Understanding the epigenetic regulation of stem cell fate in planarians using functional genomics

Divya Lakshmi Sridhar

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

Acknowledgements

My thesis is dedicated to the millions of girls who have to fight for their right to an education.

I would like to thank my supervisor Aziz Aboobaker for his help and guidance. I also thank my group members over the years, especially Damian Kao, Prasad Abhave, Alessia Di Donfrancesco, Sounak Sahu, Anish Dattani, Salha Rawas, Jakke Neiro and Zhong Yap for their assistance, advice, and scientific discussions.

Thanks to all my friends – near and far – for all their love, support, help, advice, and laughter. Special thanks to Liisa Veerus, my friend, twin, and greatest supporter. You are my constant. Tejaswini Hombale, I am thankful for your friendship and support over the years. Thanks to Diego Berdeja Suárez and all my Mertonian friends, I am fortunate to have you. A special shoutout to my Indian academic family – Rimple D'Almeida, Alok Javali, Aalok Varma, Kavana Nadahalli, Radhika Rao, Aditya Deshpande, and Das lab members for all the stimulating scientific discussions we have had over the years. Thank you for making life fun inside and outside the lab. My time in inStem holds a special place in my heart and has helped shape me into the scientist I am today.

I would not be here today without the love and support of my family. Thank you to my parents, Sridhar Sundar Raj and Sujatha Sridhar, for being the best parents a girl could ask for. Thank you for believing in me, and supporting me even when things were not easy. Thank you to Shobha Sridhar, my amazing sister and friend, I admire you. Thank you to Maria Lock, for making Oxford into a place I can call home and being the most amazing mother-in-law. A special thank you to my late grandmother, Kannamma, for inspiring me. I have always wanted to grow up to be as knowledgeable and wise as you. I miss you.

Words cannot begin to describe how grateful I am for my love, my husband, Edwin Lock. You are my rock and I am lucky to have you by my side. Thank you for your patience, understanding, support and love. You are an excellent academic, great friend and teacher. I am glad to have found a perfect partner with whom I have discovered that the kitchen's a laboratory – filled with science, laughter, love, and magic! Thank you for being you.

Abstract

Planarians owe their remarkable regenerative capabilities to somatic pluripotent stem cells or neoblasts which, as the only dividing cells in the body, are able to differentiate into all cell types. Research into planarians so far has established the cellular and molecular basis of regeneration, stem cell pluripotency and differentiation. Recent studies using single-cell transcriptomics have improved our understanding of neoblast heterogeneity and the different cell types in planarians. Yet, the epigenetic mechanisms underlying neoblast pluripotency and the determination of cell fate remain understudied. This work aims to understand the epigenetic regulation of stem cell fate in planarians using an array of next-generation sequencing methods.

Planarians have a single methyl-CpG binding domain protein, MBD2/3, which is part of the NuRD complex that regulates transcription. Planarians lack endogenous cytosine methylation. Research into the role of MBD2/3 in an organism free of DNA methylation can provide important evolutionary insights into its DNA methylation-independent role. I show that knockdown of *Smed-mbd2/3* affects both regeneration and tissue homeostasis. Using markers for stem cells and progeny, *Smed-mbd2/3* was found to be essential for differentiation but not for stem cell maintenance. The different stem cell sub-populations were also unaffected by the knockdown of *Smed-mbd2/3*. To determine the targets of MBD2/3 involved in the lineage commitment of stem cells, I analysed the transcriptome of stem cells and stem cell progeny after knockdown of *Smed-mbd2/3* at two time points. Genes misregulated after the inhibition of *Smed-mbd2/3* were identified and will form an important resource to identify new markers of stem cell differentiation.

The identity of cis-regulatory elements remains elusive in planarians. Accessible regions of chromatin in the three planarian cell compartment were identified using ATAC-seq. ChIP-seq of planarian neoblasts was performed and correlated with gene expression and open regions of chromatin to identify putative enhancers. Information from the three types of sequencing datasets was used to create a catalogue of enhancers with different combinations of enhancer properties. This lays the groundwork for further studies to provide a better understanding of the heterogeneity of planarian stem cells and identify gene regulatory networks specific to the progression of different lineages.

Author contributions

Chapter 2: Divya Sridhar (DS) and Aziz Aboobaker (AA) conceived the idea. DS designed primers for all genes and cloned *mbd2/3*, *p66* and *hdac1*. Rosanna Larter, under supervision of DS, cloned *chd3*, *chd4*, *mta1-1* and *mta1-2* (as part of her undergraduate project). DS performed all experiments including knockdown and in situ hybridisations. DS performed confocal imaging, image analysis, cell counting and statistical analysis.

Chapter 3: DS and AA conceived the idea. DS performed all experiments including knockdowns and cell dissociation. DS carried out RNA extractions, preparation of mRNA libraries, quantification of libraries, sequencing, and post-sequencing data processing. Edwin Lock (EL) provided advice on data visualisation.

Chapter 4: DS, Anish Dattani (AD) and AA conceived the idea. DS and AD performed optimisation of transposition reaction and prepared X1 and Xins libraries. DS and AD performed quantification of the libraries and sequencing. DS performed cell dissociation, preparation of libraries, quantification of libraries and sequencing for X2 cells. DS performed all laboratory processing including cell dissociation, immuno-precipitation, ChIP-library preparation, quantification of libraries and sequencing. DS processed ATAC- and ChIP-seq data. EL provided advice with computational analysis.

