://macs3-project.github.io/MACS>) (Zhang et al., 2008). The FriP score was calculated using bedtools (Quinlan et al., 2014) as

$$FRiP = \frac{n_{peak}}{n_{tot}}$$

where n_{peak} is the number of reads in the sample overlapping with called peaks and n_{tot} the total number of reads in sample. Deeptools2 bamCoverage was used to generate coverage tracks in bigwig format (reads million mapped reads, RPKM) (Ramírez et al., 2016), and Gviz was used to visualise genomic tracks (Hahne et al., 2016).

4.4.5.2 Footprint ranking of TALE targets

TALE motifs were collected from the Jaspas database and are listed in Table S4.5 (Rauluseviciute et al., 2024). ATAC-seq called in anterior and posterior blastemas were used for footprinting TOBIAS (commands ATACCorrect, ScoreBigwig, and BINDetect) was applied to calculate the absolute footprint score b_j and the log2 fold change b_{log2f_j} in footprint scores between control and TALE(RNAi) blastemas for each motif $j \in \{1, \dots, m\}$ (Bentsen et al., 2020). ChIPSeeker was used to assign genes to the nearest gene (Yu et al., 2015).

For each differentially regulated target gene $i \in \{1, \dots, n\}$, the adjusted p-value p_i and absolute value of log2 fold change $|p_{log2fc_i}|$ between control and TALE(RNAi) blastemas were obtained from differential gene expression analysis.

An inverse percentile rank was calculated for p_i :

$$R(p_i) = \text{rank}(p_i)$$

$$P(p_i) = 1 - \frac{R(p_i) - 1}{n - 1}$$

where the statistical rank $\text{rank}(p_i)$ is the position of p_i in the ordered list of adjusted p values. Thus, the lowest p value is assigned a rank of 1, while the highest p value equals 0.

Similarly, a percentile rank was calculated for the absolute value of log2 fold change $|p_{log2fc_i}|$:

$$R(|p_{log2fc_i}|) = \text{rank}(|p_{log2fc_i}|)$$

$$P(|p_{log2fc_i}|) = \frac{R(|p_{log2fc_i}|) - 1}{n - 1}$$

where the gene with the highest absolute log2 fold change in expression after *TALE(RNAi)* is assigned the value 1, and the gene with the lowest absolute log2 fold change equals 0.

For footprint scores b_j , a similar percentile ranking was performed:

$$R(b_j) = \text{rank}(b_j)$$

$$P(b_j) = \frac{R(b_j) - 1}{m - 1}$$

as well as for the log2 fold change in footprint score $b_{\log_2 f c_j}$:

$$R(b_{\log_2 f c_j}) = \text{rank}(b_{\log_2 f c_j})$$

$$P(b_{\log_2 f c_j}) = \frac{R(b_{\log_2 f c_j}) - 1}{m - 1}$$

The target score for a regulatory interaction between motif j and target gene i , $T_{i,j}$, was calculated as the average of these four percentile ranks:

$$T_{i,j} = \frac{P(p_i) + P(\log_2 f c_i) + P(b_j) + P(\log_2 f c_j)}{4}$$

4.4.5.3 Gene-specific footprint ranking

In the general footprint ranking of metazoan motifs, we downloaded all available motifs from the Jaspar database and calculated the absolute footprint score b_j in either anterior or posterior blastemas, as well as the log2 fold change $b_{\log_2 f c_j}$ in footprint scores between anterior and posterior wild-type blastemas for each motif $j \in \{1, \dots, m\}$ linked to a given gene (Bentsen et al., 2020).

Footprint scores were ranked:

$$R(b_j) = \text{rank}(b_j)$$

$$P(b_j) = \frac{R(b_j) - 1}{m - 1}$$

so that the footprint with the highest footprint score in the target blastema (anterior or posterior) receives the value 1. Similarly, the log2 fold changes in footprint score $b_{\log_2 f c_j}$ were ranked:

$$R(\log_2 f c_j) = \text{rank}(\log_2 f c_j)$$

$$P(\log_2 f c_j) = \frac{R(\log_2 f c_j) - 1}{m - 1}$$

so that the footprint with the footprint score most biased towards the target tissue (anterior or posterior) is assigned a value of 1. Lastly, the target score for motif j and its corresponding transcription factor, indicating its potential as a direct regulator of a given gene, was calculated as the average of these two percentile ranks:

$$U_j = \frac{P(b_j) + P(\log_2 f c_j)}{2}$$

4.4.6 In situ hybridisation and immunostaining

Riboprobes were prepared as in **3.4.5.4 Synthesis of dsRNA and riboprobes**. Fluorescence in situ hybridisation was performed according to the protocol devised by King et al. (2013), and immunostainings according to Forsthoefel et al. (2018).

4.5 Supplementary material

Table S4.1 Forward and reverse primer sequences used to clone genes.

Primer	Sequence
prep Forward	ACCGGTGCCGGTTCGGTTTA
prep Reverse	ACGGCAGCAGCCAAACTAGCC
pbx-1 Forward	TGAAGCCTGCGTTGTACAGTGTGT
pbx-1 Reverse	GCTGCTGTGGCAGCGGCATA
meis-2 Forward	ACGTCCTCGCCTTTGTCGGGA
meis-2 Reverse	TCCCTCCGCCACCACCAGAC
dd13065 Forward	CCGACAAACGAGCGGTTCTA
dd13065 Reverse	TTCATTTGCGCTCCAAATGGC
nkx3-2 Forward	AGATGATGCGAAAGAGCACA
nkx3-2 Reverse	ACAAGAGACGAATCGAAGCCT
tbx-1/10 Forward	TCTGGAAGGCGAATGTTTCCA
tbx-1/10 Reverse	TCACATTTGCAAGGGGCTGT

Table S4.2 RNA-seq libraries collected from 18 blastemas.

Sample	Filtered reads	Mapping rate
gfp(RNAi) Anterior 1	19.03M	78.1%
gfp(RNAi) Anterior 2	16.34M	77.3%
gfp(RNAi) Anterior 3	22.60M	71.5%
gfp(RNAi) Anterior 4	21.69M	71.3%
gfp(RNAi) Posterior 1	21.21M	74.8%
gfp(RNAi) Posterior 2	21.96M	79.4%
gfp(RNAi) Posterior 3	19.56M	72.6%
gfp(RNAi) Posterior 4	22.93M	71.1%
pbx(RNAi) Anterior 1	21.92M	79.1%
pbx(RNAi) Anterior 2	21.77M	78.6%
pbx(RNAi) Posterior 1	18.91M	76.5%
pbx(RNAi) Posterior 2	20.93M	80.3%
prep(RNAi) Anterior 1	10.70M	77.3%
prep(RNAi) Anterior 2	13.89M	76.7%
prep(RNAi) Posterior 1	25.31M	73.7%
prep(RNAi) Posterior 2	24.29M	74.8%
meis-2(RNAi) Anterior 1	32.60M	70.6%
meis-2(RNAi) Anterior 2	19.50M	72.2%
meis-2(RNAi) Posterior 1	30.52M	77.2%
meis-2(RNAi) Posterior 2	22.31M	76.1%

Table S4.3 RNA-seq libraries collected from 60 hpa blastemas.

Sample	Filtered reads	Mapping rate
gfp(RNAi) Anterior 1	24.27M	68.4%
gfp(RNAi) Anterior 2	21.01M	69.3%
gfp(RNAi) Anterior 3	21.31M	74.8%
gfp(RNAi) Posterior 1	20.87M	71.4%
gfp(RNAi) Posterior 2	23.85M	73.4%
pbx(RNAi) Anterior 1	21.8M	75.3%
pbx(RNAi) Anterior 2	29.18M	73.6%
pbx(RNAi) Anterior 3	22.91M	71.2%
pbx(RNAi) Posterior 1	20.2M	76.7%
pbx(RNAi) Posterior 2	21.4M	73.6%
prep(RNAi) Anterior 1	21.85M	72.5%
prep(RNAi) Anterior 2	20.41M	70.1%
prep(RNAi) Anterior 3	16.94M	63.3%
prep(RNAi) Posterior 1	20.55M	73.3%
prep(RNAi) Posterior 2	18.08M	67.8%
prep(RNAi) Posterior 3	28.87M	68.2%
meis-2(RNAi) Anterior 1	21.0M	77.2%
meis-2(RNAi) Anterior 2	19.45M	78.5%
meis-2(RNAi) Posterior 1	20.50M	80.7%
meis-2(RNAi) Posterior 2	22.6M	79.7%

Table S4.4 ATAC-seq libraries collected from 60 hpa blastemas.

Sample	Filtered reads	Mapping rate
gfp(RNAi) Anterior 1	61.85M	83.63%
gfp(RNAi) Anterior 2	61.08M	38.64%
gfp(RNAi) Anterior 3	90.19M	63.33%
gfp(RNAi) Posterior 1	66.53M	86.93%
gfp(RNAi) Posterior 2	31.93M	72.24%
gfp(RNAi) Posterior 3	87.9M	72.38%
pbx(RNAi) Anterior 1	68.56M	59.18%
pbx(RNAi) Anterior 2	59.55M	63.04%
pbx(RNAi) Anterior 3	87.45M	70.48%
pbx(RNAi) Posterior 1	59.01M	70.61%
pbx(RNAi) Posterior 2	35.91M	61.36%
pbx(RNAi) Posterior 3	43.93M	85.24%
prep(RNAi) Anterior 1	64.87M	85.46%
prep(RNAi) Anterior 2	42.22M	40.66%
prep(RNAi) Anterior 3	58.11M	88.14%
prep(RNAi) Posterior 1	62.35M	82.02%
prep(RNAi) Posterior 2	60.26M	56.69%
prep(RNAi) Posterior 3	67.65M	72.33%
meis-2(RNAi) Anterior 1	42,84M	51.08%
meis-2(RNAi) Anterior 2	54.80M	39.69%
meis-2(RNAi) Anterior 3	24.9M	53.25%
meis-2(RNAi) Posterior 1	36.37M	62.63%
meis-2(RNAi) Posterior 2	43.55M	54.64%
meis-2(RNAi) Posterior 3	36.37M	52.52%

Table S4.5 TALE motifs used for footprinting and predicting targets.

Motif name	JASPAR ID
exd	MA0222.1
PBX1	MA0070.1
PBX2	MA1113.1
PBX2	MA1113.2
PBX3	MA1114.1
hth	MA0227.1
Meis1	MA0498.1
MEIS1	MA0498.2
MEIS1	MA1639.1
MEIS2	MA1640.1
MEIS2	MA0774.1
MEIS3	MA0775.1
PKNOX1	MA0782.1
PKNOX1	MA0782.2
PKNOX2	MA0783.1

Table S4.6 List of motifs and inferred regulatory interactions.

Regulator	Motif name	JASPAR ID	Target gene
pbx-1	PBX2	MA1113.1	wnt1
pbx-1	PBX2	MA1113.2	wnt1
pbx-1	PBX3	MA1114.1	wnt1
meis-2	hth	MA0227.1	wnt1
meis-2	MEIS3	MA0775.1	wnt1
meis-2	MEIS1	MA1639.1	wnt1
meis-2	MEIS2	MA1640.1	wnt1
foxG	FoxG	MA1865.2	wnt1
pbx-1	PBX2	MA1113.1	ndk
pbx-1	PBX2	MA1113.2	ndk
pbx-1	PBX3	MA1114.1	ndk
prep	PKNOX1	MA0782.2	ndk
pbx-1	PBX1	MA0070.1	dd13065
pbx-1	PBX2	MA1113.2	dd13065
meis-2	MEIS1	MA1639.1	dd13065
meis-2	MEIS2	MA1640.1	dd13065
pbx-1	PBX1	MA0070.1	zic-1
zic-1/zic-2	ZIC1	MA0696.1	zic-1
zic-1/zic-2	ZIC4	MA0751.2	zic-1
zic-1/aic-2	ZIC5	MA1584.2	zic-1
prep	PKNOX1	MA0782.3	foxD
prep	PKNOX2	MA0783.1	foxD
smad2/3	SMAD2	MA1964.2	foxD
pbx-1	PBX1	MA0070.1	sFRP-1
pbx-1	PBX2	MA1113.1	sFRP-1
zic-1	ZIC1	MA0696.1	sFRP-1
zic-1	ZIC4	MA0751.2	sFRP-1
zic-1	ZIC5	MA1584.2	sFRP-1
pbx-1	PBX1	MA0070.1	ndl-4
pbx-1	exd	MA0222.1	ndl-4
pbx-1	PBX2	MA1113.1	ndl-4
pbx-1	PBX2	MA1113.2	ndl-4
pbx-1	PBX3	MA1114.1	nld-4
prep	PKNOX1	MA0782.2	nld-4
alx-3	ALX3	MA0634.2	notum
foxQ/D	FoxD-b	MA1862.2	otx-1
foxQ/D	FoxQ	MA1873.2	otx-1
tcf4	TCF7L2	MA0523.2	sp5
tcf4	TCF7	MA0769.3	sp5
tcf4	TCF7L1	MA1421.1	sp5
pbx-1	PBX2	MA1113.2	wnt11-2
meis-2	MEIS1	MA1639.1	wnt11-2
meis-2	MEIS2	MA1640.1	wnt11-2
tcf4	TCF7	MA0769.3	wnt11-1
tcf4	TCF7L1	MA1421.1	wnt11-1
tcf4	TCF7L2	MA0523.2	wnt11-1
pbx-1	PBX2	MA1113.1	fz4-1
pbx-1	PBX2	MA1113.2	fz4-1
meis-2	MEIS2	MA1640.1	fz4-1
sp5	SP8	MA0747.2	lox5a/post-2c
sp5	SP3	MA0746.3	lox5a/post-2c
foxG	FoxG	MA1865.2	hox4b

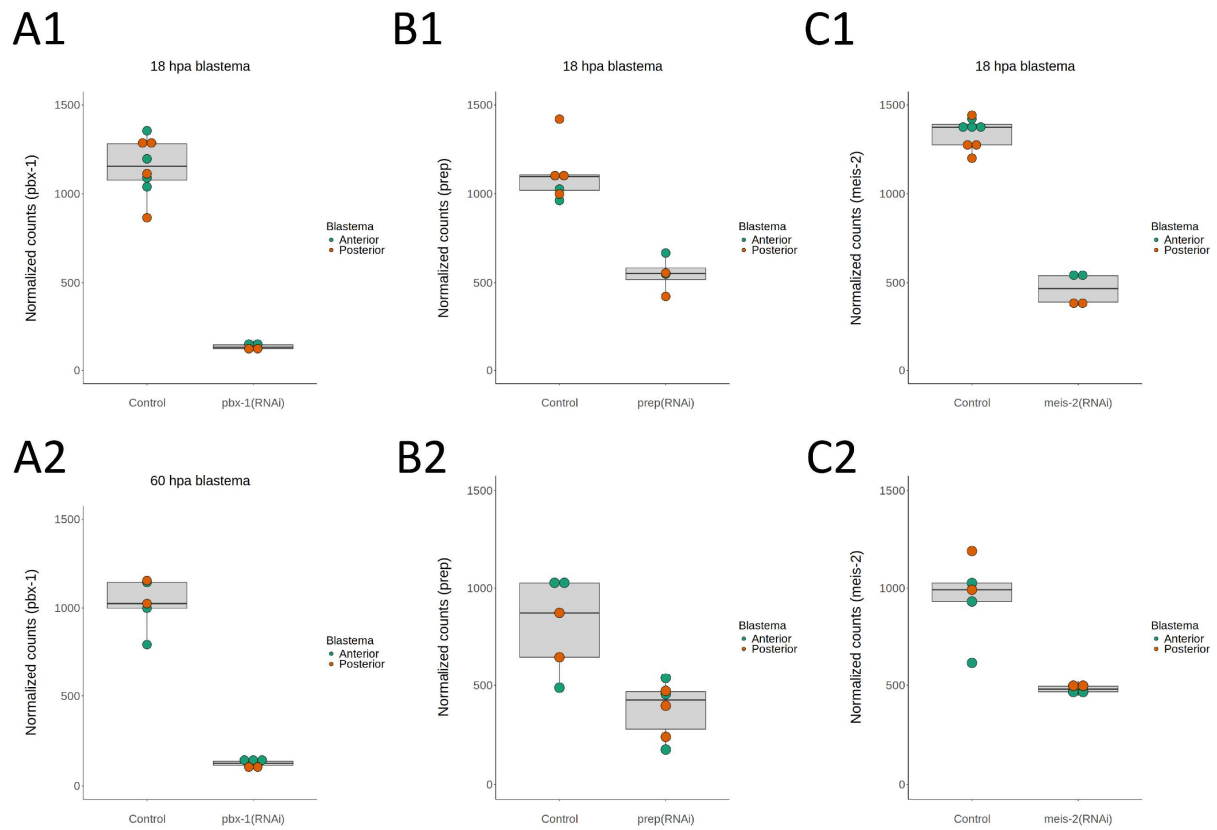


Figure S4.1 Knockdown efficiency of *pbx-1*, *prep*, and *meis-2*. (A1) Normalised counts (DEseq2) of *pbx-1* in early-stage blastemas (18 hpa) of control worms and after knockdown of *pbx-1*. (A2) Normalised counts for *pbx-1* in in late-stage blastemas (60 hpa) of control worms and after knockdown of *pbx-1*. (B1) Normalised counts of *prep* in early-stage blastemas (18 hpa) of control worms and after knockdown of *prep*. (B2) Normalised counts of *prep* in late-stage blastemas (60 hpa) of control worms and after knockdown of *prep*. (C1) Normalised counts of *meis-2* in early-stage blastemas (18 hpa) of control worms and after knockdown of *meis-2*. (C2) Normalised counts of *meis-2* in late-stage blastemas (60 hpa) of control worms and after knockdown of *meis-2*.

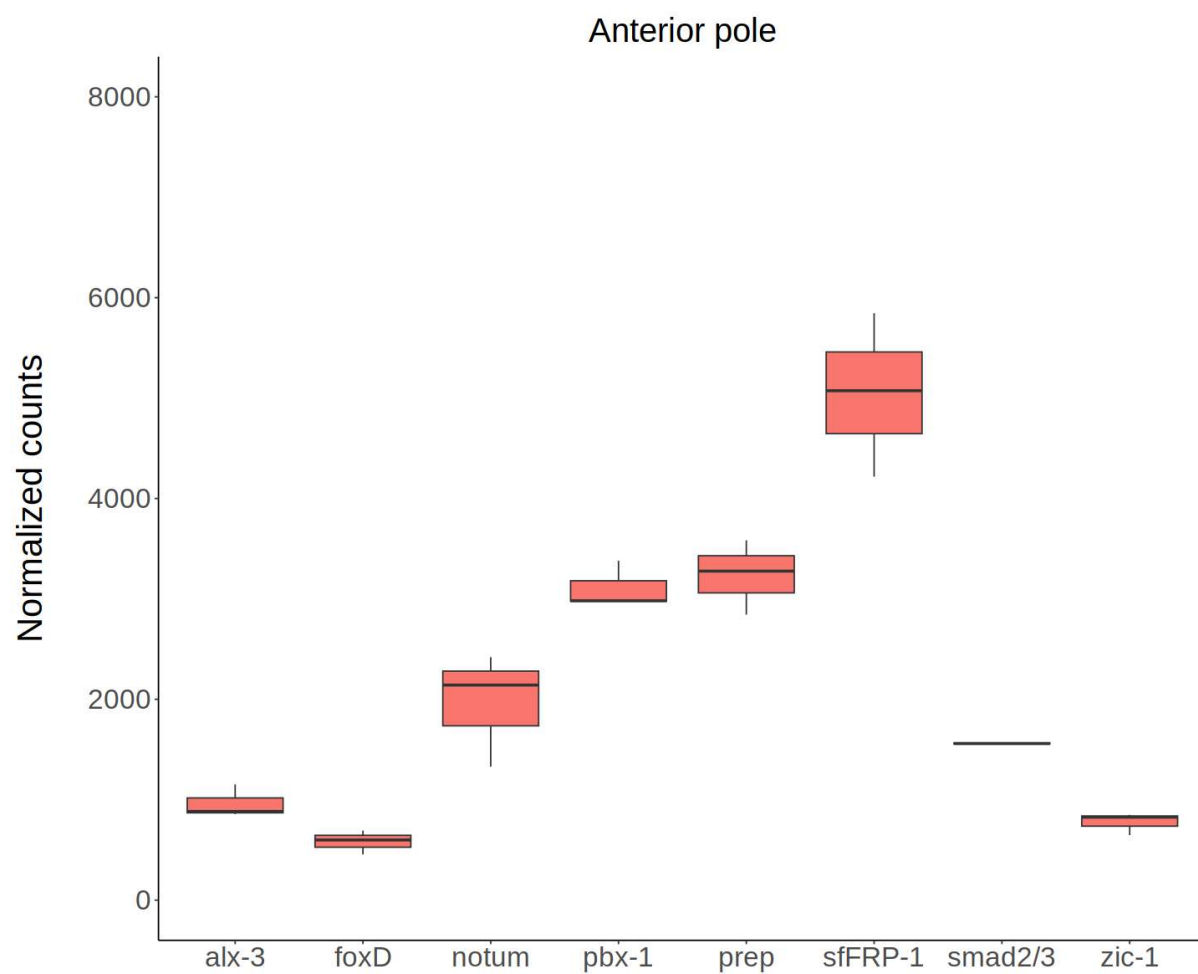


Figure S4.2 Expression of genes in the anterior pole. Normalised counts of the expression of *alx-3*, *foxD*, *notum*, *pbx-1*, *prep*, *sfFRP-1*, *smad2/3*, and *zic-1* in the anterior pole based on bulk RNA-seq data provided by Li et al. (2019).