Contents

List of Figures

1

Planarians – a model system to study the epigenetic regulation of stem cell fate

Contents

Regeneration is the capacity to re-grow lost body parts or repair cells, tissues, and organs. While this phenomenon is ubiquitous, regenerative capacity varies significantly between organisms. Mammals possess a lower regenerative capacity than invertebrates (e.g. hydra and planaria) and non-mammalian vertebrates (e.g. newt, axolotl, zebrafish)

[Zhao et al., 2016]. Different tissues and organs of the same organism often exhibit distinct degrees of regenerative capacity. Planarians have a remarkable capability of regenerating entire bodies from small fragments. Salamanders are capable of regenerating a wide array of tissues and organs, including limbs, tails, and lenses. Zebrafish can regrow hearts, lenses, spinal cords, and so forth. Unlike these animals, mammals have extremely limited regenerative capacity, which is restricted mainly to liver and bone regeneration. Severe damage to tissues or organs in mammals does not induce regenerative responses but rather wound healing with scarring [Zhao et al., 2016].

1.1 *Schmidtea mediterranea* – a model to study the epigenetic regulation of stem cell fate

Planarians are bilaterally symmetric, triploblastic, unsegmented animals that belong to the phylum Platyhelminthes, and super-phylum Lophotrochozoa.

Owing to their robust regenerative properties, stable diploid state ($2n = 8$) and relatively small genome size (~ 800 Mb; roughly a half of other common planarians), the free-living freshwater *S. mediterranea* has emerged as a good candidate for studying stem cells in vivo [Alvarado and Kang, 2005]. The extraordinary regenerative capacity of planarians can be attributed to the large population of adult stem cells, called neoblasts, that are distributed throughout the animal body. Neoblasts replace cells that are lost due to normal physiological turnover or injury. Planarians allow the study of stem cell heterogeneity and epidermal lineage progression from undifferentiated stem cells due to the availability of molecular markers for stem cells and their progeny [Eisenhoffer et al., 2008]. Identification of more lineage-specific markers is crucial for understanding the regulatory mechanisms involved in the determination of cellular fate and differentiation in planarians. An advantage of using *S. mediterranea* as a model organism for studying epigenetics is the availability of an excellent array of bioinformatics tools for this animal, which include genome assembly [Grohme et al., 2018], annotations [Cantarel et al., 2008], a genome database [Robb and Alvarado, 2014] and a transcriptome repository [Brandl et al., 2016]. Although recent reports using single-cell sequencing approaches in planarians have led to significant advances in

the understanding of planarian cell types [van Wolfswinkel et al., 2014, Fincher et al., 2018, Plass et al., 2018, Swapna et al., 2018], we are only beginning to understand the molecular principles governing the associated regulatory mechanisms. The absence of transgenic manipulation of neoblasts and the inability to maintain planarian stem cells in culture has slowed the rate of progress thus far.

1.2 Neoblasts and their role in regeneration

Neoblasts comprise a large proportion of all planarian cells ($\sim 25\%$) [Baguñà et al., 1989], and are broadly distributed throughout the body with the exception of the pharynx and the small region anterior to the eye spots [Reddien and Alvarado, 2004]. In unamputated planarians, neoblasts replace cells lost to normal physiological turnover [Newmark and Alvarado, 2000]. Upon injury, regeneration begins with rapid closing of a wounded surface, achieved by body wall muscle contraction and the subsequent protective spreading of existing epithelial cells. The neoblasts then divide rapidly throughout the body, with a peak in mitotic numbers 6 hours after injury. A second peak of neoblast proliferation is seen if the wound requires the replacement of missing tissue. This proliferation of neoblasts, which occurs at 48 hours, is concentrated at the wound site, and the neoblast progeny form an unpigmented bud of regenerated tissue called the blastema. The neoblast progeny then differentiate in a coordinated manner into the missing structures of the animal [Wenemoser and Reddien, 2010].

Sensitivity to radiation has made it possible to determine neoblast-specific genes by observing, using RNA-seq and in situ hybridisation techniques, whether expression patterns persist after radiation exposure. One limitation of this approach is the inability of excluding changes in gene expression induced by radiation that are not associated with neoblast loss [Blythe et al., 2010, Solana et al., 2012]. RNA interference (RNAi) of the neoblast-specific histone variant *Smed-histone-2B* (*Smed-H2B*) was found to genetically ablate neoblasts and not initially affect other tissues. The transcriptomic profiles of wild-type worms and *Smed-H2B* RNAi worms were then compared to identify genes specific to the neoblasts, circumventing the possibility of obtaining false positives [Solana et al., 2012]. In order to determine whether individual planarian stem cells were

capable of producing multiple cell types, a single-cell transplantation was performed [Wagner et al., 2011]. Following lethal doses of γ-irradiation, neoblast-deficient hosts were injected with a single planarian stem cell from a wild-type donor animal. Incredibly, in many cases, these individual neoblasts, termed clonogenic neoblasts ('cNeoblasts') fully re-populated the host, could restore tissue turnover and regenerative ability to irradiated animals, and were able to produce all required cell types, showing that at least some planarian stem cells are in fact pluripotent [Wagner et al., 2011].

Enrichment of the neoblast pool can be established using fluorescence-activated cell sorting (FACS). FACS analysis of animals enables the isolation of two cell populations that are lost in irradiated animals. The 'X1' gate contains cells with $> 2C$ DNA, consisting mainly of stem cells in the S/G2/M phase of the cell cycle, and the 'X2' gate contains cells with $< 2C$ DNA, mainly comprising stem cells in G1 phase and post-mitotic progenies. The 'Xins' gate contains irradiation-insensitive (Xins) cells, mainly post-mitotic differentiated cells with $2C$ DNA [Hayashi et al., 2006]. This methodology has been successfully used to isolate neoblasts, allowing for the identification of homologues of mammalian genes required for pluripotency and genes associated with the germline that have enriched expression in neoblasts [Önal et al., 2012]. An important caveat is that, unlike embryonic stem cells (ESCs), planarian neoblasts are a heterogeneous population.

1.3 Planarian neoblasts are a heterogeneous population

van Wolfswinkel et al. [2014] combined single-cell profiling with functional assays to provide evidence for the existence of three neoblast populations that make up the population of cells remaining within the cell cycle. All three groups express the pan stem cell marker *smedwi-1* transcript, orthologues of the PIWI family proteins associated with germ line maintenance. Sigma neoblasts (σ) were found to proliferate in response to injury and have broad differentiation potential consistent with the possibility that they are pluripotent. Sigma neoblasts give rise to lineage-committed subsets – zeta and gamma (Fig. 1.1). The three classes of neoblasts have high expression of a specific set of transcription factors. Zeta neoblasts (ζ) comprise specified cells that give rise

to post-mitotic lineages including epidermal cells, while gamma neoblasts (γ) express genes associated with the formation of the gut and endothelial lineages [Forsthoefel et al., 2012, Wagner et al., 2011]. Zeta and gamma neoblasts do not appear to have the capacity to self-renew [van Wolfswinkel et al., 2014].

Sigma neoblasts were not affected by knockdown of *zfp-1*. Indeed, neoblasts left after *zfp-1*(RNAi) were transplanted into irradiated (and thus neoblast-depleted) animals to show that surviving sigma neoblasts were capable of generating the zeta neoblasts [van Wolfswinkel et al., 2014]). Recently, an alternative approach was employed to provide independent evidence for defining which cycling cells can self-renew by inducing endoreplication in neoblasts. The study used condensin knockdown to find that only sigma neoblasts had increased DNA content and therefore attempted multiple rounds of cell division. This raises the question of whether zeta and gamma neoblasts should be considered stem cells at all and whether they have undiscovered heterogeneity that underpins their ability to contribute to multiple final cell types or whether they are uniform and have different cell type fates regulated without further mitoses [Lai et al., 2017].

Single-cell RNA sequencing (scRNA-seq) of stem cells and progeny from the planarian head revealed a fourth neoblast sub-population called nu neoblasts (νNeoblast) [Molinaro and Pearson, 2016]. A neoblast sub-population with distinct gene expression and neural character (νNeoblasts) was predicted by Molinaro and Pearson [2016]. Fluorescent in situ hybridisation using the predicted νNeoblast markers (*piwi-2*, *ston-2*, *elav-2*) was used to demonstrate that these neoblasts exist adjacent to the brain [Molinaro and Pearson, 2016]. However, little evidence currently exists to show that these are proliferative cells, as they have relatively low expression of known stem cell markers. These cells could be neural stem cells expressing *smedwi-2* with low *smedwi-1* expression.

Recent studies using droplet-based single-cell transcriptomics, Drop-Seq, has led to a better understanding of the different planarian cell types as well as uncovered new ones [Fincher et al., 2018, Plass et al., 2018, Swapna et al., 2018]. Furthermore, these studies indicate that the molecular markers for sigma neoblasts are also expressed in other lineage-committed progenitors, and expression of markers are not necessarily unique to cNeoblasts. Another study using scRNA-seq revealed 12 transcriptionally

distinct stem cell clusters [Zeng et al., 2018]. A novel sub-type of neoblast (Nb2) characterised by high levels of *smedwi-1* mRNA and protein and marked by *tetraspanin 1* (*tspan-1*), a conserved cell-surface protein-coding gene, was identified [Zeng et al., 2018]. An antibody against *tspan-1* allowed for pluripotent stem cells to be identified as well as purified from tissues. Single-cell transplantation of TSPAN-1$^+$ cells into lethally irradiated animals led to improved recovery efficiency compared to X1 cells [Zeng et al., 2018]. However, Zeng et al. [2018] do not address the efficacy of the 11 other neoblast clusters in single-cell transplantation, or whether they are truly pluripotent neoblasts. Nonetheless, the identification and isolation of pluripotent neoblasts provides opportunities for molecular investigations, genetic manipulation, as well as the potential establishment of individual neoblast cell lines. Together, these advances have now made it possible to study the role of epigenetic regulation in the maintenance of neoblast heterogeneity and lineage specification in planarians.

1.4 Epigenetic control of gene expression

While every cell in an organism theoretically contains the same genetic material, not all genes are active at all times. Instead they are expressed, in a highly controlled and coordinated fashion, only at times and places they are needed. This mechanism of gene regulation relies on transcriptional regulation (non-heritable) and epigenetic regulation (sometimes heritable). The term epigenetics has several meanings with independent roots. Conrad Waddington used the term for the study of epigenesis – how genotypes give rise to phenotypes during development [Waddington, 1942]. Later, epigenetics was defined as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" [Russo et al., 1996]. The third definition of epigenetics encompasses the study of the biology of chromatin, including how histone modifications and higher-order structure lead to regulatory effects. This epigenetic control is central to development, pluripotency, cell differentiation and protection against viral genomes. Epigenetic regulation involves DNA methylation, post-translational modifications of histones and small RNA mediated regulatory events, all of which are potentially heritable between cell divisions (the current work focuses