Chapter 5. General discussion

5.1 Synopsis

Regeneration necessitates regulatory mechanisms that govern both cell behaviour and the spatiotemporal organisation of cells in the regrowing tissue. The aim of this thesis is to understand these regulatory mechanisms in the highly regenerative planarian species *Schmidtea mediterranea*, thereby advancing our knowledge of stem cells and positional information systems during regeneration. Advances in our understanding of regenerative biology are a prerequisite for any translational applications in regenerative medicine. In Chapter 1, I introduce planarians and their stem cells, review comprehensively the mechanisms of TALE factors in metazoans, and attempt to synthesise our current understanding of pattern formation in regenerating planarians.

In Chapter 2, I create an expression-driven reannotation of the genome of *S. mediterranea* and compile a list of transcription factors through bioinformatic screening and manual review. Moreover, I integrate evidence from histone modifications (ChIP-seq: H3K4me1 and H3K27ac) and chromatin accessibility (ATAC-seq) to identify enhancer-like regions that are active in the genome of planarian neoblasts. By using footprints of transcription factors in these enhancer-like regions, I construct a model of gene regulatory networks in planarian neoblasts.

In Chapter 3, I continue to investigate the class of TALE homeobox genes that are ancestral to metazoans, regulate patterning and cell specification in multiple developmental contexts, and have also been implicated in the regenerative patterning of planarians. Given that our transcription factor screening uncovered several uncharacterised TALE factors and that TALE factors have recently been shown to have diversified greatly in other spiralian, I performed a phylogenetic analysis of metazoan and planarian TALE factors to understand their evolution. In the genome of *S. mediterranea*, I discovered representatives of conserved TALE classes: Irx (*irx-1*, *irx-2*, *irx-3*, *irx-4*), Pbx (*pbx-1* and *pbx-2*), Prep (*prep*), Meis (*meis-1*, *meis-2*, and *meis-like*). Furthermore, I found multiple divergent, lineage-specific TALE genes in the genome of *S. mediterranea*, including representatives of the spiralian-specific clade SPILE (*SPILE-Z1*, *SPILE-Z2*, *SPILE-E1*, *SPILE-E2*) and a platyhelminth-specific gene *PALE*. Intriguingly, preliminary functional studies suggest that *SPILE-Z1* and *PALE* have adopted functions beyond early development in both regeneration and adult physiology.

In Chapter 4, I continue to study the regulatory role of the TALE genes *pbx-1*, *prep*, and *meis-2* in anterior and posterior patterning of the regenerating planarian, aiming to uncover direct downstream targets. I characterise the knockdown phenotype of *meis-2* and demonstrate that it is required for posterior specification in regeneration. To discover the underlying regulatory mechanisms, I perform RNA-seq and ATAC-seq in anterior and posterior blastemas to predict direct targets of these TALE genes. Analysis of transcriptomic and epigenomic data along with initial functional experiments suggest that anterior pole markers and the FGFRL family are targets of *pbx-1* and *prep* in the anterior, while the posterior pole marker *wnt1* and other genes of the posterior specification program are targets of *pbx-1* and *meis-2* in the posterior. Furthermore, I leverage the blastemal ATAC-seq data alongside existing planarian literature to predict high-confidence regulatory interactions

5.2 Bioinformatics for the planarian community

The reannotation of the *S. mediterranea* genome and the transcription factor list have proven beneficial to the planarian research community. Prior to publication, I had the honour to contribute with the annotation to improve the resolution of mapping cell type composition in both *S. mediterranea* and *D. japonica* (García-Castro et al. 2021). Datasets for an expression-driven annotation of *S. mediterranea* were readily available, but the reannotation of *D. japonica* required data mining of less well-known Japanese datasets (García-Castro et al. 2021). In my opinion, this demonstrated the importance of careful management, curation and sharing

of bioinformatic resources. Therefore, I have sought to make my analysis readily available to the international planarian community, and I wish to believe that my correspondence with researchers from the University of Tokyo has helped to improve the strength of bioinformatic analyses for *D. japonica*. Unexpectedly, *D. japonica* has also caught the interest of representatives beyond academia, as some longevity entrepreneurs have contacted me regarding the reannotation. Overall, I am delighted that the methodology presented in the second chapter has contributed to other projects both within and beyond the planarian field.

Prior to chapter 2, only bioinformatically screened lists of transcription factors existed (Swapna et al., 2018). However, to better understand the results of the constructed gene regulatory networks, I recognised the need for further manual curation and the establishment of a comprehensive checklist for planarian transcription factors to ensure consistency with the existing planarian literature. With regard to the transcription factor list, I would like to thank Alberto Perez-Posada for our brief but fruitful discussions on the state of planarian bioinformatics and the necessary standards for performing and reporting computational analyses. In this vein, I am pleased to see that the building blocks generated in the second chapter have contributed to shaping the inference of gene regulatory networks in the first integration of planarian scRNA-seq and scATAC-seq data (Pérez-Posada et al., 2024).

Furthermore, it was an insightful opportunity to present the findings of the second chapter to the wider community of regenerative biologists at the 8th EMBO conference in the series *The Molecular and Cellular Basis of Regeneration and Tissue Repair* in Barcelona in September 2022 (Amiel et al., 2023). In particular, methodological discussions with Ashley Maynard helped to contrast the data and limitations of the second chapter with scRNA-seq approaches in the axolotl telencephalon (Lust et al., 2022; Amiel, et al., 2023). As the field of regenerative biology encompasses many model systems with varying degrees of experimental amenability, I believe that the development of computational pipelines integrating different experimentally available datasets will be necessary for advancing our understanding of regenerative mechanisms.

5.3 Evolution of animal stem cells

Ever since the discovery of the pluripotency network in mammalian embryonic stem cells and the Yamanaka or OSKM (Oct4, Sox2, Klf, c-Myc) factors, planarian neoblasts have been studied to understand the conservation of mechanisms in pluripotent stem cells (Takahashi and Yamanaka, 2006; Takahashi et al. 2007; Labbé et al., 2012; Lai and Aboobaker, 2018; Srivastava et al., 2021). In early 2020, we hypothesised that our epigenomic data would reveal enhancers enriched in association with neoblast-specific genes, and that the binding sites of a transcription factor conferring pluripotency would be enriched in these. The publication of the fate switch mechanism by Raz et al. (2021) was instrumental in shaping our analysis after the summer of 2021, and indeed, we uncovered that enhancer-linked genes were enriched for fate-specific transcription factors. Although we also discovered several enhancers linked to previously uncharacterised planarian transcription factors, my interpretation of the second chapter in light of current research is that the gene regulatory network active in X1 neoblasts provides support for the fate specification model (Raz et al., 2021; Reddien et al., 2022). Nonetheless, the hypothetical “pluripotency factor” could still reside within e.g. the unstudied set of zinc finger transcription factors. Further phylogenetic studies and a systematic effort to knock down all the remaining unstudied transcription factors in *S. mediterranea* would provide more clarity on this question.

Interestingly, a year after the publication of the second chapter, Poulet et al. (2023) compared the chromatin state of the promoters of neoblast-specific genes, constitutively active genes, and tissue-specific genes. While tissue-specific genes had distinct upstream accessible regions containing transcription factor motifs, the promoters of neoblast-specific genes lacked the same chromatin signature and were devoid of enriched transcription factor motifs, suggesting that no master transcription factor regulates pluripotency (Poulet et al., 2023). As the chromatin state of neoblast-specific genes resembles the chromatin organisation of

constitutive genes, Poulet et al. (2023) conclude that pluripotency is the default state of neoblasts and primarily maintained by chromatin remodelers. Therefore, Poulet et al. (2023) reproduced our finding that regulatory interactions between fate-specific transcription factors are more easily detected in epigenomic data derived from neoblasts. However, Poulet et al. (2023) restrict their analysis to promoters and proximal upstream regions and do not analyse intronic accessible regions, distal cis-regulatory regions, including downstream regions of neoblast-specific genes. Moreover, Poulet et al. (2023) take a system-wide approach, and analyse neoblast-specific genes as a group, meaning that a few restricted regulatory interactions between a pluripotency factor and neoblast-specific genes could be hidden in the dataset and might be detectable with a more reductionist method. Thus, experiments detecting the binding of transcription factors in the vicinity of neoblast-specific genes would shed more light on the matter. Nonetheless, it appears that we, Raz et al. (2021), and Poulet et al. (2023) independently converged on the same aspect of neoblast gene regulation, where the activity of fate-specific transcription factors and their cross-regulation emerge from analysis of S/G2/M neoblasts.

In addition to the embryonic pluripotency network, planarian neoblasts and their regulatory mechanisms have been compared to germ cells and the germline multipotency program, as summarised in the introduction. Thus, it appears that adult pluripotent *pimi*⁺ planarian stem cells have been compared in terms of the “pluripotency” and “germline” components, but the adult stem cell aspect has received less attention. Following my presentation of Chapter 2 at the conference *Gene Regulatory Networks in Development and Disease* in April 2023 here at the University of Oxford, Ellen Rothenberg was thrilled to see that many genes essential for hematopoietic stem cells were enriched for binding motifs in planarian neoblasts (Figure 2.2). Although presented as a passing thought, I have not been able to fully set aside this idea: Is there a shared component of potency in adult animal stem cells that is magnified to pluripotent dimensions in planarian neoblasts?

5.4 TALE factors and their tale on molecular evolution

Given their highly regenerative abilities, the conserved properties of planarian regeneration and their potential for translational therapies have been the focus of most planarian research, while many aspects of planarian biology have been overlooked. Furthermore, planarians belong to the spiralian superphylum, which has not been as well characterised as deuterostomes and ecdysozoans in general. Therefore, an understanding of the molecular evolution of transcription factors in planarians contributes to our knowledge of the evolution of spiralian body plans and the mechanisms of metazoan evolutionary developmental biology.

The hourglass model of embryonic development posits that both early and late stages of embryogenesis display more differences in comparison between taxa and are more evolvable (Duboule, 1994; Irie and Kuratani, 2014). Originally based on the conserved Hox cluster, Duboule (1994) theorised that interlinkage between Hox genes constituted a constrained module that rendered the phylotypic stage conserved. Divergence in early development has been proposed to result from selection for different reproductive strategies. Since morphological differences are difficult to quantify, comparisons in gene expression between species have attempted to assess whether the transcriptional profile over embryogenesis adheres to the hourglass model (Irie and Kuratani, 2014). In these studies, evidence for and against the hourglass model has been found, but they are still limited in the number of species examined, being mainly restricted to some vertebrates, insects, and nematodes (Irie and Kuratani, 2014). Although the third chapter only examines a single family of homeobox genes in a single species, the results can still be discussed in the wider framework of the hourglass model. In our expression analysis, we found that the expression of most novel lineage-specific factors is biased towards early development, while the expression of conserved TALE genes is strongest at mid-developmental stages, thus providing some support for the hourglass model. Overall,

the results in the second chapter underscore the need to incorporate spiralian evo-devo into the wider framework of theories on the evolution of development.

Morphological evolution has been proposed to be partially caused by modifications in the cis-regulatory code of conserved developmental genes (Davidson and Erwin, 2006; Carroll, 2008), while the diversification of spiralian TALE genes represents one of the largest known expansion of homeobox genes in metazoans (Paps et al., 2015; Holland et al., 2017; Morino et al., 2017; Phuangphong et al., 2021; Morino et al., 2022; Morino and Yoshikawa, 2023). Previously mainly implicated in early development, the third chapter extends the functional diversification of spiralian lineage-specific genes by providing evidence for additional roles in adult physiology and regenerative patterning. Therefore, spiralian-specific TALE factors increasingly exemplify the idea of evolution by gene duplication (Bridges, 1936; Stephens, 1951; Ohno, 1970); duplications in planarian TALE genes appear to create genetic redundancy and novel genetic material for evolution to tinker with.

5.5 TALE factors in the shadow of the Wnt gradient

In chapter 4, I present some evidence supporting a TALE/Wnt regulatory relationship in axial patterning. Specifically, my analysis suggests that a Pbx-1:Meis-2 dimer could be binding to intronic enhancers of the *wnt1* gene and regulate its expression in the posterior pole. In addition, the dimeric Pbx-1:Meis-2 complex could also be activating the expression of *wnt11-2* and its putative receptor *fzd4-1* (Pascual-Carreras et al., 2021). In contrast, TALE factors have been shown to make chromatin accessible for the Wnt effector LEF1 in vertebrate tailbuds, indicating that TALE factors act in parallel rather than upstream of Wnt signalling (Mariani et al., 2021). Nonetheless, *meis-2* and *pbx-1* could also act earlier to unlock the response of the early phase of *wnt1* expression, but further experiments are required to assess this possibility. Overall, although the interpretation of the fourth chapter is limited by the lack of experimental validations, it suggests a conserved role for Pbx and Meis in Wnt-directed anterior-posterior patterning.

As transgenic reporter lines are not available in planarians, the use of other model species could prove fruitful in understanding the role of TALE factors in Wnt-mediated axial regeneration and patterning. In the last decade, the acoel worm *Hofstenia miamia* has emerged as a promising model for understanding Wnt signalling dynamics of whole-body regeneration (Srivastava et al., 2014; Raz et al., 2017; Ramirez et al., 2020). As in planarians, Wnt signalling is required for posterior regeneration, is upregulated in the posterior following amputation, and the core set of genes in the posterior specification program is shared between these species (Srivastava et al., 2014; Tewari et al., 2019; Ramirez et al., 2020). As the toolset is rapidly increasing in acoels and is surpassing that of planarians (Ricci and Srivastava, 2021), acoels could prove to be a suitable model system for continuing the research direction into the interplay between Wnt, Hox, and TALE in axial regeneration and the conservation of axial patterning mechanisms.

5.6 The enigma of anterior regeneration – does *nou-darake* regulate self-organising cell adhesion?

Interestingly, the putative anterior targetome of *pbx-1* and *prep* includes the enigmatic class of FGFRL family members, whose signalling mechanisms are poorly understood in planarians and in metazoans in general (Adell et al., 2010; Agata et al., 2014). As summarised in the introduction, Fgf family members have been proposed to act as decoy receptors, capturing Fgf ligands and thus restricting the diffusion of a putative anterior morphogen. However, two decades of sequencing efforts have not found the “head activator”, while inhibitors of Wnt signalling are found in the anterior blastema (asymmetric expression of *notum* at 18 hpa) and in the anterior pole. Therefore, I find it unlikely that further sequencing efforts would uncover a prominent anterior morphogen, and the field has to be open to other mechanisms for anterior patterning.

A number of authors have proposed the idea that the anterior fate could be the default of planarian neoblasts (Adell et al., 2010; Agata et al., 2014). Recently, regeneration of planarian eyes has been shown to

be guided by self-organising principles (Reddien et al., 2021), raising the possibility that cellular self-organisation could be more widely involved in anterior patterning. In an overlooked piece of work, Yang et al. (2016) found FGFR1 to regulate cell adhesion in cultured cells. In the zebrafish embryo, a cell adhesion code has been shown to mediate the patterning of neural progenitor cells in the spinal cord (Tsai et al., 2020). Therefore, self-organisation through cell adhesion could provide further answers to the signalling of ndk and ndl-1 to ndl-5 in planarian anterior regeneration. As this idea is based on singular studies, further experiments would be required to establish rigour; yet to my knowledge, gradient-less alternatives to the FGFR1 signalling problem have not been proposed or investigated in planarian regeneration, thus providing a new possible direction for further studies.

5.7 Coda

“If there were no regeneration, there would be no life. If everything regenerated, there would be no death. All organisms exist between these two themes.” (Goss, 1969). Planarians are at one of these extremes, displaying remarkable developmental plasticity and robustness. The elusive aspects of planarian regeneration have been studied for centuries, yet we still know very little of how the components of the regenerative machinery are regulated. Time will tell how well this thesis has succeeded in understanding these regulatory mechanisms.

Acknowledgements

I wish to thank my supervisor, Aziz Aboobaker, for introducing me to the wondrous world of worm wizardry at the University of Oxford. As an undergraduate student, I wrote my Bachelor's thesis on planarian regeneration, which inspired me to choose this as the subject for my DPhil. I am grateful for the opportunity and mentorship he has provided me over the years.

I would also like to thank all past and present members of the Aboobaker laboratory for helping me conduct my research. I would especially like to thank Zhong Cheng Yap, who taught me hands-on the experimental techniques used in this thesis, and the worms he so kindly provided for several of the experiments reported in this work.

I would like to thank the Scandinavian community in Oxford for providing well-needed extracurricular activities, and the Queen's College for providing financial support and a beautiful library for my academic pursuits. I would also like to thank the spirit of the MBL 2024 course, who gave me the extra energy to finalise this thesis.

I wish to thank Tuomas and Niko for being excellent fellow jury members at the International Biology Olympiad. I am also indebted to Niko for all the fruitful discussions we have had over the years.

I wish to thank my scouting friends Joshua and Abigail for their support, and the Finnjamboree 2022 team for providing me with such a wonderful opportunity to learn how to manage a data analysis project.

Lastly, I wish to thank my family for always being supportive of my academic endeavours and for nurturing the zoological interest I had as a child. I hope this thesis shows that it was not in vain.

References

- Aboobaker A. A. (2011). Planarian stem cells: a simple paradigm for regeneration. *Trends in cell biology*, 21(5), 304–311.
- Abnave, P., Aboukhatwa, E., Kosaka, N., Thompson, J., Hill, M. A., & Aboobaker, A. A. (2017). Epithelial-mesenchymal transition transcription factors control pluripotent adult stem cell migration *in vivo* in planarians. *Development (Cambridge, England)*, 144(19), 3440–3453.
- Abramovich, C., Shen, W. F., Pineault, N., Imren, S., Montpetit, B., Largman, C., & Humphries, R. K. (2000). Functional cloning and characterization of a novel nonhomeodomain protein that inhibits the binding of PBX1-HOX complexes to DNA. *The Journal of biological chemistry*, 275(34), 26172–26177.
- Abu-Shaar, M., & Mann, R. S. (1998). Generation of multiple antagonistic domains along the proximodistal axis during *Drosophila* leg development. *Development (Cambridge, England)*, 125(19), 3821–3830.
- Abu-Shaar, M., Ryoo, H. D., & Mann, R. S. (1999). Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. *Genes & development*, 13(8), 935–945.
- Adell, T., Salò, E., Boutros, M., & Bartscherer, K. (2009). Smed-Evi/Wntless is required for beta-catenin-dependent and -independent processes during planarian regeneration. *Development (Cambridge, England)*, 136(6), 905–910.
- Adell, T., Cebrià, F., & Saló, E. (2010). Gradients in planarian regeneration and homeostasis. *Cold Spring Harbor perspectives in biology*, 2(1), a000505.
- Adler, C. E., Seidel, C. W., McKinney, S. A., & Sánchez Alvarado, A. (2014). Selective amputation of the pharynx identifies a FoxA-dependent regeneration program in planaria. *eLife*, 3, e02238.
- Adler, C. E., & Sánchez Alvarado, A. (2015). Types or States? Cellular Dynamics and Regenerative Potential. *Trends in cell biology*, 25(11), 687–696.
- Adler, C. E., & Sánchez Alvarado, A. (2017). PHRED-1 is a divergent neurexin-1 homolog that organizes muscle fibers and patterns organs during regeneration. *Developmental biology*, 427(1), 165–175.
- Agata, K., Tanaka, T., Kobayashi, C., Kato, K., & Saitoh, Y. (2003). Intercalary regeneration in planarians. *Developmental dynamics : an official publication of the American Association of Anatomists*, 226(2), 308–316.
- Agata, K., Tasaki, J., Nakajima, E., & Umesono, Y. (2014). Recent identification of an ERK signal gradient governing planarian regeneration. *Zoology (Jena, Germany)*, 117(3), 161–162.
- Agoston, Z., & Schulte, D. (2009). Meis2 competes with the Groucho co-repressor Tle4 for binding to Otx2 and specifies tectal fate without induction of a secondary midbrain-hindbrain boundary organizer. *Development (Cambridge, England)*, 136(19), 3311–3322.
- Agoston, Z., Li, N., Haslinger, A., Wizenmann, A., & Schulte, D. (2012). Genetic and physical interaction of Meis2, Pax3 and Pax7 during dorsal midbrain development. *BMC developmental biology*, 12, 10.
- Agoston, Z., Heine, P., Brill, M. S., Grebbin, B. M., Hau, A. C., Kallenborn-Gerhardt, W., Schramm, J., Götz, M., & Schulte, D. (2014). Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development (Cambridge, England)*, 141(1), 28–38.
- Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J. C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods*, 14(11), 1083–1086.
- Akheralie, Z., Scidmore, T. J., & Pearson, B. J. (2023). aristaless-like homeobox-3 is wound induced and promotes a low-Wnt environment required for planarian head regeneration. *Development (Cambridge, England)*, 150(18), dev201777.
- Almuedo-Castillo, M., Saló, E., & Adell, T. (2011). Dishevelled is essential for neural connectivity and planar cell polarity in planarians. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2813–2818.
- Almuedo-Castillo, M., Crespo-Yanez, X., Seebeck, F., Bartscherer, K., Saló, E., & Adell, T. (2014). JNK controls the onset of mitosis in planarian stem cells and triggers apoptotic cell death required for regeneration and remodeling. *PLoS genetics*, 10(6), e1004400.