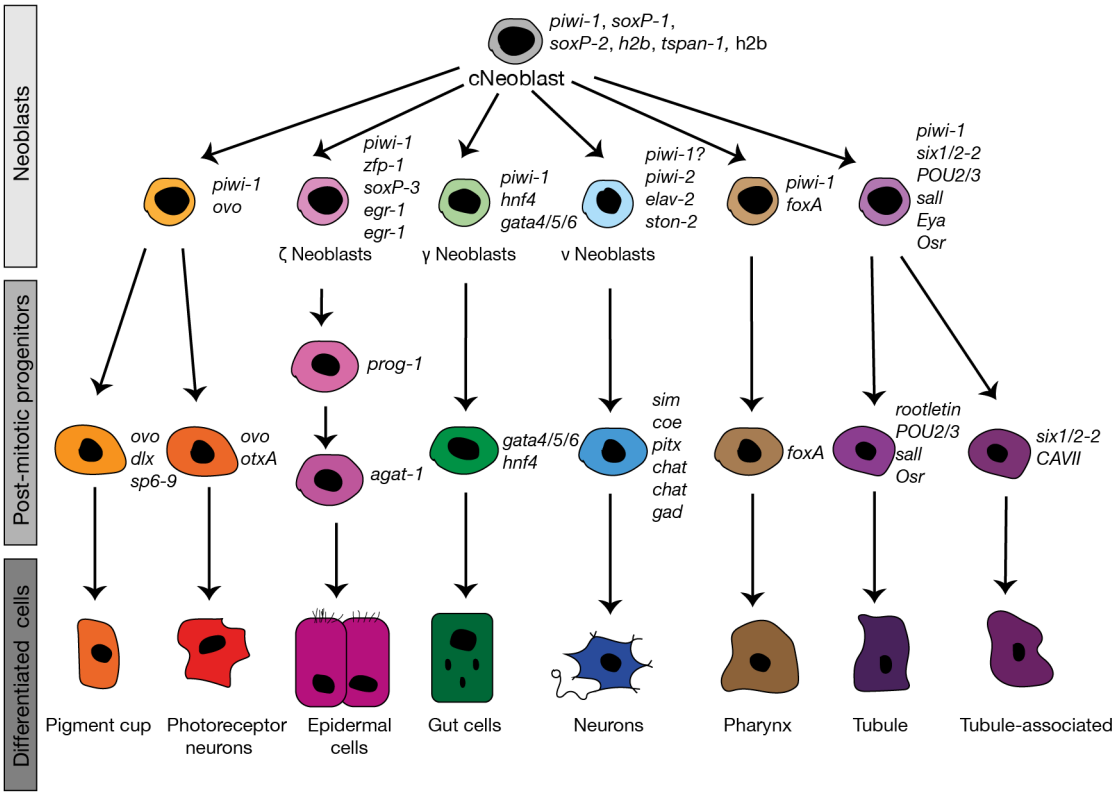

Figure 1.1: Heterogeneity and lineage commitment in planarian stem cells. Characteristic gene expression of each cell is noted alongside. The darker the colour, the more differentiated the cell. The figure is a simplified diagram with few lineages and is not an exhaustive list of all the markers identified for each cell/lineage.

on the first two.) Epigenetic modifications are closely linked and work in a concerted manner to regulate different cellular processes. The broad biological role of epigenetic mechanisms is the subject of intense study, and questions of how these mechanisms have evolved and played a major role in organismal diversity have recently become tractable, as technological advances make more species amenable to epigenetic study.

DNA methylation is an epigenetic mark involved with gene repression. DNA methylation involves the covalent transfer of a methyl group to the DNA base, usually cytosine (to form 5-methylcytosine or 5mC), by DNA methyltransferases (DNMTs) and typically occurs in CpG dinucleotides islands. Most CpG islands are sites of transcription initiation, located upstream of gene promoters, that attract repressive epigenetic modifying complexes which block transcriptional machinery from accessing these sites altogether.

DNA methylation does not function independently. DNA methylation and histone modifications work together to establish gene expression patterns. A striking feature of the core histones (H2A, H2B, H3 and H4) is the large number and variety of modified residues they possess and the patterns of histone marks, which can change rapidly. The presence of these modifications affects the higher order chromatin structure of DNA and influences the recruitment of different enzymatic complexes [Kouzarides, 2007]. These effector complexes act on chromatin in various ways to influence transcription levels as well as DNA replication, repair, recombination and condensation [Kouzarides, 2007].

1.5 Epigenetic regulation of stem cell proliferation and differentiation

Stem cells have the ability to self-renew and differentiate, generating more cells of individual or multiple lineages. Some stem cells exhibit pluripotency, allowing them to give rise to all types of cells in the adult animal, normally through the sequential production of less potent intermediates. The ability of pluripotent stem cells to maintain flexibility and then produce progeny that undergo lineage specifications is a result of both transcription factor-led gene regulatory programs as well as the overall epigenetic regulatory status across the genome. In pluripotent mouse embryonic stem cells (ESCs), for example, many promoters have been shown to contain a 'bivalent' histone modification signature of the activating H3K4me3 mark and the repressive H3K27me3 mark [Bernstein et al., 2006]. Bivalent regions are enriched in lineage specification genes encoding developmentally important transcription factors [Voigt et al., 2013]. The bivalent regions are thought to keep these developmental genes silent in ESCs while keeping them poised for activation when fate decisions are initiated. During lineage specification, the transcription of the appropriate cell type-specific genes begins, whereas the transcription of genes associated with alternative lineage fates is prevented.