- Amiel, A. R., Tsai, S. L., & Wehner, D. (2023). Embracing the diversity of model systems to deconstruct the basis of regeneration and tissue repair. *Development (Cambridge, England)*, 150(3), dev201579.
- Amin, S., Donaldson, I. J., Zannino, D. A., Hensman, J., Ratray, M., Losa, M., Spitz, F., Ladam, F., Sagerström, C., & Bobola, N. (2015). Hoxa2 selectively enhances Meis binding to change a branchial arch ground state. *Developmental cell*, 32(3), 265–277.
- Anderson, D. M., Arredondo, J., Hahn, K., Valente, G., Martin, J. F., Wilson-Rawls, J., & Rawls, A. (2006). Mohawk is a novel homeobox gene expressed in the developing mouse embryo. *Developmental dynamics : an official publication of the American Association of Anatomists*, 235(3), 792–801.
- Antosova, B., Smolikova, J., Klimova, L., Lachova, J., Bendova, M., Kozmikova, I., Machon, O., & Kozmik, Z. (2016). The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS genetics*, 12(12), e1006441.
- Auwal, M. A., Kashima, M., Nishimura, O., Hosoda, K., Motoishi, M., Kamimura, A., Okumura, A., Agata, K., & Umesono, Y. (2020). Identification and characterization of a fibroblast growth factor gene in the planarian *Dugesia japonica*. *Development, growth & differentiation*, 62(9), 527–539.
- Ariki, R., Morikawa, S., Mabuchi, Y., Suzuki, S., Nakatake, M., Yoshioka, K., Hidano, S., Nakauchi, H., Matsuzaki, Y., Nakamura, T., & Goitsuka, R. (2014). Homeodomain transcription factor Meis1 is a critical regulator of adult bone marrow hematopoiesis. *PLoS one*, 9(2), e87646.
- Asahara, H., Dutta, S., Kao, H. Y., Evans, R. M., & Montminy, M. (1999). Pbx-Hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner. *Molecular and cellular biology*, 19(12), 8219–8225.
- Aspland, S. E., & White, R. A. (1997). Nucleocytoplasmic localisation of extradenticle protein is spatially regulated throughout development in *Drosophila*. *Development (Cambridge, England)*, 124(3), 741–747.
- Atabay, K. D., LoCascio, S. A., de Hoog, T., & Reddien, P. W. (2018). Self-organization and progenitor targeting generate stable patterns in planarian regeneration. *Science (New York, N.Y.)*, 360(6387), 404–409.
- Aubrey, B. J., Cutler, J. A., Bourgeois, W., Donovan, K. A., Gu, S., Hatton, C., Perlee, S., Perner, F., Rahnamoun, H., Theall, A. C. P., Henrich, J. A., Zhu, Q., Nowak, R. P., Kim, Y. J., Parvin, S., Cremer, A., Olsen, S. N., Eleuteri, N. A., Pikman, Y., McGeehan, G. M., ... Armstrong, S. A. (2022). IKAROS and MENIN coordinate therapeutically actionable leukemogenic gene expression in MLL-r acute myeloid leukemia. *Nature cancer*, 3(5), 595–613.
- Ayyar, S., Jiang, J., Collu, A., White-Cooper, H., & White, R. A. (2003). *Drosophila* TGIF is essential for developmentally regulated transcription in spermatogenesis. *Development (Cambridge, England)*, 130(13), 2841–2852.
- Azcoitia, V., Aracil, M., Martínez-A, C., & Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Developmental biology*, 280(2), 307–320.
- Badia-I-Mompel, P., Wessels, L., Müller-Dott, S., Trimbou, R., Ramirez Flores, R. O., Argelaguet, R., & Saez-Rodriguez, J. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nature reviews. Genetics*, 24(11), 739–754.
- Baguña, J. (1976). Mitosis in the intact and regenerating planarian *Dugesia mediterranea* n. sp. II. Mitotic studies during regeneration, and a possible mechanism of blastema formation. *Journal of Experimental Zoology*, 195(1), 65–79.
- Baguña, J., & Romero, R. (1981). Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hydrobiologia*, 84, 181–194.
- Baguña, J., Saló, E., & Auladell, C. (1989). Regeneration and pattern formation in planarians: III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development*, 107(1), 77–86.
- Bailey, J. S., Rave-Harel, N., McGillivray, S. M., Coss, D., & Mellon, P. L. (2004). Activin regulation of the follicle-stimulating hormone beta-subunit gene involves Smads and the TALE homeodomain proteins Pbx1 and Prep1. *Molecular endocrinology (Baltimore, Md.)*, 18(5), 1158–1170.
- Balavoine, G., & Telford, M. J. (1995). Identification of planarian homeobox sequences indicates the antiquity of most Hox/homeotic gene subclasses. *Proceedings of the National Academy of Sciences of the United States of America*, 92(16), 7227–7231.

- Barton-Owen, T. B., Szabó, R., Somorjai, I. M. L., & Ferrier, D. E. K. (2018). A Revised Spiralian Homeobox Gene Classification Incorporating New Polychaete Transcriptomes Reveals a Diverse TALE Class and a Divergent Hox Gene. *Genome biology and evolution*, 10(9), 2151–2167.
- Baysoy, A., Bai, Z., Satija, R., & Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. *Nature reviews. Molecular cell biology*, 24(10), 695–713.
- Beane, W. S., Morokuma, J., Adams, D. S., & Levin, M. (2011). A chemical genetics approach reveals H,K-ATPase-mediated membrane voltage is required for planarian head regeneration. *Chemistry & biology*, 18(1), 77–89.
- Bely, A. E., & Nyberg, K. G. (2010). Evolution of animal regeneration: re-emergence of a field. *Trends in ecology & evolution*, 25(3), 161–170.
- Benham-Pyle, B. W., Brewster, C. E., Kent, A. M., Mann, F. G., Jr, Chen, S., Scott, A. R., Box, A. C., & Sánchez Alvarado, A. (2021). Identification of rare, transient post-mitotic cell states that are induced by injury and required for whole-body regeneration in *Schmidtea mediterranea*. *Nature cell biology*, 23(9), 939–952.
- Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., Kim, J., & Looso, M. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nature communications*, 11(1), 4267.
- Berenguer, M., Meyer, K. F., Yin, J., & Duester, G. (2020). Discovery of genes required for body axis and limb formation by global identification of retinoic acid-regulated epigenetic marks. *PLoS biology*, 18(5), e3000719.
- Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S., & Tapscott, S. J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Molecular cell*, 14(4), 465–477.
- Berry, F. B., O'Neill, M. A., Coca-Prados, M., & Walter, M. A. (2005). FOXC1 transcriptional regulatory activity is impaired by PBX1 in a filamin A-mediated manner. *Molecular and cellular biology*, 25(4), 1415–1424.
- Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F., & Blasi, F. (1998a). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *The EMBO journal*, 17(5), 1434–1445.
- Berthelsen, J., Zappavigna, V., Mavilio, F., & Blasi, F. (1998b). Prep1, a novel functional partner of Pbx proteins. *The EMBO journal*, 17(5), 1423–1433.
- Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., & Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes & development*, 13(8), 946–953.
- Bertolino, E., Reimund, B., Wildt-Perinic, D., & Clerc, R. G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *The Journal of biological chemistry*, 270(52), 31178–31188.
- Bessa, J., Gebelein, B., Pichaud, F., Casares, F., & Mann, R. S. (2002). Combinatorial control of *Drosophila* eye development by eyeless, homothorax, and teashirt. *Genes & development*, 16(18), 2415–2427.
- Bessa, J., Tavares, M. J., Santos, J., Kikuta, H., Laplante, M., Becker, T. S., Gómez-Skarmeta, J. L., & Casares, F. (2008). *meis1* regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. *Development (Cambridge, England)*, 135(5), 799–803.
- Bertrand, S., Somorjai, I., Garcia-Fernandez, J., Lamonerie, T., & Escriva, H. (2009). FGFR1 is a neglected putative actor of the FGF signalling pathway present in all major metazoan phyla. *BMC evolutionary biology*, 9, 226.
- Bharathan, G., Janssen, B. J., Kellogg, E. A., & Sinha, N. (1997). Did homeodomain proteins duplicate before the origin of angiosperms, fungi, and metazoa?. *Proceedings of the National Academy of Sciences of the United States of America*, 94(25), 13749–13753.
- Bienvenu, F., Jirawatnotai, S., Elias, J. E., Meyer, C. A., Mizeracka, K., Marson, A., Frampton, G. M., Cole, M. F., Odom, D. T., Odajima, J., Geng, Y., Zagodzón, A., Jecrois, M., Young, R. A., Liu, X. S., Cepko, C. L., Gygi, S. P., & Sicinski, P. (2010). Transcriptional role of cyclin D1 in development revealed by a genetic-proteomic screen. *Nature*, 463(7279), 374–378.

- Bisaillon, R., Wilhelm, B. T., Kros, J., & Sauvageau, G. (2011). C-terminal domain of MEIS1 converts PKNOX1 (PREP1) into a HOXA9-collaborating oncoprotein. *Blood*, 118(17), 4682–4689.
- Bjerke, G. A., Hyman-Walsh, C., & Wotton, D. (2011). Cooperative transcriptional activation by Klf4, Meis2, and Pbx1. *Molecular and cellular biology*, 31(18), 3723–3733.
- Blanco-Arias, P., Sargent, C. A., & Affara, N. A. (2002). The human-specific Yp11. 2/Xq21. 3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. *Mammalian genome*, 13(8), 463.
- Blasi, F., Bruckmann, C., Penkov, D., & Dardaei, L. (2017). A tale of TALE, PREP1, PBX1, and MEIS1: Interconnections and competition in cancer. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 39(5), 10.1002/bies.201600245.
- Blassberg, R. A., Felix, D. A., Tejada-Romero, B., & Aboobaker, A. A. (2013). PBX/extradenticle is required to re-establish axial structures and polarity during planarian regeneration. *Development (Cambridge, England)*, 140(4), 730–739.
- Bobola, N., & Sagerström, C. G. (2024). TALE transcription factors: Cofactors no more. *Seminars in cell & developmental biology*, 152-153, 76–84.
- Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. (2015). msa: an R package for multiple sequence alignment. *Bioinformatics (Oxford, England)*, 31(24), 3997–3999.
- Bohr, T. E., Shiroor, D. A., & Adler, C. E. (2021). Planarian stem cells sense the identity of the missing pharynx to launch its targeted regeneration. *eLife*, 10, e68830.
- Bonnet, C. (1745). *Traité d'insectologie ou Observations sur quelques espèces de vers d'eau douce, qui coupés par morceaux, deviennent autant d'animaux complets*, Imprint Paris : chez Durand, Libraire, M.DC.XLV
- Bonar, N. A., Gittin, D. I., & Petersen, C. P. (2022). Src acts with WNT/FGFR signaling to pattern the planarian anteroposterior axis. *Development (Cambridge, England)*, 149(7), dev200125.
- Booth, C. L., Stevens, B. C., Stubbett, C. A., Kallgren, N. T., & Davies, E. L. (2024). Axis reset is rate limiting for onset of whole-body regenerative abilities during planarian development. *bioRxiv*, 2024-07.
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2), 311–322.
- Brandl, H., Moon, H., Vila-Farré, M., Liu, S. Y., Henry, I., & Rink, J. C. (2016). PlanMine--a mineable resource of planarian biology and biodiversity. *Nucleic acids research*, 44(D1), D764–D773.
- Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., & Aerts, S. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods*, 20(9), 1355–1367.
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5), 525–527.
- Brenner S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics*, 77(1), 71–94.
- Bridges C. B. (1936). The bar "gene" a duplication. *Science (New York, N.Y.)*, 83(2148), 210–211.
- Bridoux, L., Zarrineh, P., Mallen, J., Phuycharoen, M., Latorre, V., Ladam, F., Losa, M., Baker, S. M., Sagerstrom, C., Mace, K. A., Rattray, M., & Bobola, N. (2020). HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation. *PLoS genetics*, 16(12), e1009162.
- Britten, R. J., & Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science (New York, N.Y.)*, 165(3891), 349–357.
- Brøndsted, H. V. (1955). Planarian regeneration. *Biological Reviews*, 30(1), 65-126.
- Brown, D. D. R., Molinaro, A. M., & Pearson, B. J. (2018). The planarian TCF/LEF factor Smed-tcf1 is required for the regeneration of dorsal-lateral neuronal subtypes. *Developmental biology*, 433(2), 374–383.

- Browne, E. N. (1909). The production of new hydranths in *Hydra* by the insertion of small grafts. *J Exp Zool*, 7(1), 1–23.
- Bruce, H. S., & Patel, N. H. (2020). Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. *Nature ecology & evolution*, 4(12), 1703–1712.
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), 1213–1218.
- Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., & Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561), 486–490.
- Bürglin T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic acids research*, 25(21), 4173–4180.
- Bürglin, T. R., & Affolter, M. (2016). Homeodomain proteins: an update. *Chromosoma*, 125(3), 497–521.
- Cao, Y., Xia, Y., Balowski, J. J., Ou, J., Song, L., Safi, A., ... & Cao, J. (2022). Identification of enhancer regulatory elements that direct epicardial gene expression during zebrafish heart regeneration. *Development*, 149(4), dev200133.
- Capdevila, J., Tsukui, T., Rodríguez Esteban, C., Zappavigna, V., & Izpisua Belmonte, J. C. (1999). Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Molecular cell*, 4(5), 839–849.
- Čapek, D., & Müller, P. (2019). Positional information and tissue scaling during development and regeneration. *Development (Cambridge, England)*, 146(24), dev177709.
- Capellini, T. D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V., & Selleri, L. (2006). Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development (Cambridge, England)*, 133(11), 2263–2273.
- Capellini, T. D., Zewdu, R., Di Giacomo, G., Ascitti, S., Kugler, J. E., Di Gregorio, A., & Selleri, L. (2008). Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. *Developmental biology*, 321(2), 500–514.
- Capellini, T. D., Zappavigna, V., & Selleri, L. (2011). Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Developmental dynamics : an official publication of the American Association of Anatomists*, 240(5), 1063–1086.
- Cardona, A., Hartenstein, V., & Romero, R. (2005). The embryonic development of the triclad Schmidtea polychroa. *Development genes and evolution*, 215, 109–131.
- Carroll S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1), 25–36.
- Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2013). From DNA to diversity: molecular genetics and the evolution of animal design. *John Wiley & Sons*.
- Casares, F., & Mann, R. S. (1998). Control of antennal versus leg development in *Drosophila*. *Nature*, 392(6677), 723–726.
- Cebrià, F., Kobayashi, C., Umesono, Y., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T., Itoh, M., Taira, M., Sánchez Alvarado, A., & Agata, K. (2002). FGFR-related gene *nou-darake* restricts brain tissues to the head region of planarians. *Nature*, 419(6907), 620–624.
- Cebrià, F., Guo, T., Jopek, J., & Newmark, P. A. (2007). Regeneration and maintenance of the planarian midline is regulated by a slit orthologue. *Developmental biology*, 307(2), 394–406.
- Cebrià F. (2016). Planarian Body-Wall Muscle: Regeneration and Function beyond a Simple Skeletal Support. *Frontiers in cell and developmental biology*, 4, 8.
- Cebrià, F., Adell, T., & Saló, E. (2018). Rebuilding a planarian: from early signaling to final shape. *The International journal of developmental biology*, 62(6-7-8), 537–550.

- Chai, G., Ma, C., Bao, K., Zheng, L., Wang, X., Sun, Z., Salò, E., Adell, T., & Wu, W. (2010). Complete functional segregation of planarian beta-catenin-1 and -2 in mediating Wnt signaling and cell adhesion. *The Journal of biological chemistry*, 285(31), 24120–24130.
- Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., & Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, 78(4), 603–615.
- Chan, S. K., Pöpperl, H., Krumlauf, R., & Mann, R. S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *The EMBO journal*, 15(10), 2476–2487.
- Chan, S. K., & Mann, R. S. (1996). A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proceedings of the National Academy of Sciences of the United States of America*, 93(11), 5223–5228.
- Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., & Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes & development*, 9(6), 663–674.
- Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., & Cleary, M. L. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. *Molecular and cellular biology*, 16(4), 1734–1745.
- Chang, C. P., Jacobs, Y., Nakamura, T., Jenkins, N. A., Copeland, N. G., & Cleary, M. L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Molecular and cellular biology*, 17(10), 5679–5687.
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution: Molecules, networks, populations* (pp. 207-232). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Chen, C. C., Wang, I. E., & Reddien, P. W. (2013). pbx is required for pole and eye regeneration in planarians. *Development (Cambridge, England)*, 140(4), 719–729.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, 34(17), i884–i890.
- Cheng, J., Blum, R., Bowman, C., Hu, D., Shilatfard, A., Shen, S., & Dynlacht, B. D. (2014). A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Molecular cell*, 53(6), 979–992.
- Cheng, L. C., Tu, K. C., Seidel, C. W., Robb, S. M., Guo, F., & Alvarado, A. S. (2018). Cellular, ultrastructural and molecular analyses of epidermal cell development in the planarian *Schmidtea mediterranea*. *Developmental biology*, 433(2), 357–373.
- Child, C. M. (1911). Studies on the dynamics of morphogenesis and inheritance in experimental reproduction. I. The axial gradient in planaria dorotocephala as a limiting factor in regulation. *J. Exp. Zool.*, 10, 265–320.
- Choe, S. K., Vlachakis, N., & Sagerström, C. G. (2002). Meis family proteins are required for hindbrain development in the zebrafish. *Development (Cambridge, England)*, 129(3), 585–595.
- Choe, S. K., Lu, P., Nakamura, M., Lee, J., & Sagerström, C. G. (2009). Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. *Developmental cell*, 17(4), 561–567.
- Choe, S. K., Ladam, F., & Sagerström, C. G. (2014). TALE factors poise promoters for activation by Hox proteins. *Developmental cell*, 28(2), 203–211.
- Choo, S. W., White, R., & Russell, S. (2011). Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in *Drosophila*. *PLoS one*, 6(4), e14778.
- Chou, H. C., Acevedo-Luna, N., Kuhlman, J. A., & Schneider, S. Q. (2018). PdumBase: a transcriptome database and research tool for *Platynereis dumerilii* and early development of other metazoans. *BMC genomics*, 19(1), 618.

- Cecconi, F., Proetzel, G., Alvarez-Bolado, G., Jay, D., & Gruss, P. (1997). Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. *Developmental dynamics : an official publication of the American Association of Anatomists*, 210(2), 184–190.
- Cigliola, V., Shoffner, A., Lee, N., Ou, J., Gonzalez, T. J., Hoque, J., Becker, C. J., Han, Y., Shen, G., Faw, T. D., Abd-El-Barr, M. M., Varghese, S., Asokan, A., & Poss, K. D. (2023). Spinal cord repair is modulated by the neurogenic factor Hb-egf under direction of a regeneration-associated enhancer. *Nature communications*, 14(1), 4857.
- Clark, E. G., & Petersen, C. P. (2023). BMP suppresses WNT to integrate patterning of orthogonal body axes in adult planarians. *PLoS genetics*, 19(9), e1010608.
- Cloutier, J. K., McMann, C. L., Oderberg, I. M., & Reddien, P. W. (2021). activin-2 is required for regeneration of polarity on the planarian anterior-posterior axis. *PLoS genetics*, 17(3), e1009466.
- Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S. W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risco, V. I., Kundaje, A., Khavari, P. A., Montine, T. J., ... Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods*, 14(10), 959–962.
- Cote, L. E., Simental, E., & Reddien, P. W. (2019). Muscle functions as a connective tissue and source of extracellular matrix in planarians. *Nature communications*, 10(1), 1592.
- Cowles, M. W., Brown, D. D., Nisperos, S. V., Stanley, B. N., Pearson, B. J., & Zayas, R. M. (2013). Genome-wide analysis of the bHLH gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. *Development (Cambridge, England)*, 140(23), 4691–4702.
- Cowles, M. W., Omuro, K. C., Stanley, B. N., Quintanilla, C. G., & Zayas, R. M. (2014). COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians. *PLoS genetics*, 10(10), e1004746.
- Coy, S. E., & Borycki, A. G. (2010). Expression analysis of TALE family transcription factors during avian development. *Developmental dynamics : an official publication of the American Association of Anatomists*, 239(4), 1234–1245.
- Cui, L., Li, M., Feng, F., Yang, Y., Hang, X., Cui, J., & Gao, J. (2014). MEIS1 functions as a potential AR negative regulator. *Experimental cell research*, 328(1), 58–68.
- Cui, G., Zhou, J. Y., Ge, X. Y., Sun, B. F., Song, G. G., Wang, X., Wang, X. Z., Zhang, R., Wang, H. L., Jing, Q., Koziol, M. J., Zhao, Y. L., Zeng, A., Zhang, W. Q., Han, D. L., Yang, Y. G., & Yang, Y. (2023a). m⁶A promotes planarian regeneration. *Cell proliferation*, 56(5), e13481.
- Cui, G., Dong, K., Zhou, J. Y., Li, S., Wu, Y., Han, Q., Yao, B., Shen, Q., Zhao, Y. L., Yang, Y., Cai, J., Zhang, S., & Yang, Y. G. (2023b). Spatiotemporal transcriptomic atlas reveals the dynamic characteristics and key regulators of planarian regeneration. *Nature communications*, 14(1), 3205.
- Currie, K. W., & Pearson, B. J. (2013). Transcription factors *lhx1/5-1* and *pitx* are required for the maintenance and regeneration of serotonergic neurons in planarians. *Development (Cambridge, England)*, 140(17), 3577–3588.
- Currie, K. W., Brown, D. D., Zhu, S., Xu, C., Voisin, V., Bader, G. D., & Pearson, B. J. (2016). HOX gene complement and expression in the planarian *Schmidtea mediterranea*. *EvoDevo*, 7, 7.
- Currie, K. W., Molinaro, A. M., & Pearson, B. J. (2016). Neuronal sources of *hedgehog* modulate neurogenesis in the adult planarian brain. *eLife*, 5, e19735.
- Cvejic, A., Serbanovic-Canic, J., Stemple, D. L., & Ouwehand, W. H. (2011). The role of *meis1* in primitive and definitive hematopoiesis during zebrafish development. *Haematologica*, 96(2), 190–198.
- Dalyell, J. G. (1814). Observations on some interesting phenomena in animal physiology, exhibited by several species of Planariae: illustrated by coloured figures of living animals. Museum of Comparative Zoology--Biodiversity Heritage Library digitization project.
- Dambly-Chaudière, C., & Leyns, L. (1992). The determination of sense organs in *Drosophila*: a search for interacting genes. *The International journal of developmental biology*, 36(1), 85–91.