Another example of the combined importance of transcriptional and epigenetic regulation is the reprogramming of differentiated cells to induced pluripotent stem cells (iPSCs) [Takahashi and Yamanaka, 2006]. This can be achieved by the forced

expression of transcription factors (OKSM, i.e. *Oct4*, *Sox2*, *Klf4*, *c-Myc* or other combinations) [Takahashi and Yamanaka, 2006]. This initiates a cascade of transcriptional and epigenetic changes (such as demethylation of OKSM promoters) that leads to cells treated this way becoming pluripotent [Yamanaka and Blau, 2010, Orkin and Hochedlinger, 2016]. Loss of *Oct4* function early in development leads to increased DNA methylation and closure of previously accessible chromatin and results in ESCs defaulting to a trophectodermal fate [Nichols et al., 1998].

Several epigenetic regulators have been shown to be essential for lineage specification in mouse ESCs, and are therefore crucial for accurate embryonic development [Luis et al., 2012, Morey et al., 2015]. However, the role of epigenetic regulation in the function of adult tissue stem cells is less understood. One of the challenges in this regard has been generating in vivo models to study the molecular mechanisms that regulate adult stem cells. Most studies carried out on the epigenetic regulation of stem cells utilise either ESCs or iPSCs. However, evidence suggests that long-term ESCs in culture may acquire epigenetic and transcriptional changes that are not reflective of their in vivo condition [O'Neill et al., 2006] and not all histone marks are efficiently reprogrammed in iPSCs [Hawkins et al., 2010, Bar-Nur et al., 2011]. For both cell types, experiments have demonstrated that they are pluripotent but many developmentally important genes are differentially epigenetically marked between iPSCs and ESCs [Van Den Hurk et al., 2016]. The use of other in vivo models to study epigenetics will not only add to our understanding of conserved mechanisms but will also help us gain an alternative perspective. Towards this goal, planarians offer themselves as a remarkable experimental model with pluripotent adult stem cells for in vivo studies that circumvent the common concerns associated with artefacts of cell culture conditions and provide an opportunity to study pluripotency in a distinct evolutionary context.

1.6 DNA methylation is patchy across invertebrates

DNA methylation is one of several epigenetic mechanisms involved in the control gene of expression. Alongside being a regulator of gene expression, cytosine methylation has also been implicated in the transcriptional silencing of transposons, genomic imprinting

and X-chromosome inactivation [Bird, 2002]. It also contributes to diseases like cancer when misregulated [Bird, 2002].

In general, vertebrate CpG methylation is highly prevalent in gene bodies. This global methylation correlates with transcriptional silencing of the DNA segment in question. CpG islands found upstream of gene promoters are the exceptions to genome-wide methylation [Li and Zhang, 2014]. DNA methylation varies from undetectable or very low levels to mosaic patterns of substantial methylation in different invertebrate genomes, although never reaching the high levels of global methylation of vertebrates, and the function of DNA methylation in these organisms remains unclear.

Alongside vertebrates, DNA methylation is abundant in algae, moss, plants, as well as in the Ciona, ants, honeybees, beetles and the crustacean *Parhyale* [Feng et al., 2010, Jeltsch, 2010, Zemach et al., 2010, Kao et al., 2016]. In model organisms, such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Schmidtea mediterranea*, cytosine DNA methylation is lost altogether [Jeltsch, 2010, Jaber-Hijazi et al., 2013] (Fig. 1.2). The patchy evolutionary distribution of DNA methylation has remained enigmatic despite considerable attention. DNA methylation has been lost numerous times over the course of evolution, including in lineages leading to *D. melanogaster* and *C. elegans* and *Oikopleura dioica* [Zemach and Zilberman, 2010]. DNA methylation is seen in crustaceans like *Parhyale hawaiensis* and *Daphnia pulex* but not in the crab *Cancer pagurus* [Regev et al., 1998, Strepetkaitė et al., 2015, Kao et al., 2016]. There is no correlation between repetitive sequences in invertebrate genomes and the percentage of methylation, which might have been the case if the main role of DNA methylation was to suppress genomic parasites [Regev et al., 1998]. How closely related invertebrates can differ in the amount of DNA methylation content across phylogeny is yet to be understood, and will require further studies in more closely related groups.

Methyl-CpG binding domain proteins (MBDs) recognise and bind to methylated CpG in DNA. This leads to the recruitment of chromatin remodelling and histone modifcation complexes to methylated promoter regions, thereby resulting in transcriptional silencing [Li and Zhang, 2014]. MBD2/3 and MBD4/MeCP2 are the ancestral group of MBD genes in animals which, following two rounds of whole genome duplication (2R), resulted

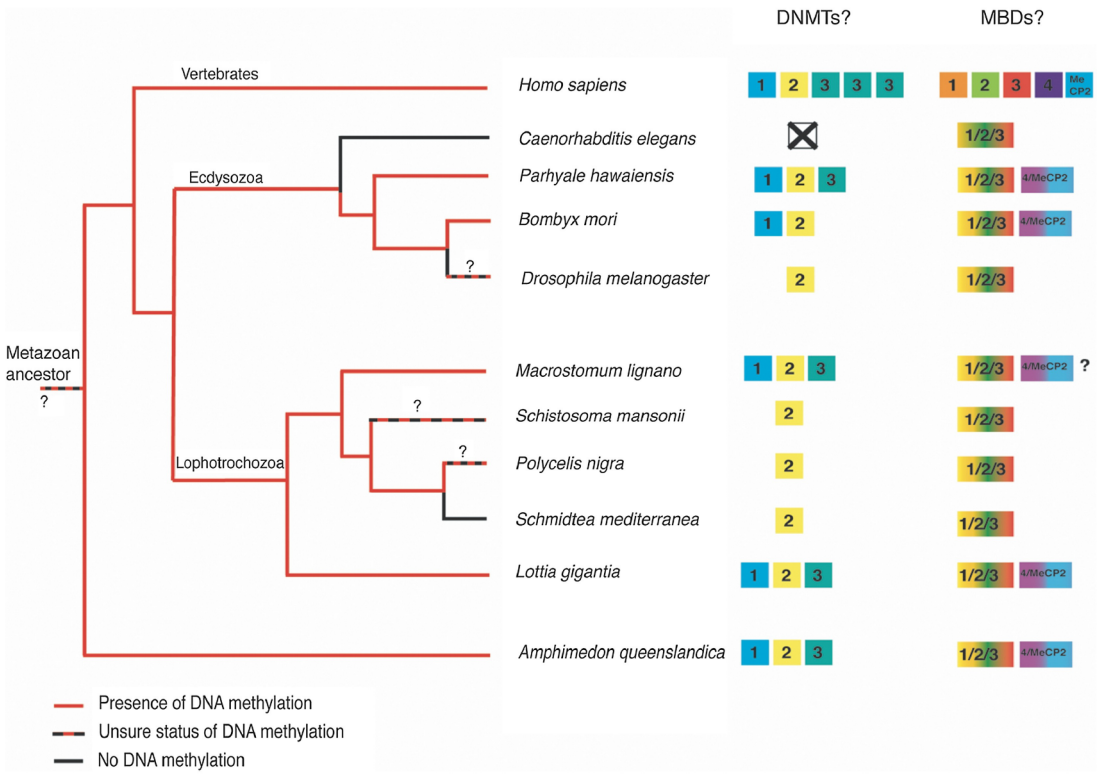

Figure 1.2: DNA methylation in flatworms and other species. The methylation status of the metazoan ancestor is unknown, but DNA methylation is not present in many invertebrates. DNA methylation has been lost several times in the course of invertebrate evolution with concomitant loss of genes encoding for one or both DNA methyltransferases (*Dnmt1* and *Dnmt3*). Alongside the diversity of DNMT and MBD genes, the extent of 5-methyl cytosine and its genomic distribution also differ among invertebrates. As many MBDs are found to lack the ability to bind methylated DNA, it is plausible that the ancestral role of MBD2/3 did not involve binding to methylated DNA. Figure from Dattani et al. [2018b].

in the paralogs MBD1, MBD2, MBD3, MBD4 and MeCP2 in vertebrates [Albalat, 2008, Albalat et al., 2012, Dattani et al., 2018b].

1.7 Lack of DNA methylation in planarians

DNA methylation studies in flatworms have been controversial. Early studies argued that DNA methylation was absent in the phylum Platyhelminthes on the basis of methylation-based restriction endonucleases followed by amplification of restriction fragments [Regev et al., 1998, Fantappie et al., 2001]. The basal Platyhemlminth *Macrostomum lignano* has low levels of DNA methylation [Wasik et al., 2015]. Parasitic flatworms were initially found to lack DNA methylation [Regev et al., 1998]. Surprisingly, Geyer et al.

[2011] suggested the existence of DNMT2-mediated cytosine methylation in the parasite *Schistosoma mansoni* and also believed they had demonstrated that DNA methylation in this organism plays a role in the regulation of oviposition. However, their drug-based experiments may have targeted DNMT2-mediated RNA methylation. Another study in Platyhelminthes by the same group revealed that representative species from all four Platyhelminth classes (Monogenea, Trematoda, Turbellaria and Cestoda) have only a DNMT2 orthologue and detectable levels of cytosine methylation [Geyer et al., 2013]. Using methylation-sensitive amplified polymorphism (MSAP), it was shown that *Polycelis nigra*, a Turbellarian species of flatworm closely related to *S. mediterranea*, had detectable levels of 5-methylcytosine [Geyer et al., 2013]. In contradiction to these findings, a comprehensive study utilising whole-genome bisulfite sequencing showed that the *S. mansoni* genome was not methylated and incompletely converted cytosines following bisulfite treatment were a likely source of artefactual calling of methylated bases [Raddatz et al., 2013]. *S. mediterranea* was found to have a conserved DNMT2 with no role in regeneration and a Smed-MBD2/3 protein sequence that does not have the conserved residues essential to contact methylated DNA [Jaber-Hijazi et al., 2013, Ohki et al., 2001]. Absence of 5-methylcytosine in *S. mediterranea* was also confirmed in various ways: by amplifying restriction fragments after digestion using methylation dependent restriction enzymes, the absence of antibody staining against 5-methylcytosine and undetectable levels of 5-methylcytosine in HPLC-MS [Jaber-Hijazi et al., 2013].

1.8 Methyl-CpG-binding domain proteins

DNA methylation is 'read' by methyl-CpG-binding domain (MBD) proteins. The MBD protein family consists of eleven MBD domain containing proteins (Fig. 1.3). In vertebrates, the core MBD proteins, MeCP2, MBD1, MBD2, MBD3, MBD4, are involved in transcriptional silecing [Hendrich and Bird, 1998, Jaenisch and Bird, 2003]. An additional 6 proteins with the MBD domain were found in the mammalian genome – MBD5, MBD6, BAZ2A (TIP5) and BAZ2B; SETDB1 and SETDB2 [Hendrich and Tweedie, 2003, Roloff et al., 2003]. In plants, 17 MBD proteins have been identified in

rice while 13 were found in *Arabidopsis thaliana*, respectively [Grafi et al., 2007]. On the other hand, invertebrates mostly contain one MBD gene, and in some cases two.