- Dard, A., Reboulet, J., Jia, Y., Bleicher, F., Duffraisse, M., Vanaker, J. M., Forcet, C., & Merabet, S. (2018). Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors. *Cell reports*, 22(11), 3058–3071.
- Dard, A., Jia, Y., Reboulet, J., Bleicher, F., Lavau, C., & Merabet, S. (2019). The human HOXA9 protein uses paralog-specific residues of the homeodomain to interact with TALE-class cofactors. *Scientific reports*, 9(1), 5664.
- Dardaei, L., Longobardi, E., & Blasi, F. (2014). Prep1 and Meis1 competition for Pbx1 binding regulates protein stability and tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10), E896–E905.
- Dardaei, L., Penkov, D., Mathiasen, L., Bora, P., Morelli, M. J., & Blasi, F. (2015). Tumorigenesis by Meis1 overexpression is accompanied by a change of DNA target-sequence specificity which allows binding to the AP-1 element. *Oncotarget*, 6(28), 25175–25187.
- Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. j., Schilstra, M. J., Clarke, P. J., Arnone, M. I., ... Bolouri, H. (2002). A genomic regulatory network for development. *Science (New York, N.Y.)*, 295(5560), 1669–1678.
- Davidson, E. H., & Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science (New York, N.Y.)*, 311(5762), 796–800.
- Davidson, E. H., & Levine, M. S. (2008). Properties of developmental gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51), 20063–20066.
- Davies, E. L., Lei, K., Seidel, C. W., Kroesen, A. E., McKinney, S. A., Guo, L., Robb, S. M., Ross, E. J., Gotting, K., & Alvarado, A. S. (2017). Embryonic origin of adult stem cells required for tissue homeostasis and regeneration. *eLife*, 6, e21052.
- Deflorian, G., Tiso, N., Ferretti, E., Meyer, D., Blasi, F., Bortolussi, M., & Argenton, F. (2004). Prep1.1 has essential genetic functions in hindbrain development and cranial neural crest cell differentiation. *Development (Cambridge, England)*, 131(3), 613–627.
- De Kumar, B., Parker, H. J., Paulson, A., Parrish, M. E., Pushel, I., Singh, N. P., Zhang, Y., Slaughter, B. D., Unruh, J. R., Florens, L., Zeitlinger, J., & Krumlauf, R. (2017). HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets. *Genome research*, 27(9), 1501–1512.
- Delgado, I., López-Delgado, A. C., Roselló-Díez, A., Giovinnazzo, G., Cadenas, V., Fernández-de-Manuel, L., Sánchez-Cabo, F., Anderson, M. J., Lewandoski, M., & Torres, M. (2020). Proximo-distal positional information encoded by an Fgf-regulated gradient of homeodomain transcription factors in the vertebrate limb. *Science advances*, 6(23), eaaz0742.
- Delgado, I., Giovinnazzo, G., Temiño, S., Gauthier, Y., Balsalobre, A., Drouin, J., & Torres, M. (2021). Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors. *Nature communications*, 12(1), 3086.
- de Mendoza, A., Sebé-Pedrós, A., Šestak, M. S., Matejic, M., Torruella, G., Domazet-Lošo, T., & Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences of the United States of America*, 110(50), E4858–E4866.
- Derelle, R., Lopez, P., Le Guyader, H., & Manuel, M. (2007). Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evolution & development*, 9(3), 212–219.
- Díaz, V. M., Mori, S., Longobardi, E., Menendez, G., Ferrai, C., Keough, R. A., Bachi, A., & Blasi, F. (2007). p160 Myb-binding protein interacts with Prep1 and inhibits its transcriptional activity. *Molecular and cellular biology*, 27(22), 7981–7990.
- Dibner, C., Elias, S., & Frank, D. (2001). XMeis3 protein activity is required for proper hindbrain patterning in *Xenopus laevis* embryos. *Development (Cambridge, England)*, 128(18), 3415–3426.
- Dibner, C., Elias, S., Ofir, R., Souopgui, J., Kolm, P. J., Sive, H., Pieler, T., & Frank, D. (2004). The Meis3 protein and retinoid signaling interact to pattern the *Xenopus* hindbrain. *Developmental biology*, 271(1), 75–86.

- Di Giacomo, G., Koss, M., Capellini, T. D., Brendolan, A., Pöpperl, H., & Selleri, L. (2006). Spatio-temporal expression of Pbx3 during mouse organogenesis. *Gene expression patterns : GEP*, 6(7), 747–757.
- DiMartino, J. F., Selleri, L., Traver, D., Firpo, M. T., Rhee, J., Warnke, R., O'Gorman, S., Weissman, I. L., & Cleary, M. L. (2001). The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver. *Blood*, 98(3), 618–626.
- Dildrop, R., & Rüther, U. (2004). Organization of Iroquois genes in fish. *Development genes and evolution*, 214(6), 267–276.
- Di Rocco, G., Mavilio, F., & Zappavigna, V. (1997). Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex. *The EMBO journal*, 16(12), 3644–3654.
- Di Rosa, P., Villaescusa, J. C., Longobardi, E., Iotti, G., Ferretti, E., Diaz, V. M., Miccio, A., Ferrari, G., & Blasi, F. (2007). The homeodomain transcription factor Prep1 (pKnox1) is required for hematopoietic stem and progenitor cell activity. *Developmental biology*, 311(2), 324–334.
- Doddihall, V., Mann, F. G., Jr, Ross, E. J., McKinney, M. C., Guerrero-Hernández, C., Brewster, C. E., McKinney, S. A., & Sánchez Alvarado, A. (2024). A PAK family kinase and the Hippo/Yorkie pathway modulate WNT signaling to functionally integrate body axes during regeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 121(20), e2321919121.
- Donaldson, I. J., Amin, S., Hensman, J. J., Kutejova, E., Rattray, M., Lawrence, N., Hayes, A., Ward, C. M., & Bobola, N. (2012). Genome-wide occupancy links Hoxa2 to Wnt- β -catenin signaling in mouse embryonic development. *Nucleic acids research*, 40(9), 3990–4001.
- Dubois F. (1949). Contribution à l'étude de la migration des cellules de régénération chez les *Planaires dulcicoles*. *Bull Biol Fr Belg* 83, 213–283.
- Duboule D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development (Cambridge, England). Supplement*, 135–142.
- DuBuc, T. Q., Stephenson, T. B., Rock, A. Q., & Martindale, M. Q. (2018). Hox and Wnt pattern the primary body axis of an anthozoan cnidarian before gastrulation. *Nature communications*, 9(1), 2007.
- Dupacova, N., Antosova, B., Paces, J., & Kozmik, Z. (2021). Meis homeobox genes control progenitor competence in the retina. *Proceedings of the National Academy of Sciences of the United States of America*, 118(12), e2013136118.
- Duncan, E. M., Chitsazan, A. D., Seidel, C. W., & Sánchez Alvarado, A. (2015). Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *Cell reports*, 13(12), 2741–2755.
- Durant, F., Lobo, D., Hammelman, J., & Levin, M. (2016). Physiological controls of large-scale patterning in planarian regeneration: a molecular and computational perspective on growth and form. *Regeneration (Oxford, England)*, 3(2), 78–102.
- Durant, F., Bischof, J., Fields, C., Morokuma, J., LaPalme, J., Hoi, A., & Levin, M. (2019). The Role of Early Bioelectric Signals in the Regeneration of Planarian Anterior/Posterior Polarity. *Biophysical journal*, 116(5), 948–961.
- Durston, A. J., Jansen, H. J., In der Rieden, P., & Hooiveld, M. H. (2011). Hox collinearity - a new perspective. *The International journal of developmental biology*, 55(10-12), 899–908.
- Ebner, A., Cabernard, C., Affolter, M., & Merabet, S. (2005). Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development (Cambridge, England)*, 132(7), 1591–1600.
- Eisenhoffer, G. T., Kang, H., & Sánchez Alvarado, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell stem cell*, 3(3), 327–339.
- Elkouby, Y. M., Elias, S., Casey, E. S., Blythe, S. A., Tsabar, N., Klein, P. S., Root, H., Liu, K. J., & Frank, D. (2010). Mesodermal Wnt signaling organizes the neural plate via Meis3. *Development (Cambridge, England)*, 137(9), 1531–1541.

- Elliott, S. A., & Sánchez Alvarado, A. (2013). The history and enduring contributions of planarians to the study of animal regeneration. *Wiley interdisciplinary reviews. Developmental biology*, 2(3), 301–326.
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20, 1-14.
- Evans, D. J., Owlarn, S., Tejada Romero, B., Chen, C., & Aboobaker, A. A. (2011). Combining classical and molecular approaches elaborates on the complexity of mechanisms underpinning anterior regeneration. *PloS one*, 6(11), e27927.
- Fan, R., Bonde, S., Gao, P., Sotomayor, B., Chen, C., Mouw, T., Zavazava, N., & Tan, K. (2012). Dynamic HoxB4-regulatory network during embryonic stem cell differentiation to hematopoietic cells. *Blood*, 119(19), e139–e147.
- Fan, Y., Chai, C., Li, P., Zou, X., Ferrell, J. E., Jr, & Wang, B. (2023). Ultrafast distant wound response is essential for whole-body regeneration. *Cell*, 186(17), 3606–3618.e16.
- Fang, X., Li, L., Luo, R. Xu. F., Wang, X., Zhu, Y., Yang, L., Huang, Z. (2012). Genomic data from the Pacific oyster (*Crassostrea gigas*). *GigaScience*.
- Feijóo, C. G., Manzanares, M., de la Calle-Mustienes, E., Gómez-Skarmeta, J. L., & Allende, M. L. (2004). The *Irx* gene family in zebrafish: genomic structure, evolution and initial characterization of *irx5b*. *Development genes and evolution*, 214(6), 277–284.
- Feng, S., Rastogi, C., Loker, R., Glassford, W. J., Tomas Rube, H., Bussemaker, H. J., & Mann, R. S. (2022). Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors. *Nature communications*, 13(1), 3808.
- Fernandez-Diaz, L. C., Laurent, A., Girasoli, S., Turco, M., Longobardi, E., Iotti, G., Jenkins, N. A., Fiorenza, M. T., Copeland, N. G., & Blasi, F. (2010). The absence of *Prep1* causes p53-dependent apoptosis of mouse pluripotent epiblast cells. *Development (Cambridge, England)*, 137(20), 3393–3403.
- Ferretti, E., Schulz, H., Talarico, D., Blasi, F., & Berthelsen, J. (1999). The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. *Mechanisms of development*, 83(1-2), 53–64.
- Ferretti, E., Marshall, H., Pöpperl, H., Maconochie, M., Krumlauf, R., & Blasi, F. (2000). Segmental expression of *Hoxb2* in r4 requires two separate sites that integrate cooperative interactions between *Prep1*, *Pbx* and *Hox* proteins. *Development (Cambridge, England)*, 127(1), 155–166.
- Ferretti, E., Cambronero, F., Tümpel, S., Longobardi, E., Wiedemann, L. M., Blasi, F., & Krumlauf, R. (2005). *Hoxb1* enhancer and control of rhombomere 4 expression: complex interplay between PREP1-PBX1-HOXB1 binding sites. *Molecular and cellular biology*, 25(19), 8541–8552.
- Ferretti, E., Villaescusa, J. C., Di Rosa, P., Fernandez-Diaz, L. C., Longobardi, E., Mazzieri, R., Miccio, A., Micali, N., Selleri, L., Ferrari, G., & Blasi, F. (2006). Hypomorphic mutation of the TALE gene *Prep1* (*pKnox1*) causes a major reduction of *Pbx* and *Meis* proteins and a pleiotropic embryonic phenotype. *Molecular and cellular biology*, 26(15), 5650–5662.
- Ferrier D. E. (2014). The Hox-TALE has been wagging for a long time. *eLife*, 3, e02515.
- Ficara, F., Murphy, M. J., Lin, M., & Cleary, M. L. (2008). *Pbx1* regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell stem cell*, 2(5), 484–496.
- Ficara, F., Crisafulli, L., Lin, C., Iwasaki, M., Smith, K. S., Zammataro, L., & Cleary, M. L. (2013). *Pbx1* restrains myeloid maturation while preserving lymphoid potential in hematopoietic progenitors. *Journal of cell science*, 126(Pt 14), 3181–3191.
- Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., & Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science (New York, N.Y.)*, 360(6391), eaaq1736.
- Fleck, J. S., Jansen, S. M. J., Wollny, D., Zenk, F., Seimiya, M., Jain, A., Okamoto, R., Santel, M., He, Z., Camp, J. G., & Treutlein, B. (2023). Inferring and perturbing cell fate regulomes in human brain organoids. *Nature*, 621(7978), 365–372.

- Fognani, C., Kilstrup-Nielsen, C., Berthelsen, J., Ferretti, E., Zappavigna, V., & Blasi, F. (2002). Characterization of PREP2, a paralog of PREP1, which defines a novel sub-family of the MEINOX TALE homeodomain transcription factors. *Nucleic acids research*, 30(9), 2043–2051.
- Foos, N., Maurel-Zaffran, C., Maté, M. J., Vincentelli, R., Hainaut, M., Berenger, H., Pradel, J., Saurin, A. J., Ortiz-Lombardía, M., & Graba, Y. (2015). A flexible extension of the *Drosophila* ultrabithorax homeodomain defines a novel Hox/PBC interaction mode. *Structure (London, England : 1993)*, 23(2), 270–279.
- Forsthoefel, D. J., Park, A. E., & Newmark, P. A. (2011). Stem cell-based growth, regeneration, and remodeling of the planarian intestine. *Developmental biology*, 356(2), 445–459.
- Forsthoefel, D. J., Ross, K. G., Newmark, P. A., & Zayas, R. M. (2018). Fixation, Processing, and Immunofluorescent Labeling of Whole Mount Planarians. *Methods in molecular biology (Clifton, N.J.)*, 1774, 353–366.
- Forsthoefel, D. J., Cejda, N. I., Khan, U. W., & Newmark, P. A. (2020). Cell-type diversity and regionalized gene expression in the planarian intestine. *eLife*, 9, e52613.
- French V. (1976). Leg regeneration in the cockroach, *Blattella germanica*. II. Regeneration from a non-congruent tibial graft/host junction. *Journal of embryology and experimental morphology*, 35(2), 267–301.
- Fujimura, N., Vacik, T., Machon, O., Vlcek, C., Scalabrin, S., Speth, M., Diep, D., Krauss, S., & Kozmik, Z. (2007). Wnt-mediated down-regulation of Sp1 target genes by a transcriptional repressor Sp5. *The Journal of biological chemistry*, 282(2), 1225–1237.
- Gahan, J. M., Bradshaw, B., Flici, H., & Frank, U. (2016). The interstitial stem cells in *Hydractinia* and their role in regeneration. *Current opinion in genetics & development*, 40, 65–73.
- Galant, R., Walsh, C. M., & Carroll, S. B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development (Cambridge, England)*, 129(13), 3115–3126.
- García-Castro, H., Kenny, N. J., Iglesias, M., Álvarez-Campos, P., Mason, V., Elek, A., ... & Solana, J. (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome biology*, 22, 1–34.
- Garcia-Fernández, J., Baguña, J., & Saló, E. (1991). Planarian homeobox genes: cloning, sequence analysis, and expression. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), 7338–7342.
- Gaviño, M. A., & Reddien, P. W. (2011). A Bmp/Admp regulatory circuit controls maintenance and regeneration of dorsal-ventral polarity in planarians. *Current biology : CB*, 21(4), 294–299.
- Gaviño, M. A., Wenemoser, D., Wang, I. E., & Reddien, P. W. (2013). Tissue absence initiates regeneration through follistatin-mediated inhibition of activin signaling. *eLife*, 2, e00247.
- Gehrke, A. R., & Srivastava, M. (2016). Neoblasts and the evolution of whole-body regeneration. *Current opinion in genetics & development*, 40, 131–137.
- Gehrke, A. R., Neverett, E., Luo, Y. J., Brandt, A., Ricci, L., Hulett, R. E., Gompers, A., Ruby, J. G., Rokhsar, D. S., Reddien, P. W., & Srivastava, M. (2019). Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science (New York, N.Y.)*, 363(6432), eaau6173.
- Gierer, A., & Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik*, 12(1), 30–39.
- Giresi, P. G., Kim, J., McDaniel, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*, 17(6), 877–885.
- Glazov, E. A., Pheasant, M., McGraw, E. A., Bejerano, G., & Mattick, J. S. (2005). Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing. *Genome research*, 15(6), 800–808.
- Goldman, J. A., Kuzu, G., Lee, N., Karasik, J., Gemberling, M., Foglia, M. J., Karra, R., Dickson, A. L., Sun, F., Tolstorukov, M. Y., & Poss, K. D. (2017). Resolving Heart Regeneration by Replacement Histone Profiling. *Developmental cell*, 40(4), 392–404.e5.
- Goldman, J. A., & Poss, K. D. (2020). Gene regulatory programmes of tissue regeneration. *Nature reviews. Genetics*, 21(9), 511–525.

- González-Crespo, S., Abu-Shaar, M., Torres, M., Martínez-A, C., Mann, R. S., & Morata, G. (1998). Antagonism between extradenticle function and Hedgehog signalling in the developing limb. *Nature*, 394(6689), 196–200.
- Grebbin, B. M., Hau, A. C., Groß, A., Anders-Maurer, M., Schramm, J., Koss, M., Wille, C., Mittelbronn, M., Selleri, L., & Schulte, D. (2016). Pbx1 is required for adult subventricular zone neurogenesis. *Development (Cambridge, England)*, 143(13), 2281–2291.
- Grebbin, B. M., & Schulte, D. (2017). PBX1 as Pioneer Factor: A Case Still Open. *Frontiers in cell and developmental biology*, 5, 9.
- Green, J. B., & Sharpe, J. (2015). Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development (Cambridge, England)*, 142(7), 1203–1211.
- Gutkovich, Y. E., Ofir, R., Elkouby, Y. M., Dibner, C., Gefen, A., Elias, S., & Frank, D. (2010). Xenopus Meis3 protein lies at a nexus downstream to Zic1 and Pax3 proteins, regulating multiple cell-fates during early nervous system development. *Developmental biology*, 338(1), 50–62.
- Gordon, J. A., Hassan, M. Q., Saini, S., Montecino, M., van Wijnen, A. J., Stein, G. S., Stein, J. L., & Lian, J. B. (2010). Pbx1 represses osteoblastogenesis by blocking Hoxa10-mediated recruitment of chromatin remodeling factors. *Molecular and cellular biology*, 30(14), 3531–3541.
- Gordon, J. A., Hassan, M. Q., Koss, M., Montecino, M., Selleri, L., van Wijnen, A. J., Stein, J. L., Stein, G. S., & Lian, J. B. (2011). Epigenetic regulation of early osteogenesis and mineralized tissue formation by a HOXA10-PBX1-associated complex. *Cells, tissues, organs*, 194(2-4), 146–150.
- Goss, R. J. (1969). *Principles of regeneration*. Elsevier.
- Goudet, G., Delhalle, S., Biemar, F., Martial, J. A., & Peers, B. (1999). Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter. *The Journal of biological chemistry*, 274(7), 4067–4073.
- Grieder, N. C., Marty, T., Ryoo, H. D., Mann, R. S., & Affolter, M. (1997). Synergistic activation of a Drosophila enhancer by HOM/EXD and DPP signaling. *The EMBO journal*, 16(24), 7402–7410.
- Grohme, M. A., Schloissnig, S., Rozanski, A., Pippel, M., Young, G. R., Winkler, S., Brandl, H., Henry, I., Dahl, A., Powell, S., Hiller, M., Myers, E., & Rink, J. C. (2018). The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms. *Nature*, 554(7690), 56–61.
- Guo, T., Peters, A. H., & Newmark, P. A. (2006). A Bruno-like gene is required for stem cell maintenance in planarians. *Developmental cell*, 11(2), 159–169.
- Gurley, K. A., Rink, J. C., & Sánchez Alvarado, A. (2008). Beta-catenin defines head versus tail identity during planarian regeneration and homeostasis. *Science (New York, N.Y.)*, 319(5861), 323–327.
- Gurley, K. A., Elliott, S. A., Simakov, O., Schmidt, H. A., Holstein, T. W., & Sánchez Alvarado, A. (2010). Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Developmental biology*, 347(1), 24–39.
- Hahne, F. & Ivanek, R. (2016). “Statistical Genomics: Methods and Protocols.” In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335–351. Springer New York, New York, NY.
- Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (New York, N.Y.)*, 267(5205), 1788–1792.
- Haller, K., Rambaldi, I., Kovács, E. N., Daniels, E., & Featherstone, M. (2002). Prep2: cloning and expression of a new prep family member. *Developmental dynamics : an official publication of the American Association of Anatomists*, 225(3), 358–364.
- Haller, K., Rambaldi, I., Daniels, E., & Featherstone, M. (2004). Subcellular localization of multiple PREP2 isoforms is regulated by actin, tubulin, and nuclear export. *The Journal of biological chemistry*, 279(47), 49384–49394.
- Hanna, J. H., Saha, K., & Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell*, 143(4), 508–525.