MBD proteins have a critical role in determining the transcriptional state of the genome [Du et al., 2015]. MBD proteins coordinate the complex interplay between DNA methylation, histone modifications and chromatin organisation to establish patterns of gene expression [Du et al., 2015]. MeCP2 was the first MBD-containing protein to be discovered and is characterised by the presence of a transcriptional repression domain (TRD) [Meehan et al., 1992]. MeCP2 is involved in heterochromatin formation and chromatin organisation [Nan et al., 1998, Fuks et al., 2003, Agarwal et al., 2007, Dhasarathy and Wade, 2008, Kernohan et al., 2014, Du et al., 2015]. MBD1 is involved in transcriptional repression and maintenance of heterochromatin [Fujita et al., 2003]. MBD1 mutations do not appear to cause any severe developmental defects [Zhao et al., 2003]. MBD2 and MBD3 proteins are paralogs and hence exhibit the greatest sequence similarity outside the MBD motif [Hendrich and Bird, 1998]. MBD2 can bind methylated DNA while MBD3 cannot, except in *Xenopus laevis*, due to a mutation in its MBD motif [Iwano et al., 2004, Saito and Ishikawa, 2002]. MBD2 and MBD3 associate with the nucleosome remodelling and deacetylation (NuRD) complex, in a mutually exclusive manner, and mediate transcriptional repression [Hendrich et al., 2001, Guezennec et al., 2006]. MBD2 and MBD3, together with NuRD, are involved in epigenetic regulation of cell fate [Xue et al., 1998, Zhang et al., 1999, Denslow and Wade, 2007]. MBD3 is shown to be indispensable for embryonic development, while MBD2-knockout mice are viable with mild phenotypes [Hendrich et al., 2001]. ESCs lacking MBD3 were viable, failed to downregulate progenitor genes, and fail to differentiate [Reynolds et al., 2012a]. MBD3 has been shown to be required for the formation of the NuRD complex in ESCs [Kaji et al., 2006]. MBD4 is a thymine glycosylase involved in base excision repair (involved in correcting G/T mismatches following deamination of 5-methylcytosine) [Hendrich et al., 1999]. MBD5 and MBD6 interact with the polycomb deubiquitinase complex, which catalyses deubiquitination of H2AK119 [Baymaz et al., 2014, Du et al., 2015]. MBD5 has high expression levels in the brain and oocytes while MBD5 and

MBD6 are highly expressed in the testes and their role in developmental functions requires further research [Laget et al., 2010, Du et al., 2015].

SETDB1/2 and BAZ2A/B are the only proteins in the MBD family that possess catalytic activity essential for modifying proteins without relying on protein partners [Du et al., 2015]. SETDB1 and SETDB2 contain the SET domain with H3-specific methyltransferase activity, whereas BAZ2A and BAZ2B contain the bromo-domain with acetylated histone binding capacity [Du et al., 2015]. SETDB1 associates with MBD1, and is involved in transcriptional repression and heterochromatin formation [Ichimura et al., 2005, Loyola et al., 2009, Du et al., 2015]. BAZ2A/TIP5 forms a part of the nucleolar remodelling complex that establishes ribosomal DNA silencing [Santoro and Grummt, 2005]. Recently, an association of BAZ2A with H3K14ac has been found to be essential for the de-differentiation of cancer cells into a cancer stem-like state in the prostate [Peña-Hernández et al., 2020]. SETDB2 and BAZ2A are under-studied and their functions are less known. SETDB2 is involved in chromosome condensation and segregation during mitosis [Falandry et al., 2010]. A recent bioinformatics study has shown a functional convergence between BAZ2B and genes known to cause neurodevelopmental disorders during foetal cortical development [Scott et al., 2020].

In invertebrates such as *D. melanogaster* [Roder et al., 2000], *C. elegans* [Gutierrez and Sommer, 2004] and *S. mediterranea* [Jaber-Hijazi et al., 2013], a single MBD2/3 gene has been identified. MBD2/3 is similar to its vertebrate MBD2 and MBD3 homologues and is considered to be ancestral [Hendrich and Tweedie, 2003]. In *D. melanogaster*, which lacks genomic 5-methylcytosine, MBD2/3 was found to interact with the Mi-2/NuRD complex to mediate transcriptional repression [Marhold et al., 2004].

1.9 The multi-subunit NuRD complex

Nucleosome positioning and chromatin remodelling can determine the accessibility of regulatory proteins interacting with DNA [Becker and Workman, 2013]. The nucleosome remodelling and deacetylation (NuRD) complex is a transcriptional repressor essential for embryonic development and ESC function [Mcdonel et al., 2009]. The NuRD complex has at least three distinct enzymatic activities – chromatin remodelling, deacetylase and