- Harris, R. E., Setiawan, L., Saul, J., & Hariharan, I. K. (2016). Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature *Drosophila* imaginal discs. *eLife*, 5, e11588.
- Harris M. P. (2021). Bioelectric signaling as a unique regulator of development and regeneration. *Development (Cambridge, England)*, 148(10), dev180794.
- Hau, A. C., Grebbin, B. M., Agoston, Z., Anders-Maurer, M., Müller, T., Groß, A., Kolb, J., Langer, J. D., Döring, C., & Schulte, D. (2017). MEIS homeodomain proteins facilitate PARP1/ARTD1-mediated eviction of histone H1. *The Journal of cell biology*, 216(9), 2715–2729.
- Hayashi, T., Asami, M., Higuchi, S., Shibata, N., & Agata, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development, growth & differentiation*, 48(6), 371–380.
- Heine, P., Dohle, E., Bumsted-O'Brien, K., Engelkamp, D., & Schulte, D. (2008). Evidence for an evolutionary conserved role of homothorax/Meis1/2 during vertebrate retina development. *Development (Cambridge, England)*, 135(5), 805–811.
- Henry J. Q. (2014). Spiralian model systems. *The International journal of developmental biology*, 58(6-8), 389–401.
- Hill, E. M., & Petersen, C. P. (2018). Positional information specifies the site of organ regeneration and not tissue maintenance in planarians. *eLife*, 7, e33680.
- Hisa, T., Spence, S. E., Rachel, R. A., Fujita, M., Nakamura, T., Ward, J. M., Devor-Henneman, D. E., Saiki, Y., Kutsuna, H., Tessarollo, L., Jenkins, N. A., & Copeland, N. G. (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *The EMBO journal*, 23(2), 450–459.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular biology and evolution*, 35(2), 518–522.
- Holland, P. W., Booth, H. A. F., & Bruford, E. A. (2007). Classification and nomenclature of all human homeobox genes. *BMC biology*, 5, 1-28.
- Holland, P. W., Marlétaz, F., Maeso, I., Dunwell, T. L., & Paps, J. (2017). New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372(1713), 20150480.
- Holstein T. W. (2022). The role of cnidarian developmental biology in unraveling axis formation and Wnt signaling. *Developmental biology*, 487, 74–98.
- Holzem, M., Boutros, M., & Holstein, T. W. (2024). The origin and evolution of Wnt signalling. *Nature Reviews Genetics*, 1-13.
- Hu, Y., Yu, H., Shaw, G., Renfree, M. B., & Pask, A. J. (2011). Differential roles of TGIF family genes in mammalian reproduction. *BMC developmental biology*, 11, 58.
- Huang, H., Paliouras, M., Rambaldi, I., Lasko, P., & Featherstone, M. (2003). Nonmuscle myosin promotes cytoplasmic localization of PBX. *Molecular and cellular biology*, 23(10), 3636–3645.
- Huang, H., Rastegar, M., Bodner, C., Goh, S. L., Rambaldi, I., & Featherstone, M. (2005). MEIS C termini harbor transcriptional activation domains that respond to cell signaling. *The Journal of biological chemistry*, 280(11), 10119–10127.
- Huang, Y., Sitwala, K., Bronstein, J., Sanders, D., Dandekar, M., Collins, C., Robertson, G., MacDonald, J., Cezard, T., Bilenky, M., Thiessen, N., Zhao, Y., Zeng, T., Hirst, M., Hero, A., Jones, S., & Hess, J. L. (2012). Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood*, 119(2), 388–398.
- Hubert, A., Henderson, J. M., Ross, K. G., Cowles, M. W., Torres, J., & Zayas, R. M. (2013). Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone methyltransferases. *Epigenetics*, 8(1), 79-91.

- Hudry, B., Remacle, S., Delfini, M. C., Rezsohazy, R., Graba, Y., & Merabet, S. (2012). Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. *PLoS biology*, 10(6), e1001351.
- Hudry, B., Thomas-Chollier, M., Volovik, Y., Duffraisse, M., Dard, A., Frank, D., Technau, U., & Merabet, S. (2014). Molecular insights into the origin of the Hox-TALE patterning system. *eLife*, 3, e01939.
- Hurtado, R., Zewdu, R., Mtui, J., Liang, C., Aho, R., Kurylo, C., Selleri, L., & Herzlinger, D. (2015). Pbx1-dependent control of VMC differentiation kinetics underlies gross renal vascular patterning. *Development (Cambridge, England)*, 142(15), 2653–2664.
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS one*, 5(9), e12776.
- Hyman, C. A., Bartholin, L., Newfeld, S. J., & Wotton, D. (2003). Drosophila TGIF proteins are transcriptional activators. *Molecular and cellular biology*, 23(24), 9262–9274.
- Iglesias, M., Gomez-Skarmeta, J. L., Saló, E., & Adell, T. (2008). Silencing of Smed-betacatenin1 generates radial-like hypercephalized planarians. *Development (Cambridge, England)*, 135(7), 1215–1221.
- Iglesias, M., Almuedo-Castillo, M., Aboobaker, A. A., & Saló, E. (2011). Early planarian brain regeneration is independent of blastema polarity mediated by the Wnt/ β -catenin pathway. *Developmental biology*, 358(1), 68–78.
- Iotti, G., Longobardi, E., Masella, S., Dardaei, L., De Santis, F., Micali, N., & Blasi, F. (2011). Homeodomain transcription factor and tumor suppressor Prep1 is required to maintain genomic stability. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29), E314–E322.
- Iotti, G., Mejetta, S., Modica, L., Penkov, D., Ponzoni, M., & Blasi, F. (2012). Reduction of Prep1 levels affects differentiation of normal and malignant B cells and accelerates Myc driven lymphomagenesis. *PLoS one*, 7(10), e48353.
- Irie, N., & Kuratani, S. (2014). The developmental hourglass model: a predictor of the basic body plan?. *Development (Cambridge, England)*, 141(24), 4649–4655.
- Ivanković, M., Brand, J. N., Pandolfini, L., Brown, T., Pippel, M., Rozanski, A., Schubert, T., Grohme, M. A., Winkler, S., Robledillo, L., Zhang, M., Codino, A., Gustincich, S., Vila-Farré, M., Zhang, S., Papantonis, A., Marques, A., & Rink, J. C. (2024). A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *Nature communications*, 15(1), 8215.
- Jaber-Hijazi, F., Lo, P. J., Mihaylova, Y., Foster, J. M., Benner, J. S., Romero, B. T., ... & Aboobaker, A. A. (2013). Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Developmental biology*, 384(1), 141–153.
- Jacobs, Y., Schnabel, C. A., & Cleary, M. L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Molecular and cellular biology*, 19(7), 5134–5142.
- Jagannathan-Bogdan, M., & Zon, L. I. (2013). Hematopoiesis. *Development (Cambridge, England)*, 140(12), 2463–2467.
- Jaw, T. J., You, L. R., Knoepfler, P. S., Yao, L. C., Pai, C. Y., Tang, C. Y., Chang, L. P., Berthelsen, J., Blasi, F., Kamps, M. P., & Sun, Y. H. (2000). Direct interaction of two homeoproteins, homothorax and extradenticle, is essential for EXD nuclear localization and function. *Mechanisms of development*, 91(1-2), 279–291.
- Johnson, F. B., Parker, E., & Krasnow, M. A. (1995). Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 92(3), 739–743.
- Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, 316(5830), 1497–1502.
- Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., & Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, 527(7578), 384–388.

- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9), 1236–1240.
- Joo, S., Wang, M. H., Lui, G., Lee, J., Barnas, A., Kim, E., Sudek, S., Worden, A. Z., & Lee, J. H. (2018). Common ancestry of heterodimerizing TALE homeobox transcription factors across Metazoa and Archaeplastida. *BMC biology*, 16(1), 136.
- Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., & Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, 131(3), 530–543.
- Juliano, C. E., Swartz, S. Z., & Wessel, G. M. (2010). A conserved germline multipotency program. *Development (Cambridge, England)*, 137(24), 4113–4126.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: II. Zygotic loci on the third chromosome. *Wilhelm Roux's archives of developmental biology*, 193(5), 283–295.
- Kakebeen, A. D., Chitsazan, A. D., Williams, M. C., Saunders, L. M., & Wills, A. E. (2020). Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of regeneration in neural progenitors. *eLife*, 9, e52648.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6), 587–589.
- Kamal, A., Arnold, C., Claringbould, A., Moussa, R., Servaas, N. H., Kholmatov, M., Daga, N., Nogina, D., Mueller-Dott, S., Reyes-Palomares, A., Palla, G., Sigalova, O., Bunina, D., Pabst, C., & Zaugg, J. B. (2023). GRaNIE and GRaNPA: inference and evaluation of enhancer-mediated gene regulatory networks. *Molecular systems biology*, 19(6), e11627.
- Kamps, M. P., Murre, C., Sun, X. H., & Baltimore, D. (1990). A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell*, 60(4), 547–555.
- Kang, J., Hu, J., Karra, R., Dickson, A. L., Tornini, V. A., Nachtrab, G., Gemberling, M., Goldman, J. A., Black, B. L., & Poss, K. D. (2016). Modulation of tissue repair by regeneration enhancer elements. *Nature*, 532(7598), 201–206.
- Kanzleiter, T., Rath, M., Penkov, D., Puchkov, D., Schulz, N., Blasi, F., & Schürmann, A. (2014). Pknox1/Prep1 regulates mitochondrial oxidative phosphorylation components in skeletal muscle. *Molecular and cellular biology*, 34(2), 290–298.
- Kao, D., Felix, D., & Aboobaker, A. (2013). The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. *BMC genomics*, 14(1), 797.
- Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature reviews. Molecular cell biology*, 9(10), 770–780.
- Kato, K., Orii, H., Watanabe, K., & Agata, K. (2001). Dorsal and ventral positional cues required for the onset of planarian regeneration may reside in differentiated cells. *Developmental Biology*, 233(1), 109–121.
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160–1166.
- Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature communications*, 10(1), 1930.
- Kerner, P., Ikmi, A., Coen, D., & Vervoort, M. (2009). Evolutionary history of the iroquois/Irx genes in metazoans. *BMC evolutionary biology*, 9, 74.

- Kicheva, A., & Briscoe, J. (2023). Control of Tissue Development by Morphogens. *Annual review of cell and developmental biology*, 39, 91–121.
- Kilstrup-Nielsen, C., Alessio, M., & Zappavigna, V. (2003). PBX1 nuclear export is regulated independently of PBX-MEINOX interaction by PKA phosphorylation of the PBC-B domain. *The EMBO journal*, 22(1), 89–99.
- Kim, S. K., Selleri, L., Lee, J. S., Zhang, A. Y., Gu, X., Jacobs, Y., & Cleary, M. L. (2002). Pbx1 inactivation disrupts pancreas development and in *Ipfl1*-deficient mice promotes diabetes mellitus. *Nature genetics*, 30(4), 430–435.
- Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular cell*, 83(3), 373–392.
- King, R. S., & Newmark, P. A. (2013). In situ hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea*. *BMC developmental biology*, 13, 8.
- King, H. O., Owusu-Boaitey, K. E., Fincher, C. T., & Reddien, P. W. (2024). A transcription factor atlas of stem cell fate in planarians. *Cell reports*, 43(3), 113843.
- Kitisin, K., Saha, T., Blake, T., Golestaneh, N., Deng, M., Kim, C., Tang, Y., Shetty, K., Mishra, B., & Mishra, L. (2007). Tgf-Beta signaling in development. *Science's STKE : signal transduction knowledge environment*, 2007(399), cm1.
- Knoepfler, P. S., & Kamps, M. P. (1995). The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Molecular and cellular biology*, 15(10), 5811–5819.
- Knoepfler, P. S., Lu, Q., & Kamps, M. P. (1996). Pbx-1 Hox heterodimers bind DNA on inseparable half-sites that permit intrinsic DNA binding specificity of the Hox partner at nucleotides 3' to a TAAT motif. *Nucleic acids research*, 24(12), 2288–2294.
- Knoepfler, P. S., Calvo, K. R., Chen, H., Antonarakis, S. E., & Kamps, M. P. (1997). Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. *Proceedings of the National Academy of Sciences of the United States of America*, 94(26), 14553–14558.
- Knoepfler, P. S., & Kamps, M. P. (1997). The highest affinity DNA element bound by Pbx complexes in t(1;19) leukemic cells fails to mediate cooperative DNA-binding or cooperative transactivation by E2a-Pbx1 and class I Hox proteins - evidence for selective targeting of E2a-Pbx1 to a subset of Pbx-recognition elements. *Oncogene*, 14(21), 2521–2531.
- Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-Korytko, I., Sun, Y. H., Wright, W. E., Tapscott, S. J., & Kamps, M. P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. *Nucleic acids research*, 27(18), 3752–3761.
- Kobayashi, C., Watanabe, K., & Agata, K. (1999). The process of pharynx regeneration in planarians. *Developmental biology*, 211(1), 27–38.
- Kobayashi, M., Fujioka, M., Tolkunova, E. N., Deka, D., Abu-Shaar, M., Mann, R. S., & Jaynes, J. B. (2003). Engrailed cooperates with extradenticle and homothorax to repress target genes in *Drosophila*. *Development (Cambridge, England)*, 130(4), 741–751.
- Kocabas, F., Zheng, J., Thet, S., Copeland, N. G., Jenkins, N. A., DeBerardinis, R. J., Zhang, C., & Sadek, H. A. (2012). Meis1 regulates the metabolic phenotype and oxidant defense of hematopoietic stem cells. *Blood*, 120(25), 4963–4972.
- Koinuma, S., Umesono, Y., Watanabe, K., & Agata, K. (2003). The expression of planarian brain factor homologs, DjFoxG and DjFoxD. *Gene expression patterns : GEP*, 3(1), 21–27.
- Kojima T. (2004). The mechanism of *Drosophila* leg development along the proximodistal axis. *Development, growth & differentiation*, 46(2), 115–129.

- Kolb, J., Anders-Maurer, M., Müller, T., Hau, A. C., Grebbin, B. M., Kallenborn-Gerhardt, W., Behrends, C., & Schulte, D. (2018). Arginine Methylation Regulates MEIS2 Nuclear Localization to Promote Neuronal Differentiation of Adult SVZ Progenitors. *Stem cell reports*, 10(4), 1184–1192.
- Krumlauf, R., & Wilkinson, D. G. (2021). Segmentation and patterning of the vertebrate hindbrain. *Development (Cambridge, England)*, 148(15), dev186460.
- Kurant, E., Pai, C. Y., Sharf, R., Halachmi, N., Sun, Y. H., & Salzberg, A. (1998). Dorsotonal/homothorax, the *Drosophila* homologue of *meis1*, interacts with extradenticle in patterning of the embryonic PNS. *Development (Cambridge, England)*, 125(6), 1037–1048.
- Kwiatkowski, D., Blaxter, M., Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, & Darwin Tree of Life Consortium (2021). The genome sequence of the bootlace worm, *Lineus longissimus* (Gunnerus, 1770). *Wellcome open research*, 6, 272.
- Labbé, R. M., Irimia, M., Currie, K. W., Lin, A., Zhu, S. J., Brown, D. D., Ross, E. J., Voisin, V., Bader, G. D., Blencowe, B. J., & Pearson, B. J. (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem cells (Dayton, Ohio)*, 30(8), 1734–1745.
- Ladam, F., Stanney, W., Donaldson, I. J., Yildiz, O., Bobola, N., & Sagerström, C. G. (2018). TALE factors use two distinct functional modes to control an essential zebrafish gene expression program. *eLife*, 7, e36144.
- Lai, A. G., & Aboobaker, A. A. (2018). EvoRegen in animals: Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes. *Developmental biology*, 433(2), 118–131.
- Lambert J. D. (2010). Developmental patterns in spiralian embryos. *Current biology : CB*, 20(2), R72–R77.
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, 172(4), 650–665.
- Lander, R., & Petersen, C. P. (2016). Wnt, Ptk7, and FGFR1 expression gradients control trunk positional identity in planarian regeneration. *eLife*, 5, e12850.
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9, 1-13.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357–359.
- Lapan, S. W., & Reddien, P. W. (2011). *dlx* and *sp6-9* Control optic cup regeneration in a prototypic eye. *PLoS genetics*, 7(8), e1002226.
- Lapan, S. W., & Reddien, P. W. (2012). Transcriptome analysis of the planarian eye identifies *ovo* as a specific regulator of eye regeneration. *Cell reports*, 2(2), 294–307.
- LaRonde-LeBlanc, N. A., & Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes & development*, 17(16), 2060–2072.
- Laumer, C. E., Bekkouché, N., Kerbl, A., Goetz, F., Neves, R. C., Sørensen, M. V., Kristensen, R. M., Hejnol, A., Dunn, C. W., Giribet, G., & Worsaae, K. (2015). Spiralian phylogeny informs the evolution of microscopic lineages. *Current biology : CB*, 25(15), 2000–2006.
- Laurent, A., Bihan, R., Deschamps, S., Guerrier, D., Dupé, V., Omilli, F., Burel, A., & Pellerin, I. (2007). Identification of a new type of PBX1 partner that contains zinc finger motifs and inhibits the binding of HOXA9-PBX1 to DNA. *Mechanisms of development*, 124(5), 364–376.
- Laurent, A., Masse, J., Deschamps, S., Burel, A., Omilli, F., Richard-Parpaillon, L., & Pellerin, I. (2009). Interaction of ZFPIP with PBX1 is crucial for proper expression of neural genetic markers during *Xenopus* development. *Development, growth & differentiation*, 51(8), 699–706.
- Laurent, A., Masse, J., Omilli, F., Deschamps, S., Richard-Parpaillon, L., Chartrain, I., & Pellerin, I. (2009). ZFPIP/Zfp462 is maternally required for proper early *Xenopus laevis* development. *Developmental biology*, 327(1), 169–176.

- Laurent, A., Calabrese, M., Warnatz, H. J., Yaspo, M. L., Tkachuk, V., Torres, M., Blasi, F., & Penkov, D. (2015). ChIP-Seq and RNA-Seq analyses identify components of the Wnt and Fgf signaling pathways as Prep1 target genes in mouse embryonic stem cells. *PLoS one*, 10(4), e0122518.
- Lavoie, H., Gagnon, J., & Therrien, M. (2020). ERK signalling: a master regulator of cell behaviour, life and fate. *Nature reviews. Molecular cell biology*, 21(10), 607–632.
- Lawrence, H. J., Rozenfeld, S., Cruz, C., Matsukuma, K., Kwong, A., Kömüves, L., Buchberg, A. M., & Largman, C. (1999). Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. *Leukemia*, 13(12), 1993–1999.
- LeBrun, D. P., & Cleary, M. L. (1994). Fusion with E2A alters the transcriptional properties of the homeodomain protein PBX1 in t(1;19) leukemias. *Oncogene*, 9(6), 1641–1647.
- Lee, R. T., Zhao, Z., & Ingham, P. W. (2016). Hedgehog signalling. *Development (Cambridge, England)*, 143(3), 367–372.
- Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic acids research*, 52(W1), W78–W82.
- Leyns, L., Gómez-Skarmeta, J. L., & Dambly-Chaudière, C. (1996). iroquois: a prepattern gene that controls the formation of bristles on the thorax of *Drosophila*. *Mechanisms of development*, 59(1), 63–72.
- Lewallen, M., & Burggren, W. (2020). Metabolic physiology of the freshwater Planaria *Girardia dorotocephala* and *Schmidtea mediterranea*: Reproductive mode, specific dynamic action, and temperature. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 319(4), R428–R438.
- Li, T., Stark, M. R., Johnson, A. D., & Wolberger, C. (1995). Crystal structure of the MATA1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science (New York, N.Y.)*, 270(5234), 262–269.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079.
- Li, D., Taylor, D. H., & van Wolfswinkel, J. C. (2021). PIWI-mediated control of tissue-specific transposons is essential for somatic cell differentiation. *Cell reports*, 37(1), 109776.
- Liberali, P., & Schier, A. F. (2024). The evolution of developmental biology through conceptual and technological revolutions. *Cell*, 187(14), 3461–3495.
- Lin, H., & Spradling, A. C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development (Cambridge, England)*, 124(12), 2463–2476.
- Lin, G., & Slack, J. M. (2008). Requirement for Wnt and FGF signaling in *Xenopus* tadpole tail regeneration. *Developmental biology*, 316(2), 323–335.
- Lin, A. Y., & Pearson, B. J. (2014). Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. *Development (Cambridge, England)*, 141(6), 1197–1208.
- Lin, A. Y. T., & Pearson, B. J. (2017). Yorkie is required to restrict the injury responses in planarians. *PLoS genetics*, 13(7), e1006874.
- Liu, Y., MacDonald, R. J., & Swift, G. H. (2001). DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex. *The Journal of biological chemistry*, 276(21), 17985–17993.
- Liu, S. Y., Selck, C., Friedrich, B., Lutz, R., Vila-Farré, M., Dahl, A., Brandl, H., Lakshmanaperumal, N., Henry, I., & Rink, J. C. (2013). Reactivating head regrowth in a regeneration-deficient planarian species. *Nature*, 500(7460), 81–84.

- Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., & Ren, J. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics (Oxford, England)*, 31(20), 3359–3361.
- Liu, Y., Ao, X., Zhou, X., Du, C., & Kuang, S. (2022). The regulation of PBXs and their emerging role in cancer. *Journal of cellular and molecular medicine*, 26(5), 1363–1379.
- LoCascio, S. A., Lapan, S. W., & Reddien, P. W. (2017). Eye Absence Does Not Regulate Planarian Stem Cells during Eye Regeneration. *Developmental cell*, 40(4), 381–391.e3.
- Loh, K. M., van Amerongen, R., & Nusse, R. (2016). Generating cellular diversity and spatial form: Wnt signaling and the evolution of multicellular animals. *Developmental cell*, 38(6), 643–655.
- Longobardi, E., Iotti, G., Di Rosa, P., Mejetta, S., Bianchi, F., Fernandez-Diaz, L. C., Micali, N., Nuciforo, P., Lenti, E., Ponzoni, M., Doglioni, C., Caniatti, M., Di Fiore, P. P., & Blasi, F. (2010). Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1. *Molecular oncology*, 4(2), 126–134.
- Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., & Blasi, F. (2014). Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Developmental dynamics : an official publication of the American Association of Anatomists*, 243(1), 59–75.
- Lopes, C. S., & Casares, F. (2010). hth maintains the pool of eye progenitors and its downregulation by Dpp and Hh couples retinal fate acquisition with cell cycle exit. *Developmental biology*, 339(1), 78–88.
- Lopes, C. S., & Casares, F. (2015). Eye selector logic for a coordinated cell cycle exit. *PLoS genetics*, 11(2), e1004981.
- López-Delgado, A. C., Delgado, I., Cadenas, V., Sánchez-Cabo, F., & Torres, M. (2021). Axial skeleton anterior-posterior patterning is regulated through feedback regulation between Meis transcription factors and retinoic acid. *Development (Cambridge, England)*, 148(1), dev193813.
- Losa, M., Barozzi, I., Osterwalder, M., Hermosilla-Aguayo, V., Morabito, A., Chacón, B. H., Zarrineh, P., Girdziusaitė, A., Benazet, J. D., Zhu, J., Mackem, S., Capellini, T. D., Dickel, D., Bobola, N., Zuniga, A., Visel, A., Zeller, R., & Selleri, L. (2023). A spatio-temporally constrained gene regulatory network directed by PBX1/2 acquires limb patterning specificity via HAND2. *Nature communications*, 14(1), 3993.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15, 1–21.
- Lu, Q., Knoepfler, P. S., Scheele, J., Wright, D. D., & Kamps, M. P. (1995). Both Pbx1 and E2A-Pbx1 bind the DNA motif ATCAATCAA cooperatively with the products of multiple murine Hox genes, some of which are themselves oncogenes. *Molecular and cellular biology*, 15(7), 3786–3795.
- Lu, Q., & Kamps, M. P. (1996). Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex. *Molecular and cellular biology*, 16(4), 1632–1640.
- Lu, Q., & Kamps, M. P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specificities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. *Oncogene*, 14(1), 75–83.
- Luo, Y. J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M., Yamasaki, S. I., Endo, K., & Satoh, N. (2015). The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature communications*, 6, 8301.
- Luo, Y. J., Kanda, M., Koyanagi, R., Hisata, K., Akiyama, T., Sakamoto, H., Sakamoto, T., & Satoh, N. (2018). Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nature ecology & evolution*, 2(1), 141–151.
- Lust, K., Maynard, A., Gomes, T., Fleck, J. S., Camp, J. G., Tanaka, E. M., & Treutlein, B. (2022). Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. *Science (New York, N.Y.)*, 377(6610), eabp9262.

- Ma, S., Zhang, B., LaFave, L. M., Earl, A. S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V. K., Tay, T., Law, T., Lareau, C., Hsu, Y. C., Regev, A., & Buenrostro, J. D. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*, 183(4), 1103–1116.e20.
- Machon, O., Masek, J., Machonova, O., Krauss, S., & Kozmik, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. *BMC developmental biology*, 15, 40.
- Maconochie, M. K., Nonchev, S., Studer, M., Chan, S. K., Pöpperl, H., Sham, M. H., Mann, R. S., & Krumlauf, R. (1997). Cross-regulation in the mouse HoxB complex: the expression of Hoxb2 in rhombomere 4 is regulated by Hoxb1. *Genes & development*, 11(14), 1885–1895.
- Maeda, R., Mood, K., Jones, T. L., Aruga, J., Buchberg, A. M., & Daar, I. O. (2001). Xmeis1, a protooncogene involved in specifying neural crest cell fate in *Xenopus* embryos. *Oncogene*, 20(11), 1329–1342.
- Maeda, R., Ishimura, A., Mood, K., Park, E. K., Buchberg, A. M., & Daar, I. O. (2002). Xpbx1b and Xmeis1b play a collaborative role in hindbrain and neural crest gene expression in *Xenopus* embryos. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5448–5453.
- Magnani, L., Ballantyne, E. B., Zhang, X., & Lupien, M. (2011). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS genetics*, 7(11), e1002368.
- Mahmoud, A. I., Kocabas, F., Muralidhar, S. A., Kimura, W., Koura, A. S., Thet, S., Porrello, E. R., & Sadek, H. A. (2013). Meis1 regulates postnatal cardiomyocyte cell cycle arrest. *Nature*, 497(7448), 249–253.
- Mann, R. S., & Abu-Shaar, M. (1996). Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. *Nature*, 383(6601), 630–633.
- Mann, R. S., & Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends in genetics : TIG*, 12(7), 258–262.
- Manning, S. A., Kroeger, B., & Harvey, K. F. (2020). The regulation of Yorkie, YAP and TAZ: new insights into the Hippo pathway. *Development (Cambridge, England)*, 147(8), dev179069.
- Marcos, S., González-Lázaro, M., Beccari, L., Carramolino, L., Martín-Bermejo, M. J., Amarie, O., Mateos-San Martín, D., Torroja, C., Bogdanović, O., Doohan, R., Puk, O., Hrabě de Angelis, M., Graw, J., Gomez-Skarmeta, J. L., Casares, F., Torres, M., & Bovolenta, P. (2015). Meis1 coordinates a network of genes implicated in eye development and microphthalmia. *Development (Cambridge, England)*, 142(17), 3009–3020.
- Mariani, L., Guo, X., Menezes, N. A., Drozd, A. M., Çakal, S. D., Wang, Q., & Ferretti, E. (2021). A TALE/HOX code unlocks WNT signalling response towards paraxial mesoderm. *Nature communications*, 12(1), 5136.
- Marlétaz, F., Peijnenburg, K. T. C. A., Goto, T., Satoh, N., & Rokhsar, D. S. (2019). A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Current biology : CB*, 29(2), 312–318.e3.
- Martín-Durán, J. M., Monjo, F., & Romero, R. (2012). Planarian embryology in the era of comparative developmental biology. *The International journal of developmental biology*, 56(1-3), 39–48.
- Martín-Durán, J. M., & Marlétaz, F. (2020). Unravelling spiral cleavage. *Development (Cambridge, England)*, 147(1), dev181081.
- Martín-Zamora, F. M., Davies, B. E., Donnellan, R. D., Guynes, K., & Martín-Durán, J. M. (2023). Functional genomics in Spiralia. *Briefings in functional genomics*, 22(6), 487–497.
- März, M., Seebeck, F., & Bartscherer, K. (2013). A Pitx transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. *Development (Cambridge, England)*, 140(22), 4499–4509.
- Mathiasen, L., Valentini, E., Boivin, S., Cattaneo, A., Blasi, F., Svergun, D. I., & Bruckmann, C. (2016). The flexibility of a homeodomain transcription factor heterodimer and its allosteric regulation by DNA binding. *The FEBS journal*, 283(16), 3134–3154.
- Mauer, K., Hellmann, S. L., Groth, M., Fröbisch, A. C., Zischler, H., Hankeln, T., & Herlyn, H. (2020). The genome, transcriptome, and proteome of the fish parasite *Pomphorhynchus laevis* (Acanthocephala). *PloS one*, 15(6), e0232973.

- Maves, L., Waskiewicz, A. J., Paul, B., Cao, Y., Tyler, A., Moens, C. B., & Tapscott, S. J. (2007). Pbx homeodomain proteins direct MyoD activity to promote fast-muscle differentiation. *Development (Cambridge, England)*, 134(18), 3371–3382.
- McNeill, H., Yang, C. H., Brodsky, M., Ungos, J., & Simon, M. A. (1997). mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the *Drosophila* eye. *Genes & development*, 11(8), 1073–1082.
- Merabet, S., & Galliot, B. (2015). The TALE face of Hox proteins in animal evolution. *Frontiers in genetics*, 6, 267.
- Merabet, S., & Mann, R. S. (2016). To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends in genetics : TIG*, 32(6), 334–347.
- Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martínez, C., & Torres, M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature*, 402(6760), 425–429.
- Mercader, N., Leonardo, E., Piedra, M. E., Martínez-A, C., Ros, M. A., & Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development (Cambridge, England)*, 127(18), 3961–3970.
- Mercader, N., Tanaka, E. M., & Torres, M. (2005). Proximodistal identity during vertebrate limb regeneration is regulated by Meis homeodomain proteins. *Development (Cambridge, England)*, 132(18), 4131–4142.
- Micali, N., Ferrai, C., Fernandez-Diaz, L. C., Blasi, F., & Crippa, M. P. (2009). Prep1 directly regulates the intrinsic apoptotic pathway by controlling Bcl-XL levels. *Molecular and cellular biology*, 29(5), 1143–1151.
- Mihaylova, Y., Abnave, P., Kao, D., Hughes, S., Lai, A., Jaber-Hijazi, F., Kosaka, N., & Aboobaker, A. A. (2018). Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Nature communications*, 9(1), 3633.
- Mikhailov, K. V., Slyusarev, G. S., Nikitin, M. A., Logacheva, M. D., Penin, A. A., Aleoshin, V. V., & Panchin, Y. V. (2016). The Genome of *Intoshia linei* Affirms Orthonectids as Highly Simplified Spiralian. *Current biology : CB*, 26(13), 1768–1774.
- Mikkola, I., Bruun, J. A., Holm, T., & Johansen, T. (2001). Superactivation of Pax6-mediated transactivation from paired domain-binding sites by dna-independent recruitment of different homeodomain proteins. *The Journal of biological chemistry*, 276(6), 4109–4118.
- Milech, N., Kees, U. R., & Watt, P. M. (2001). Novel alternative PBX3 isoforms in leukemia cells with distinct interaction specificities. *Genes, chromosomes & cancer*, 32(3), 275–280.
- Milech, N., Gottardo, N. G., Ford, J., D'Souza, D., Greene, W. K., Kees, U. R., & Watt, P. M. (2010). MEIS proteins as partners of the TLX1/HOX11 oncoprotein. *Leukemia research*, 34(3), 358–363.
- Miller, J. A. (1938). Studies on heteroplastic transplantation in triclads. I. Cephalic grafts between *Euplanaria dorotocephala* and *E. tigrina*. *Physiol. Zed.* XI, 214–47.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution*, 37(5), 1530–1534.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature methods*, 19(6), 679–682.
- Modica, L., Iotti, G., D'Avola, A., & Blasi, F. (2014). Prep1 (pKnox1) regulates mouse embryonic HSC cycling and self-renewal affecting the Stat1-Sca1 IFN-dependent pathway. *PLoS one*, 9(9), e107916.
- Moens, C. B., & Selleri, L. (2006). Hox cofactors in vertebrate development. *Developmental biology*, 291(2), 193–206.
- Molina, M. D., Saló, E., & Cebrià, F. (2007). The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Developmental biology*, 311(1), 79–94.

- Molina, M. D., Saló, E., & Cebrià, F. (2009). Expression pattern of the expanded noggin gene family in the planarian *Schmidtea mediterranea*. *Gene expression patterns : GEP*, 9(4), 246–253.
- Molina, M. D., Neto, A., Maeso, I., Gómez-Skarmeta, J. L., Saló, E., & Cebrià, F. (2011). Noggin and noggin-like genes control dorsoventral axis regeneration in planarians. *Current biology : CB*, 21(4), 300–305.
- Molinaro, A. M., & Pearson, B. J. (2016). In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome biology*, 17, 87.
- Monica, K., Galili, N., Nourse, J., Saltman, D., & Cleary, M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Molecular and cellular biology*, 11(12), 6149–6157.
- Moreland, R. T., Nguyen, A. D., Ryan, J. F., & Baxevas, A. D. (2020). The Mnemiopsis Genome Project Portal: integrating new gene expression resources and improving data visualization. *Database : the journal of biological databases and curation*, 2020, baaa029.
- Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, 24(3), 319–324.
- Morgan, T. H. (1898). Experimental studies of the regeneration of *Planaria maculata* (Vol. 2). W. Engelmann.
- Morgan, T. H. (1900). Regeneration in Planarians. *Archiv für Entwicklungsmechanik der Organismen* 10, 58–119.
- Morgan, T. H. (1901a). Growth and regeneration in *Planaria lugubris*. *Archiv für Entwicklungsmechanik der Organismen* 13, 179–212.
- Morgan, T. H. (1901b). Regeneration (No. 7). Macmillan.
- Morgan, T. H. (1904). Polarity and axial heteromorphosis. *Am Nat* 38, 502–505
- Morino, Y., Hashimoto, N., & Wada, H. (2017). Expansion of TALE homeobox genes and the evolution of spiralian development. *Nature ecology & evolution*, 1(12), 1942–1949.
- Morino Y. (2022). Dynamic evolutionary history of spiralian-specific TALE homeobox genes in mollusks. *Development, growth & differentiation*, 64(4), 198–209.
- Morino, Y., & Yoshikawa, H. (2023). Role of maternal spiralian-specific homeobox gene SPILE-E in the specification of blastomeres along the animal-vegetal axis during the early cleavage stages of mollusks. *Development, growth & differentiation*, 65(7), 384–394.
- Moskow, J. J., Bullrich, F., Huebner, K., Daar, I. O., & Buchberg, A. M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Molecular and cellular biology*, 15(10), 5434–5443.
- Mukherjee, K., & Bürglin, T. R. (2007). Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *Journal of molecular evolution*, 65(2), 137–153.
- Murad, R., Macias-Muñoz, A., Wong, A., Ma, X., & Mortazavi, A. (2021). Coordinated Gene Expression and Chromatin Regulation during Hydra Head Regeneration. *Genome biology and evolution*, 13(12), evab221.
- Nail, A. N., Smith, J. J., Peterson, M. L., & Spear, B. T. (2020). Evolutionary Analysis of the Zinc Finger and Homeoboxes Family of Proteins Identifies Multiple Conserved Domains and a Common Early Chordate Ancestor. *Genome biology and evolution*, 12(3), 174–184.
- Nakamura, T., Largaespada, D. A., Shaughnessy, J. D., Jr, Jenkins, N. A., & Copeland, N. G. (1996a). Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature genetics*, 12(2), 149–153.
- Nakamura, T., Jenkins, N. A., & Copeland, N. G. (1996b). Identification of a new family of Pbx-related homeobox genes. *Oncogene*, 13(10), 2235–2242.
- Nelson, S. M., Park, L., & Stenkamp, D. L. (2009). Retinal homeobox 1 is required for retinal neurogenesis and photoreceptor differentiation in embryonic zebrafish. *Developmental biology*, 328(1), 24–39.
- Neto, M., Naval-Sánchez, M., Potier, D., Pereira, P. S., Geerts, D., Aerts, S., & Casares, F. (2017). Nuclear receptors connect progenitor transcription factors to cell cycle control. *Scientific reports*, 7(1), 4845.

- Neuteboom, S. T., Peltenburg, L. T., van Dijk, M. A., & Murre, C. (1995). The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 92(20), 9166–9170.
- Neuteboom, S. T., & Murre, C. (1997). Pbx raises the DNA binding specificity but not the selectivity of antennapedia Hox proteins. *Molecular and cellular biology*, 17(8), 4696–4706.
- Newmark, P. A., & Sánchez Alvarado, A. (2000). Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. *Developmental biology*, 220(2), 142–153.
- Newmark, P. A., Wang, Y., & Chong, T. (2008). Germ cell specification and regeneration in planarians. *Cold Spring Harbor symposia on quantitative biology*, 73, 573–581.
- Niehrs, C. (2010). On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development*, 137(6), 845–857.
- Nogi, T., & Levin, M. (2005). Characterization of innexin gene expression and functional roles of gap-junctional communication in planarian regeneration. *Developmental biology*, 287(2), 314–335.
- Noro, B., Culi, J., McKay, D. J., Zhang, W., & Mann, R. S. (2006). Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax. *Genes & development*, 20(12), 1636–1650.
- Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D., & Cleary, M. L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, 60(4), 535–545.
- Oderberg, I. M., Li, D. J., Scimone, M. L., Gaviño, M. A., & Reddien, P. W. (2017). Landmarks in Existing Tissue at Wounds Are Utilized to Generate Pattern in Regenerating Tissue. *Current biology : CB*, 27(5), 733–742.
- Ohno, S. (1970). Evolution by gene duplication. New York: Springer Verlag.
- Okada, Y. K., & Sugino, H. (1934). Transplantation experiments in Planaria gonocephala, I. *Proceedings of the Imperial Academy*, 10(1), 37–40.
- Okada, Y. K., & Sugino, H. (1937). Transplantation experiments in planaria gonocephala Duges. *Jap. J. Zool*, 7, 373–439.
- Onal, P., Grün, D., Adamidi, C., Rybak, A., Solana, J., Mastrobuoni, G., Wang, Y., Rahn, H. P., Chen, W., Kempa, S., Ziebold, U., & Rajewsky, N. (2012). Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *The EMBO journal*, 31(12), 2755–2769.
- Oriente, F., Fernandez Diaz, L. C., Miele, C., Iovino, S., Mori, S., Diaz, V. M., Troncone, G., Cassese, A., Formisano, P., Blasi, F., & Beguinot, F. (2008). Prep1 deficiency induces protection from diabetes and increased insulin sensitivity through a p160-mediated mechanism. *Molecular and cellular biology*, 28(18), 5634–5645.
- Orii, H., & Watanabe, K. (2007). Bone morphogenetic protein is required for dorso-ventral patterning in the planarian Dugesia japonica. *Development, growth & differentiation*, 49(4), 345–349.
- Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P., & Dollé, P. (1997). Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. *Developmental dynamics : an official publication of the American Association of Anatomists*, 210(2), 173–183.
- Oviedo, N. J., & Levin, M. (2007). smedinx-11 is a planarian stem cell gap junction gene required for regeneration and homeostasis. *Development (Cambridge, England)*, 134(17), 3121–3131.
- Owen, J. H., Wagner, D. E., Chen, C. C., Petersen, C. P., & Reddien, P. W. (2015). teashirt is required for head-versus-tail regeneration polarity in planarians. *Development (Cambridge, England)*, 142(6), 1062–1072.
- Owlarn, S., Klenner, F., Schmidt, D., Rabert, F., Tomasso, A., Reuter, H., Mulaw, M. A., Moritz, S., Gentile, L., Weidinger, G., & Bartscherer, K. (2017). Generic wound signals initiate regeneration in missing-tissue contexts. *Nature communications*, 8(1), 2282.

- Pai, C. Y., Kuo, T. S., Jaw, T. J., Kurant, E., Chen, C. T., Bessarab, D. A., Salzberg, A., & Sun, Y. H. (1998). The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in *Drosophila*. *Genes & development*, 12(3), 435–446.
- Palakodeti, D., Smielewska, M., Lu, Y. C., Yeo, G. W., & Graveley, B. R. (2008). The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA (New York, N.Y.)*, 14(6), 1174–1186.
- Pallas, P.S. (1774). *Spicilegia zoologica : quibus novae imprimis et obscurae animalium species iconibus, descriptionibus atque commentariis illustrantur*. Berolini: Prostant apud Gottl. August. Lange
- Palmigiano, A., Santaniello, F., Cerutti, A., Penkov, D., Purushothaman, D., Makhija, E., Luzi, L., di Fagagna, F. D., Pelicci, P. G., Shivashankar, V., Dellino, G. I., & Blasi, F. (2018). PREP1 tumor suppressor protects the late-replicating DNA by controlling its replication timing and symmetry. *Scientific reports*, 8(1), 3198.
- Paps, J., Xu, F., Zhang, G., & Holland, P. W. (2015). Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. *Genome biology and evolution*, 7(3), 677–688.
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.
- Park, C., Owusu-Boaitey, K. E., Valdes, G. M., & Reddien, P. W. (2023). Fate specification is spatially intermingled across planarian stem cells. *Nature Communications*, 14(1), 7422.
- Pascual Carreras, E. (2020). Autonomous and non-autonomous regulation on planarian growth and regeneration: Smed-bls, canonical Wnt signalling and Fox family.
- Pascual-Carreras, E., Sureda-Gómez, M., Barrull-Mascaró, R., Jordà, N., Gelabert, M., Coronel-Córdoba, P., Saló, E., & Adell, T. (2021). WNT-FRIZZLED-LRP5/6 Signaling Mediates Posterior Fate and Proliferation during Planarian Regeneration. *Genes*, 12(1), 101.
- Pascual-Carreras, E., Marín-Barba, M., Castillo-Lara, S., Coronel-Córdoba, P., Magri, M. S., Wheeler, G. N., Gómez-Skarmeta, J. L., Abril, J. F., Saló, E., & Adell, T. (2023). Wnt/ β -catenin signalling is required for pole-specific chromatin remodeling during planarian regeneration. *Nature communications*, 14(1), 298.
- Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., & Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*, 397(6721), 714–719.
- Peifer, M., & Wieschaus, E. (1990). Mutations in the *Drosophila* gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. *Genes & development*, 4(7), 1209–1223.
- Peers, B., Sharma, S., Johnson, T., Kamps, M., & Montminy, M. (1995). The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain. *Molecular and cellular biology*, 15(12), 7091–7097.
- Pellettieri, J., Fitzgerald, P., Watanabe, S., Mancuso, J., Green, D. R., & Sánchez Alvarado, A. (2010). Cell death and tissue remodeling in planarian regeneration. *Developmental biology*, 338(1), 76–85.
- Peltenburg, L. T., & Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *The EMBO journal*, 15(13), 3385–3393.
- Penkov, D., Di Rosa, P., Fernandez Diaz, L., Basso, V., Ferretti, E., Grassi, F., Mondino, A., & Blasi, F. (2005). Involvement of Prep1 in the alphabeta T-cell receptor T-lymphocytic potential of hematopoietic precursors. *Molecular and cellular biology*, 25(24), 10768–10781.
- Penkov, D., Palazzolo, M., Mondino, A., & Blasi, F. (2008). Cytosolic sequestration of Prep1 influences early stages of T cell development. *PloS one*, 3(6), e2424.
- Penkov, D., Mateos San Martín, D., Fernandez-Díaz, L. C., Rosselló, C. A., Torroja, C., Sánchez-Cabo, F., Warnatz, H. J., Sultan, M., Yaspo, M. L., Gabrieli, A., Tkachuk, V., Brendolan, A., Blasi, F., & Torres, M. (2013). Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell reports*, 3(4), 1321–1333.
- Pérez-Posada, A., García-Castro, H., Emili, E., Vanni, V., Arias-Baldrich, C., Frölich, S., van Heeringen, S. J., Kenny, N., & Solana, J. (2024). The Regulatory Logic of Planarian Stem Cell Differentiation. *bioRxiv*, 2024-08.

- Persikov, A. V., Osada, R., & Singh, M. (2009). Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics (Oxford, England)*, 25(1), 22–29.
- Persikov, A. V., & Singh, M. (2014). De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic acids research*, 42(1), 97–108.
- Peter, I. S., & Davidson, E. H. (2015). Genomic control process: development and evolution. Academic Press.
- Petersen, C. P., & Reddien, P. W. (2008). Smed-betacatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science (New York, N.Y.)*, 319(5861), 327–330.
- Petersen, C. P., & Reddien, P. W. (2009a). Wnt signaling and the polarity of the primary body axis. *Cell*, 139(6), 1056–1068.
- Petersen, C. P., & Reddien, P. W. (2009b). A wound-induced Wnt expression program controls planarian regeneration polarity. *Proceedings of the National Academy of Sciences*, 106(40), 17061–17066.
- Petersen, C. P., & Reddien, P. W. (2011). Polarized notum activation at wounds inhibits Wnt function to promote planarian head regeneration. *Science (New York, N.Y.)*, 332(6031), 852–855.
- Petersen, C. P. (2023). Wnt signaling in whole-body regeneration. In *Current Topics in Developmental Biology* (Vol. 153, pp. 347–380). Academic Press.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science : a publication of the Protein Society*, 30(1), 70–82.
- Phelan, M. L., Rambaldi, I., & Featherstone, M. S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Molecular and cellular biology*, 15(8), 3989–3997.
- Phelan, M. L., & Featherstone, M. S. (1997). Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers. *The Journal of biological chemistry*, 272(13), 8635–8643.
- Phuangphong, S., Tsunoda, J., Wada, H., & Morino, Y. (2021). Duplication of spiralian-specific TALE genes and evolution of the blastomere specification mechanism in the bivalve lineage. *EvoDevo*, 12(1), 11.
- Pillay, L. M., Forrester, A. M., Erickson, T., Berman, J. N., & Waskiewicz, A. J. (2010). The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. *Developmental biology*, 340(2), 306–317.
- Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., & Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell*, 96(4), 587–597.
- Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F. J., Kocks, C., & Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science (New York, N.Y.)*, 360(6391), eaaq1723.
- Pöpperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S., & Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell*, 81(7), 1031–1042.
- Pöpperl, H., Rikhof, H., Chang, H., Haffter, P., Kimmel, C. B., & Moens, C. B. (2000). *lazarus* is a novel *pbx* gene that globally mediates *hox* gene function in zebrafish. *Molecular cell*, 6(2), 255–267.
- Poss, K. D., & Tanaka, E. M. (2024). Hallmarks of regeneration. *Cell stem cell*, 31(9), 1244–1261.
- Preissl, S., Gaulton, K. J., & Ren, B. (2023). Characterizing cis-regulatory elements using single-cell epigenomics. *Nature reviews. Genetics*, 24(1), 21–43.
- Purushothaman, D., Bianchi, L. F., Penkov, D., Poli, A., Li, Q., Vermezovic, J., Pramotton, F. M., Choudhary, R., Pennacchio, F. A., Sommariva, E., Foiani, M., Gauthier, N., Maiuri, P., & Blasi, F. (2022). The transcription factor PREP1(PKNOX1) regulates nuclear stiffness, the expression of LINC complex proteins and mechanotransduction. *Communications biology*, 5(1), 456.

- Quinlan A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics*, 47, 11.12.1–11.12.34.
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*, 44(W1), W160–W165.
- Ramirez, A. N., Loubet-Seneor, K., & Srivastava, M. (2020). A Regulatory Program for Initiation of Wnt Signaling during Posterior Regeneration. *Cell reports*, 32(9), 108098.
- Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J. A., Ferenc, K., Kumar, V., Lemma, R. B., Lucas, J., Chèneby, J., Baranasic, D., Khan, A., Fornes, O., Gundersen, S., Johansen, M., Hovig, E., Lenhard, B., Sandelin, A., Wasserman, W. W., Parcy, F., & Mathelier, A. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic acids research*, 52(D1), D174–D182.
- Rauskolb, C., Peifer, M., & Wieschaus, E. (1993). extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. *Cell*, 74(6), 1101–1112.
- Rauskolb, C., & Wieschaus, E. (1994). Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins. *The EMBO journal*, 13(15), 3561–3569.
- Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C. O., Forrest, A. R., Gough, J., Grimmond, S., Han, J. H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., ... Hayashizaki, Y. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, 140(5), 744–752.
- Rave-Harel, N., Givens, M. L., Nelson, S. B., Duong, H. A., Coss, D., Clark, M. E., Hall, S. B., Kamps, M. P., & Mellon, P. L. (2004). TALE homeodomain proteins regulate gonadotropin-releasing hormone gene expression independently and via interactions with Oct-1. *The Journal of biological chemistry*, 279(29), 30287–30297.
- Raz, A. A., Srivastava, M., Salvamoser, R., & Reddien, P. W. (2017). Acoel regeneration mechanisms indicate an ancient role for muscle in regenerative patterning. *Nature communications*, 8(1), 1260.
- Raz, A. A., Wurtzel, O., & Reddien, P. W. (2021). Planarian stem cells specify fate yet retain potency during the cell cycle. *Cell stem cell*, 28(7), 1307–1322.e5.
- Reddien, P. W., & Sánchez Alvarado, A. (2004). Fundamentals of planarian regeneration. *Annual review of cell and developmental biology*, 20, 725–757.
- Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R., & Sánchez Alvarado, A. (2005). Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental cell*, 8(5), 635–649.
- Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C., & Sánchez Alvarado, A. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science (New York, N.Y.)*, 310(5752), 1327–1330.
- Reddien, P. W., Bermange, A. L., Kicza, A. M., & Sánchez Alvarado, A. (2007). BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. *Development (Cambridge, England)*, 134(22), 4043–4051.
- Reddien P. W. (2013). Specialized progenitors and regeneration. *Development (Cambridge, England)*, 140(5), 951–957.
- Reddien P. W. (2018). The Cellular and Molecular Basis for Planarian Regeneration. *Cell*, 175(2), 327–345.
- Reddien, P. W. (2021). Principles of regeneration revealed by the planarian eye. *Current opinion in cell biology*, 73, 19–25.
- Reddien P. W. (2022). Positional Information and Stem Cells Combine to Result in Planarian Regeneration. *Cold Spring Harbor perspectives in biology*, 14(4), a040717.
- Reichlmeir, M., Elias, L., & Schulte, D. (2021). Posttranslational Modifications in Conserved Transcription Factors: A Survey of the TALE-Homeodomain Superclass in Human and Mouse. *Frontiers in cell and developmental biology*, 9, 648765.
- Rentzsch, F., & Holstein, T. W. (2018). Making head or tail of cnidarian hox gene function. *Nature Communications*, 9(1), 2187.

- Reuter, H., März, M., Vogg, M. C., Eccles, D., Grífol-Boldú, L., Wehner, D., Owlarn, S., Adell, T., Weidinger, G., & Bartscherer, K. (2015). B-catenin-dependent control of positional information along the AP body axis in planarians involves a teashirt family member. *Cell reports*, 10(2), 253–265.
- Rhee, J. W., Arata, A., Selleri, L., Jacobs, Y., Arata, S., Onimaru, H., & Cleary, M. L. (2004). Pbx3 deficiency results in central hypoventilation. *The American journal of pathology*, 165(4), 1343–1350.
- Ricci, L., & Srivastava, M. (2021). Transgenesis in the acoel worm *Hofstenia miamia*. *Developmental cell*, 56(22), 3160–3170.e4.
- Rieckhof, G. E., Casares, F., Ryoo, H. D., Abu-Shaar, M., & Mann, R. S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*, 91(2), 171–183.
- Rink, J. C., Gurley, K. A., Elliott, S. A., & Sánchez Alvarado, A. (2009). Planarian Hh signaling regulates regeneration polarity and links Hh pathway evolution to cilia. *Science (New York, N.Y.)*, 326(5958), 1406–1410.
- Roberts-Galbraith, R. H., & Newmark, P. A. (2013). Follistatin antagonizes activin signaling and acts with notum to direct planarian head regeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4), 1363–1368.
- Roberts-Galbraith, R. H., Brubacher, J. L., & Newmark, P. A. (2016). A functional genomics screen in planarians reveals regulators of whole-brain regeneration. *eLife*, 5, e17002.
- Rodríguez-Esteban, G., González-Sastre, A., Rojo-Laguna, J. I., Saló, E., & Abril, J. F. (2015). Digital gene expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in *Schmidtea mediterranea*. *BMC genomics*, 16(1), 361.
- Romero, B. T., Evans, D. J., & Aboobaker, A. A. (2012). FACS analysis of the planarian stem cell compartment as a tool to understand regenerative mechanisms. *Methods in molecular biology (Clifton, N.J.)*, 916, 167–179.
- Ross, K. G., Currie, K. W., Pearson, B. J., & Zayas, R. M. (2017). Nervous system development and regeneration in freshwater planarians. *Wiley interdisciplinary reviews. Developmental biology*, 6(3), 10.1002/wdev.266.
- Ross, K. G., Molinaro, A. M., Romero, C., Dockter, B., Cable, K. L., Gonzalez, K., Zhang, S., Collins, E. S., Pearson, B. J., & Zayas, R. M. (2018). SoxB1 Activity Regulates Sensory Neuron Regeneration, Maintenance, and Function in Planarians. *Developmental cell*, 47(3), 331–347.e5.
- Rossi, L., Salvetti, A., Lena, A., Batistoni, R., Deri, P., Pugliesi, C., Loreti, E., & Gremigni, V. (2006). DjPiwi-1, a member of the PAZ-Piwi gene family, defines a subpopulation of planarian stem cells. *Development genes and evolution*, 216(6), 335–346.
- Rossi, L., Salvetti, A., Marincola, F. M., Lena, A., Deri, P., Mannini, L., Batistoni, R., Wang, E., & Gremigni, V. (2007). Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile. *Genome biology*, 8(4), R62.
- Rouhana, L., Shibata, N., Nishimura, O., & Agata, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Developmental biology*, 341(2), 429–443.
- Rowan, S., Siggers, T., Lachke, S. A., Yue, Y., Bulyk, M. L., & Maas, R. L. (2010). Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes & development*, 24(10), 980–985.
- Ryan, J. F., Pang, K., NISC Comparative Sequencing Program, Mullikin, J. C., Martindale, M. Q., & Baxevanis, A. D. (2010). The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *EvoDevo*, 1(1), 9.
- Ryoo, H. D., & Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes & development*, 13(13), 1704–1716.
- Ryoo, H. D., Marty, T., Casares, F., Affolter, M., & Mann, R. S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development (Cambridge, England)*, 126(22), 5137–5148.
- Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., Brena, C., Pradel, J., & Graba, Y. (2011). Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. *Proceedings of the*

National Academy of Sciences of the United States of America, 108(6), 2276–2281.

Saito, Y., Koinuma, S., Watanabe, K., & Agata, K. (2003). Mediolateral intercalation in planarians revealed by grafting experiments. *Developmental dynamics: an official publication of the American Association of Anatomists*, 226(2), 334–340.

Saleh, M., Rambaldi, I., Yang, X. J., & Featherstone, M. S. (2000a). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. *Molecular and cellular biology*, 20(22), 8623–8633.

Saleh, M., Huang, H., Green, N. C., & Featherstone, M. S. (2000b). A conformational change in PBX1A is necessary for its nuclear localization. *Experimental cell research*, 260(1), 105–115.

Saló, E., & Baguñà, J. (1984). Regeneration and pattern formation in planarians I. The pattern of mitosis in anterior and posterior regeneration in *Dugesia (G) tigrina*, and a new proposal for blastema formation. *Development*, 83(1), 63–80.

Saló, E., & Baguñà, J. (1985). Proximal and distal transformation during intercalary regeneration in the planarian *Dugesia (S) mediterranea*: Evidence using a chromosomal marker. *Wilhelm Roux's archives of developmental biology*, 194, 364–368.

Salvetti, A., Rossi, L., Lena, A., Batistoni, R., Deri, P., Rainaldi, G., Locci, M. T., Evangelista, M., & Gremigni, V. (2005). DjPum, a homologue of *Drosophila* Pumilio, is essential to planarian stem cell maintenance. *Development (Cambridge, England)*, 132(8), 1863–1874.

Salzberg, A., Elias, S., Nachaliel, N., Bonstein, L., Henig, C., & Frank, D. (1999). A Meis family protein caudalizes neural cell fates in *Xenopus*. *Mechanisms of development*, 80(1), 3–13.

Samad, O. A., Geisen, M. J., Caronia, G., Varlet, I., Zappavigna, V., Ericson, J., Goridis, C., & Rijli, F. M. (2004). Integration of anteroposterior and dorsoventral regulation of Phox2b transcription in cranial motoneuron progenitors by homeodomain proteins. *Development (Cambridge, England)*, 131(16), 4071–4083.

Sánchez Alvarado, A., & Newmark, P. A. (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), 5049–5054.

Sánchez-Guardado, L. Ó., Irimia, M., Sánchez-Arrones, L., Burguera, D., Rodríguez-Gallardo, L., García-Fernández, J., Puelles, L., Ferran, J. L., & Hidalgo-Sánchez, M. (2011). Distinct and redundant expression and transcriptional diversity of MEIS gene paralogs during chicken development. *Developmental dynamics: an official publication of the American Association of Anatomists*, 240(6), 1475–1492.

Santos, F.V. (1929). Studies on transplantation in planarian. *Biol. Bull.* 57, 188–97.

Santos FV. (1931). Studies on transplantation in planaria. *Phys. Zool.* 4, 111–64.

Sanyal, M., Tung, J. W., Karsunky, H., Zeng, H., Selleri, L., Weissman, I. L., Herzenberg, L. A., & Cleary, M. L. (2007). B-cell development fails in the absence of the Pbx1 proto-oncogene. *Blood*, 109(10), 4191–4199.

Sarkar, S. R., Dubey, V. K., Jahagirdar, A., Lakshmanan, V., Haroon, M. M., Sowndarya, S., Sowdhamini, R., & Palakodeti, D. (2022). DDX24 is required for muscle fiber organization and the suppression of wound-induced Wnt activity necessary for pole re-establishment during planarian regeneration. *Developmental biology*, 488, 11–29.

Sartorelli, V., & Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nature structural & molecular biology*, 27(6), 521–528.

Sato, K., Shibata, N., Orii, H., Amikura, R., Sakurai, T., Agata, K., Kobayashi, S., & Watanabe, K. (2006). Identification and origin of the germline stem cells as revealed by the expression of nanos-related gene in planarians. *Development, growth & differentiation*, 48(9), 615–628.

Schad, E. G., & Petersen, C. P. (2020). STRIPAK Limits Stem Cell Differentiation of a WNT Signaling Center to Control Planarian Axis Scaling. *Current biology: CB*, 30(2), 254–263.e2.

Schewtschenko, N. N. (1936a). Der Einfluss des KBrperteilea von hoher physiologischer Aktivität auf den Regenerationsvorgang bei Planarien. *Proc. Inst. Zool., Kiev*, 43–64.

- Schewtschenko, N. N. (1936b). Polaritätsumkehr bei der Unterdrückung der dominierenden Region bei Planuria. 2001. *Anz.* 115, 232-44.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), 676–682.
- Schnabel, C. A., Selleri, L., Jacobs, Y., Warnke, R., & Cleary, M. L. (2001). Expression of Pbx1b during mammalian organogenesis. *Mechanisms of development*, 100(1), 131–135.
- Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S. W., Dimitrova, E., Dimond, A., Edelman, L. B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., LeProust, E., ... Fraser, P. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome research*, 25(4), 582–597.
- Schulte, D., & Frank, D. (2014). TALE transcription factors during early development of the vertebrate brain and eye. *Developmental dynamics : an official publication of the American Association of Anatomists*, 243(1), 99–116.
- Schulte, D., & Geerts, D. (2019). MEIS transcription factors in development and disease. *Development (Cambridge, England)*, 146(16), dev174706.
- Schultz, D. T., Haddock, S. H., Bredeson, J. V., Green, R. E., Simakov, O., & Rokhsar, D. S. (2023). Ancient gene linkages support ctenophores as sister to other animals. *Nature*, 618(7963), 110-117.
- Scimone, M. L., Meisel, J., & Reddien, P. W. (2010). The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian Schmidtea mediterranea. *Development (Cambridge, England)*, 137(8), 1231–1241.
- Scimone, M. L., Srivastava, M., Bell, G. W., & Reddien, P. W. (2011). A regulatory program for excretory system regeneration in planarians. *Development (Cambridge, England)*, 138(20), 4387–4398.
- Scimone, M. L., Lapan, S. W., & Reddien, P. W. (2014a). A forkhead transcription factor is wound-induced at the planarian midline and required for anterior pole regeneration. *PLoS genetics*, 10(1), e1003999.
- Scimone, M. L., Kravarik, K. M., Lapan, S. W., & Reddien, P. W. (2014b). Neoblast specialization in regeneration of the planarian Schmidtea mediterranea. *Stem cell reports*, 3(2), 339–352.
- Scimone, M. L., Cote, L. E., Rogers, T., & Reddien, P. W. (2016). Two FGFR1-Wnt circuits organize the planarian anteroposterior axis. *eLife*, 5, e12845.
- Scimone, M. L., Cote, L. E., & Reddien, P. W. (2017). Orthogonal muscle fibres have different instructive roles in planarian regeneration. *Nature*, 551(7682), 623–628.
- Scimone, M. L., Wurtzel, O., Malecek, K., Fincher, C. T., Oderberg, I. M., Kravarik, K. M., & Reddien, P. W. (2018). foxF-1 Controls Specification of Non-body Wall Muscle and Phagocytic Cells in Planarians. *Current biology : CB*, 28(23), 3787–3801.e6.
- Scimone, M. L., Atabay, K. D., Fincher, C. T., Bonneau, A. R., Li, D. J., & Reddien, P. W. (2020). Muscle and neuronal guidepost-like cells facilitate planarian visual system regeneration. *Science*, 368(6498), eaba3203.
- Sea Urchin Genome Sequencing Consortium, Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Etensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., ... Wright, R. (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science (New York, N.Y.)*, 314(5801), 941–952.
- Selleri, L., Depew, M. J., Jacobs, Y., Chanda, S. K., Tsang, K. Y., Cheah, K. S., Rubenstein, J. L., O'Gorman, S., & Cleary, M. L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development (Cambridge, England)*, 128(18), 3543–3557.
- Selleri, L., Zappavigna, V., & Ferretti, E. (2019). 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks. *Genes & development*, 33(5-6), 258-275.
- Shanmugam, K., Green, N. C., Rambaldi, I., Saragovi, H. U., & Featherstone, M. S. (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. *Molecular and cellular biology*, 19(11), 7577–7588.

- Shen, W. F., Chang, C. P., Rozenfeld, S., Sauvageau, G., Humphries, R. K., Lu, M., Lawrence, H. J., Cleary, M. L., & Largman, C. (1996). Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. *Nucleic acids research*, 24(5), 898–906.
- Shen, W. F., Rozenfeld, S., Lawrence, H. J., & Largman, C. (1997a). The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *The Journal of biological chemistry*, 272(13), 8198–8206.
- Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M., & Largman, C. (1997b). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Molecular and cellular biology*, 17(11), 6448–6458.
- Shen, W. F., Rozenfeld, S., Kwong, A., Köm ves, L. G., Lawrence, H. J., & Largman, C. (1999). HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. *Molecular and cellular biology*, 19(4), 3051–3061.
- Shi, Z., Jiao, S., & Zhou, Z. (2016). STRIPAK complexes in cell signaling and cancer. *Oncogene*, 35(35), 4549–4557.
- Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., & Agata, K. (1999). Expression of vasa(vas)-related genes in germline cells and totipotent somatic stem cells of planarians. *Developmental biology*, 206(1), 73–87.
- Shine, I., & Wrobel, S. (2014). *Thomas Hunt Morgan: pioneer of genetics*. University Press of Kentucky.
- Sikes, J. M., & Newmark, P. A. (2013). Restoration of anterior regeneration in a planarian with limited regenerative ability. *Nature*, 500(7460), 77–80.
- Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D. H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J. A., Shapiro, H., Aerts, A., Otilar, R. P., Terry, A. Y., Boore, J. L., ... Rokhsar, D. S. (2013). Insights into bilaterian evolution from three spiralian genomes. *Nature*, 493(7433), 526–531.
- Simakov, O., Marlétaz, F., Yue, J. X., O'Connell, B., Jenkins, J., Brandt, A., Calef, R., Tung, C. H., Huang, T. K., Schmutz, J., Satoh, N., Yu, J. K., Putnam, N. H., Green, R. E., & Rokhsar, D. S. (2020). Deeply conserved syntenic resolves early events in vertebrate evolution. *Nature ecology & evolution*, 4(6), 820–830.
- Singh, A., Tare, M., Kango-Singh, M., Son, W. S., Cho, K. O., & Choi, K. W. (2011). Opposing interactions between homothorax and Lobe define the ventral eye margin of Drosophila eye. *Developmental biology*, 359(2), 199–208.
- Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J., & Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*, 147(6), 1270–1282.
- Sluys, R., & Riutort, M. (2018). Planarian diversity and phylogeny. *Planarian regeneration: methods and protocols*, 1–56.
- Smith, J. E., Jr, Bollekens, J. A., Inghirami, G., & Takeshita, K. (1997). Cloning and mapping of the MEIS1 gene, the human homolog of a murine leukemogenic gene. *Genomics*, 43(1), 99–103.
- Solana, J., Lasko, P., & Romero, R. (2009). Spoltud-1 is a chromatoid body component required for planarian long-term stem cell self-renewal. *Developmental biology*, 328(2), 410–421.
- Solana, J., Kao, D., Mihaylova, Y., Jaber-Hijazi, F., Malla, S., Wilson, R., & Aboobaker, A. (2012). Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome biology*, 13(3), R19.
- Solana J. (2013). Closing the circle of germline and stem cells: the Primordial Stem Cell hypothesis. *EvoDevo*, 4(1), 2.
- Soutourina J. (2018). Transcription regulation by the Mediator complex. *Nature reviews. Molecular cell biology*, 19(4), 262–274.
- Spallanzani, L. (1768). Prodrómo di un'opera da imprimersi sopra le riproduzioni animali dato in luce dall'abate Spallanzani. Modena: Nella stamperia di Giovanni Montanari.

- Spemann, H. & Mangold, H. (1924). Über induktion von embryonalanlagen durch implantation artfremder organisatoren. *Wilhelm Roux Arch. Entw. Mech. Org.* 100, 599–638.
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T. H., Plachetzki, D. C., Zhai, Y., Adamski, M., ... Rokhsar, D. S. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature*, 466(7307), 720–726.
- Srivastava, M., Mazza-Curll, K. L., van Wolfswinkel, J. C., & Reddien, P. W. (2014). Whole-body acoeel regeneration is controlled by Wnt and Bmp-Admp signaling. *Current biology : CB*, 24(10), 1107–1113.
- Srivastava M. (2021). Beyond Casual Resemblance: Rigorous Frameworks for Comparing Regeneration Across Species. *Annual review of cell and developmental biology*, 37, 415–440.
- Stankunas, K., Shang, C., Twu, K. Y., Kao, S. C., Jenkins, N. A., Copeland, N. G., Sanyal, M., Selleri, L., Cleary, M. L., & Chang, C. P. (2008). Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circulation research*, 103(7), 702–709.
- Stanney, W., 3rd, Ladam, F., Donaldson, I. J., Parsons, T. J., Maehr, R., Bobola, N., & Sagerström, C. G. (2020). Combinatorial action of NF-Y and TALE at embryonic enhancers defines distinct gene expression programs during zygotic genome activation in zebrafish. *Developmental biology*, 459(2), 161–180.
- Stapornwongkul, K. S., & Vincent, J. P. (2021). Generation of extracellular morphogen gradients: the case for diffusion. *Nature reviews. Genetics*, 22(6), 393–411.
- Steelman, S., Moskow, J. J., Muzynski, K., North, C., Druck, T., Montgomery, J. C., Huebner, K., Daar, I. O., & Buchberg, A. M. (1997). Identification of a conserved family of Meis1-related homeobox genes. *Genome research*, 7(2), 142–156.
- Stelman, C. R., Smith, B. M., Chandra, B., & Roberts-Galbraith, R. H. (2021). CBP/p300 homologs CBP2 and CBP3 play distinct roles in planarian stem cell function. *Developmental Biology*, 473, 130–143.
- Stephens, S. G. (1951). Possible significances of duplication in evolution. *Advances in genetics*, 4, 247–265.
- Stevens, K. E., & Mann, R. S. (2007). A balance between two nuclear localization sequences and a nuclear export sequence governs extradenticle subcellular localization. *Genetics*, 175(4), 1625–1636.
- Stückemann, T., Cleland, J. P., Werner, S., Thi-Kim Vu, H., Bayersdorf, R., Liu, S. Y., Friedrich, B., Jülicher, F., & Rink, J. C. (2017). Antagonistic Self-Organizing Patterning Systems Control Maintenance and Regeneration of the Anteroposterior Axis in Planarians. *Developmental cell*, 40(3), 248–263.e4.
- Subramaniam, N., Campión, J., Rafter, I., & Okret, S. (2003). Cross-talk between glucocorticoid and retinoic acid signals involving glucocorticoid receptor interaction with the homeodomain protein Pbx1. *The Biochemical journal*, 370(Pt 3), 1087–1095.
- Sugino, H. (1940). Influence of the head pieces implanted in the prepharyngeal region of planarian. *Ann. Zool. Jap.* 19, 245–52.
- Sugino, H. (1941). Homopolar union in *Planaria gonocephalu*. *Jap. J. 2001.* 9, 175–83.
- Sun, F., Ou, J., Shoffner, A. R., Luan, Y., Yang, H., Song, L., Safi, A., Cao, J., Yue, F., Crawford, G. E., & Poss, K. D. (2022). Enhancer selection dictates gene expression responses in remote organs during tissue regeneration. *Nature cell biology*, 24(5), 685–696.
- Sureda-Gómez, M., Pascual-Carreras, E., & Adell, T. (2015). Posterior Wnts Have Distinct Roles in Specification and Patterning of the Planarian Posterior Region. *International journal of molecular sciences*, 16(11), 26543–26554.
- Sureda-Gómez, M., Martín-Durán, J. M., & Adell, T. (2016). Localization of planarian β -CATENIN-1 reveals multiple roles during anterior-posterior regeneration and organogenesis. *Development (Cambridge, England)*, 143(22), 4149–4160.
- Swapna, L. S., Molinaro, A. M., Lindsay-Mosher, N., Pearson, B. J., & Parkinson, J. (2018). Comparative transcriptomic analyses and single-cell RNA sequencing of the freshwater planarian *Schmidtea mediterranea* identify major cell types and pathway conservation. *Genome biology*, 19(1), 124.

- Swift, G. H., Liu, Y., Rose, S. D., Bischof, L. J., Steelman, S., Buchberg, A. M., Wright, C. V., & MacDonald, R. J. (1998). An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2). *Molecular and cellular biology*, 18(9), 5109–5120.
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), 663–676.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5), 861–872.
- Takeuchi, J. K., & Bruneau, B. G. (2007). Irx11, a divergent Iroquois homeobox family transcription factor gene. *Gene Expression Patterns*, 7(1-2), 51-56.
- Takeuchi, T., Koyanagi, R., Gyoja, F., Kanda, M., Hisata, K., Fujie, M., Goto, H., Yamasaki, S., Nagai, K., Morino, Y., Miyamoto, H., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., Satoh, N., & Kawashima, T. (2016). Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological letters*, 2, 3.
- Tan, T. C., Rahman, R., Jaber-Hijazi, F., Felix, D. A., Chen, C., Louis, E. J., & Aboobaker, A. (2012). Telomere maintenance and telomerase activity are differentially regulated in asexual and sexual worms. *Proceedings of the National Academy of Sciences of the United States of America*, 109(11), 4209–4214.
- Tarazona, O. A., Lopez, D. H., Slota, L. A., & Cohn, M. J. (2019). Evolution of limb development in cephalopod mollusks. *eLife*, 8, e43828.
- Tasaki, J., Shibata, N., Nishimura, O., Itomi, K., Tabata, Y., Son, F., Suzuki, N., Araki, R., Abe, M., Agata, K., & Umesono, Y. (2011a). ERK signaling controls blastema cell differentiation during planarian regeneration. *Development (Cambridge, England)*, 138(12), 2417–2427.
- Tasaki, J., Shibata, N., Sakurai, T., Agata, K., & Umesono, Y. (2011b). Role of c-Jun N-terminal kinase activation in blastema formation during planarian regeneration. *Development, growth & differentiation*, 53(3), 389–400.
- Tejada-Romero, B., Carter, J. M., Mihaylova, Y., Neumann, B., & Aboobaker, A. A. (2015). JNK signalling is necessary for a Wnt- and stem cell-dependent regeneration programme. *Development (Cambridge, England)*, 142(14), 2413–2424.
- Tewari, A. G., Stern, S. R., Oderberg, I. M., & Reddien, P. W. (2018). Cellular and molecular responses unique to major injury are dispensable for planarian regeneration. *Cell reports*, 25(9), 2577-2590.
- Tewari, A. G., Owen, J. H., Petersen, C. P., Wagner, D. E., & Reddien, P. W. (2019). A small set of conserved genes, including sp5 and Hox, are activated by Wnt signaling in the posterior of planarians and acoels. *PLoS genetics*, 15(10), e1008401.
- Thi-Kim Vu, H., Rink, J. C., McKinney, S. A., McClain, M., Lakshmanaperumal, N., Alexander, R., & Sánchez Alvarado, A. (2015). Stem cells and fluid flow drive cyst formation in an invertebrate excretory organ. *eLife*, 4, e07405.
- Thompson, J. D., Ou, J., Lee, N., Shin, K., Cigliola, V., Song, L., Crawford, G. E., Kang, J., & Poss, K. D. (2020). Identification and requirements of enhancers that direct gene expression during zebrafish fin regeneration. *Development (Cambridge, England)*, 147(14), dev191262.
- Trembley, A. (1744). X. translation of a letter from Mr. Abraham Trembley, FRS to the President, with observations upon several newly discover'd species of fresh-water Polypi. *Philosophical Transactions of the Royal Society of London*, 43(474), 169-183.
- Trueb, B. (2011). Biology of FGFR1, the fifth fibroblast growth factor receptor. *Cellular and molecular life sciences*, 68, 951-964.
- Tsai, I. J., Zarowiecki, M., Holroyd, N., Garcíarrubio, A., Sánchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R. J., Frago, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H. M., Cai, X., Camicia, F., Clark, R., Cucher, M., De Silva,

- N., Day, T. A., Deplazes, P., ... Berriman, M. (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, 496(7443), 57–63.
- Tsai, T. Y., Sikora, M., Xia, P., Colak-Champollion, T., Knaut, H., Heisenberg, C. P., & Megason, S. G. (2020). An adhesion code ensures robust pattern formation during tissue morphogenesis. *Science (New York, N.Y.)*, 370(6512), 113–116.
- Tümpel, S., Cambroner, F., Ferretti, E., Blasi, F., Wiedemann, L. M., & Krumlauf, R. (2007). Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. *Developmental biology*, 302(2), 646–660.
- Turing, A. M., (1952). The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B*, 23, 737–72.
- Umesono, Y., Tasaki, J., Nishimura, Y., Hrouda, M., Kawaguchi, E., Yazawa, S., Nishimura, O., Hosoda, K., Inoue, T., & Agata, K. (2013). The molecular logic for planarian regeneration along the anterior-posterior axis. *Nature*, 500(7460), 73–76.
- Vaccari, E., Deflorian, G., Bernardi, E., Pauls, S., Tiso, N., Bortolussi, M., & Argenton, F. (2010). prep1.2 and aldh1a2 participate to a positive loop required for branchial arches development in zebrafish. *Developmental biology*, 343(1-2), 94–103.
- Van Aiken, K., Weaver, D., Robertson, B., Sundaram, M., Saldi, T., Edgar, L., Elling, U., Lee, M., Boese, Q., & Wood, W. B. (2002). Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in *C. elegans* embryogenesis. *Development (Cambridge, England)*, 129(22), 5255–5268.
- van Dijk, M. A., & Murre, C. (1994). extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell*, 78(4), 617–624.
- Van Wolfswinkel, J. C., Wagner, D. E., & Reddien, P. W. (2014). Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell stem cell*, 15(3), 326–339.
- Varley, Á., Horkan, H. R., McMahon, E. T., Krasovec, G., & Frank, U. (2023). Pluripotent, germ cell competent adult stem cells underlie cnidarian regenerative ability and clonal growth. *Current biology : CB*, 33(10), 1883–1892.e3. <https://doi.org/10.1016/j.cub.2023.03.039>
- Vásquez-Doorman, C., & Petersen, C. P. (2014). zic-1 Expression in Planarian neoblasts after injury controls anterior pole regeneration. *PLoS genetics*, 10(7), e1004452.
- Vásquez-Doorman, C., & Petersen, C. P. (2016). The NuRD complex component p66 suppresses photoreceptor neuron regeneration in planarians. *Regeneration (Oxford, England)*, 3(3), 168–178.
- Vila-Farré, M., Rozanski, A., Ivanković, M., Cleland, J., Brand, J. N., Thalen, F., Grohme, M. A., von Kannen, S., Grosbusch, A. L., Vu, H. T., Prieto, C. E., Carbayo, F., Egger, B., Bleidorn, C., Rasko, J. E. J., & Rink, J. C. (2023). Evolutionary dynamics of whole-body regeneration across planarian flatworms. *Nature ecology & evolution*, 7(12), 2108–2124.
- Villaescusa, J. C., Buratti, C., Penkov, D., Mathiasen, L., Planagumà, J., Ferretti, E., & Blasi, F. (2009). Cytoplasmic Prep1 interacts with 4EHP inhibiting Hoxb4 translation. *PLoS one*, 4(4), e5213.
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), 854–858.
- Vitobello, A., Ferretti, E., Lampe, X., Vilain, N., Ducret, S., Ori, M., Spetz, J. F., Selleri, L., & Rijli, F. M. (2011). Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation. *Developmental cell*, 20(4), 469–482.
- Vlachakis, N., Ellstrom, D. R., & Sagerström, C. G. (2000). A novel pbx family member expressed during early zebrafish embryogenesis forms trimeric complexes with Meis3 and Hoxb1b. *Developmental dynamics : an official publication of the American Association of Anatomists*, 217(1), 109–119.
- Vlachakis, N., Choe, S. K., & Sagerström, C. G. (2001). Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development (Cambridge, England)*, 128(8), 1299–1312.

- Vogg, M. C., Owlarn, S., Pérez Rico, Y. A., Xie, J., Suzuki, Y., Gentile, L., Wu, W., & Bartscherer, K. (2014). Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. *Developmental biology*, 390(2), 136–148.
- Wagner, K., Mincheva, A., Korn, B., Lichter, P., & Pöpperl, H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mechanisms of development*, 103(1-2), 127–131.
- Wagner, D. E., Wang, I. E., & Reddien, P. W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science (New York, N.Y.)*, 332(6031), 811–816.
- Wagner, D. E., Ho, J. J., & Reddien, P. W. (2012). Genetic regulators of a pluripotent adult stem cell system in planarians identified by RNAi and clonal analysis. *Cell stem cell*, 10(3), 299–311.
- Wang, Y., Yin, L., & Hillgartner, F. B. (2001). The homeodomain proteins PBX and MEIS1 are accessory factors that enhance thyroid hormone regulation of the malic enzyme gene in hepatocytes. *The Journal of biological chemistry*, 276(26), 23838–23848.
- Wang, Z., & Mann, R. S. (2003). Requirement for two nearly identical TGIF-related homeobox genes in Drosophila spermatogenesis. *Development (Cambridge, England)*, 130(13), 2853–2865.
- Wang, Y., Zayas, R. M., Guo, T., & Newmark, P. A. (2007). nanos function is essential for development and regeneration of planarian germ cells. *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), 5901–5906.
- Wang, W., Hu, C. K., Zeng, A., Alegre, D., Hu, D., Gotting, K., Ortega Granillo, A., Wang, Y., Robb, S., Schnittker, R., Zhang, S., Alegre, D., Li, H., Ross, E., Zhang, N., Brunet, A., & Sánchez Alvarado, A. (2020). Changes in regeneration-responsive enhancers shape regenerative capacities in vertebrates. *Science (New York, N.Y.)*, 369(6508), eaaz3090.
- Wang, L. G., Lam, T. T., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular biology and evolution*, 37(2), 599–603.
- Waskiewicz, A. J., Rikhof, H. A., Hernandez, R. E., & Moens, C. B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development (Cambridge, England)*, 128(21), 4139–4151.
- Waskiewicz, A. J., Rikhof, H. A., & Moens, C. B. (2002). Eliminating zebrafish pbx proteins reveals a hindbrain ground state. *Developmental cell*, 3(5), 723–733.
- Weidinger, G., Thorpe, C. J., Wuennenberg-Stapleton, K., Ngai, J., & Moon, R. T. (2005). The Sp1-related transcription factors sp5 and sp5-like act downstream of Wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. *Current biology : CB*, 15(6), 489–500.
- Wenemoser, D., & Reddien, P. W. (2010). Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Developmental biology*, 344(2), 979–991.
- Wenemoser, D., Lapan, S. W., Wilkinson, A. W., Bell, G. W., & Reddien, P. W. (2012). A molecular wound response program associated with regeneration initiation in planarians. *Genes & development*, 26(9), 988–1002.
- Weston, C. R., & Davis, R. J. (2007). The JNK signal transduction pathway. *Current opinion in cell biology*, 19(2), 142–149.
- Wieschaus, E., Nüsslein-Volhard, C., & Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* : III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's archives of developmental biology*, 193(5), 296–307.
- Wiggans, M., Zhu, S. J., Molinaro, A. M., & Pearson, B. J. (2023). The BAF chromatin remodeling complex licenses planarian stem cells access to ectodermal and mesodermal cell fates. *BMC biology*, 21(1), 227.
- Williams, T. M., Williams, M. E., & Innis, J. W. (2005). Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. *Developmental biology*, 277(2), 457–471.

- Witchley, J. N., Mayer, M., Wagner, D. E., Owen, J. H., & Reddien, P. W. (2013). Muscle cells provide instructions for planarian regeneration. *Cell reports*, 4(4), 633–641.
- Wolpert L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, 25(1), 1–47.
- Wurtzel, O., Cote, L. E., Poirier, A., Satija, R., Regev, A., & Reddien, P. W. (2015). A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians. *Developmental cell*, 35(5), 632–645.
- Wurtzel, O., Oderberg, I. M., & Reddien, P. W. (2017). Planarian Epidermal Stem Cells Respond to Positional Cues to Promote Cell-Type Diversity. *Developmental cell*, 40(5), 491–504.e5.
- Xie, Y., Su, N., Yang, J., Tan, Q., Huang, S., Jin, M., Ni, Z., Zhang, B., Zhang, D., Luo, F., Chen, H., Sun, X., Feng, J. Q., Qi, H., & Chen, L. (2020). FGF/FGFR signaling in health and disease. *Signal transduction and targeted therapy*, 5(1), 181.
- Xu, Q., Georgiou, G., Frölich, S., van der Sande, M., Veenstra, G. J. C., Zhou, H., & van Heeringen, S. J. (2021). ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic acids research*, 49(14), 7966–7985.
- Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome biology*, 21(1), 22.
- Yang, X., Steinberg, F., Zhuang, L., Bessey, R., & Trueb, B. (2016). Receptor FGFR1 does not promote cell proliferation but induces cell adhesion. *International journal of molecular medicine*, 38(1), 30–38.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13, 134.
- Yoshida-Kashikawa, M., Shibata, N., Takechi, K., & Agata, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Developmental dynamics : an official publication of the American Association of Anatomists*, 236(12), 3436–3450.
- Yoshioka, K., Oda, A., Notsu, C., Ohtsuka, T., Kawai, Y., Suzuki, S., Nakamura, T., Mabuchi, Y., Matsuzaki, Y., & Goitsuka, R. (2015). Loss of the Homeodomain Transcription Factor Prep1 Perturbs Adult Hematopoiesis in the Bone Marrow. *PLoS one*, 10(8), e0136107.
- Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics (Oxford, England)*, 31(14), 2382–2383.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36.
- Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science (New York, N.Y.)*, 309(5734), 626–630.
- Yuan, Q., & Duren, Z. (2024). Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nature Biotechnology*, 1–11.
- Zeng, A., Li, H., Guo, L., Gao, X., McKinney, S., Wang, Y., Yu, Z., Park, J., Semerad, C., Ross, E., Cheng, L. C., Davies, E., Lei, K., Wang, W., Perera, A., Hall, K., Peak, A., Box, A., & Sánchez Alvarado, A. (2018). Prospectively Isolated Tetraspanin⁺ Neoblasts Are Adult Pluripotent Stem Cells Underlying Planaria Regeneration. *Cell*, 173(7), 1593–1608.e20.
- Zhang, X., Friedman, A., Heaney, S., Purcell, P., & Maas, R. L. (2002). Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. *Genes & development*, 16(16), 2097–2107.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), R137.
- Zhang, D., Chan, J. D., Nogi, T., & Marchant, J. S. (2011). Opposing roles of voltage-gated Ca²⁺ channels in neuronal control of regenerative patterning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(44), 15983–15995.

Zhang, Y., & Beachy, P. A. (2023). Cellular and molecular mechanisms of Hedgehog signalling. *Nature reviews. Molecular cell biology*, 24(9), 668–687.

Zhu, S. J., Hallows, S. E., Currie, K. W., Xu, C., & Pearson, B. J. (2015). A mex3 homolog is required for differentiation during planarian stem cell lineage development. *eLife*, 4, e07025.