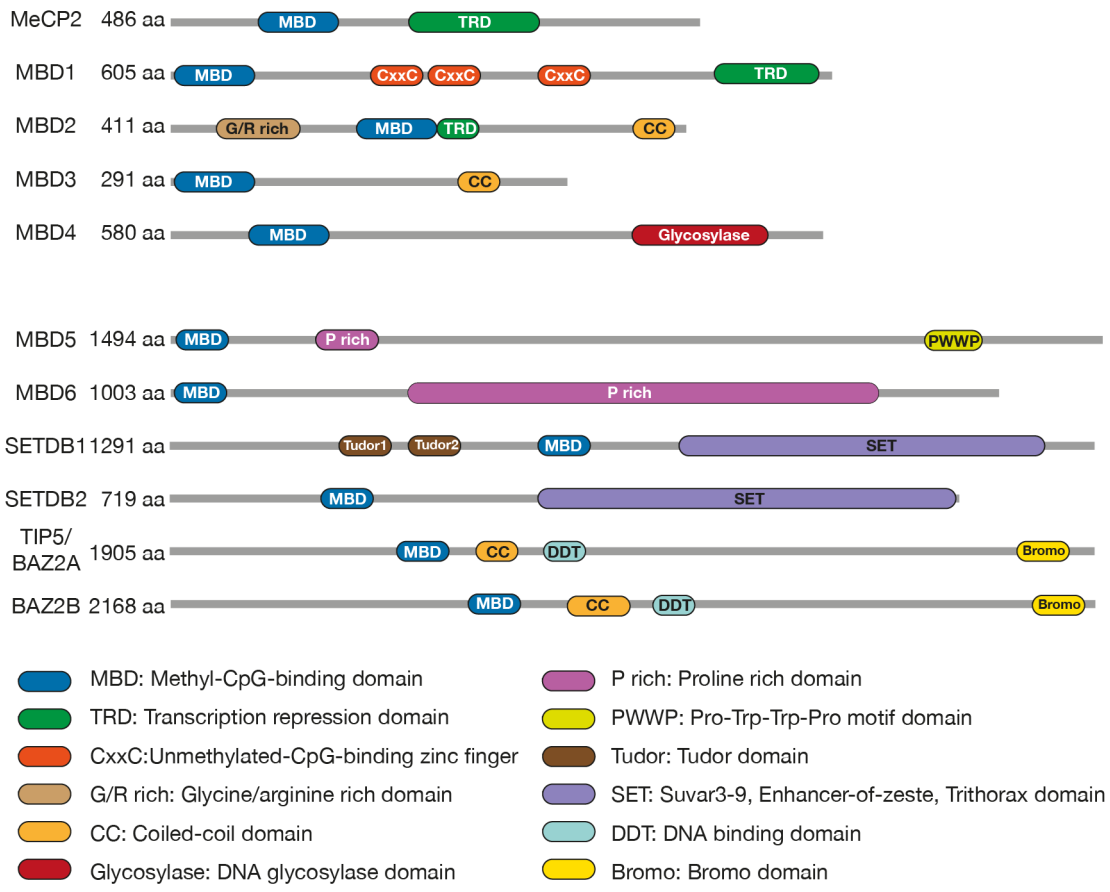

Figure 1.3: Methyl-CpG-binding domain protein family. MBD family proteins owe their function to the combined activity of their DNA binding domains and catalytic activity of protein partners. Core MBD proteins (MeCP2 and MBD1–4) are similar in size. Figure adapted from Du et al. [2015].

lysine-specific demethylation. Chromodomain-helicase-DNA-binding proteins paralogues CHD3 (Mi-2α), CHD4(Mi-2β) and CHD5 have ATP-dependent chromatin remodelling activity while histone deacetylases, HDAC1 and HDAC2, catalyse the removal of acetyl groups from amino acids of histones [Lai and Wade, 2011]. Lysine-specific histone demethylase 1A (LSD1) has also been shown to be associated with the NuRD complex and is involved in the demethylation of lysine 4 of histone 3 (H3K4) [Wang et al., 2009]. The non-enzymatic subunits of the NuRD complex include MBD (MBD2 and MBD3), metastasis-associated genes (MTA1, MTA2 and MTA3), histone chaperones or retinoblastoma-binding proteins (RBBP4/RbAp48 and RBBP7/RbAp46), and a GATA-type zinc finger domain protein (GATAD2A/p66α and GATAD2B/p66β) [Xue et al., 1998, Lai and Wade, 2011, Dattani et al., 2018b]. The GATAD2 subunit is likely

involved in sequence-specific DNA binding to target the NuRD complex to specific loci [Brackertz et al., 2002, Feng et al., 2002]. The MBD subunits target the NuRD complex to different genomic loci by associating with methylated DNA [Hendrich and Bird, 1998] while MTA subunits achieve this by associating with transcription factors [Fujita et al., 2003, Lai and Wade, 2011].

1.10 Roles of the NuRD complex in cell differentiation

The different enzymatic components of the NuRD complex have been shown to be crucial for developmental processes among eukaryotes. The ATP-dependent chromatin remodelling activity of NuRD is a result of the mutually-exclusive chromodomain helicase DNA binding paralogous subunits CHD3 (Mi-2α), CHD4 (Mi-2β) and CHD5. CHD3/4/5 utilise the energy released from the ATP hydrolysis to induce nucleosome sliding which either enables the recruitment of transcriptional complexes or suppresses transcription entirely, thereby modulating gene expression. CHD4 has been shown to associate with histone acetyltransferases and methyl transferases, resulting in transcriptional activation [Denslow and Wade, 2007]. Early studies in *A. thaliana* and *C. elegans* indicated that CHD subunits are involved in the silencing of embryonic genes during differentiation. In *A. thaliana*, the CHD3 homologue PKL (PICKLE) is required to transition from embryonic to post-embryonic development. PKL is essential for the repression of LEC1, a transcription factor that promotes embryonic identity [Ogas et al., 1999]. In *C. elegans*, Mi-2 homologues, LET-418 and CHD3 play essential roles in vulval cell fate determination [von Zelewsky et al., 2000]. More recent studies in mammalian systems have since shown that CHD guides lineage-specific gene transcription. For example, the CHD3/4/5 proteins regulate distinct genes essential for brain development as well as distinct and non-redundant aspects of mouse embryonic cortical differentiation [Nitarska et al., 2016]. CHD4 also is required for maintenance and multi-lineage differentiation in early hematopoiesis [Yoshida et al., 2008, Ng et al., 2009].

NuRD has histone deacetylase activity through the activities of the HDAC subunits. HDAC1 and HDAC2 remove acetyl groups from lysine residues on histones and other

proteins. HDAC also forms a part of other repressor complexes and is highly conserved and present in all eukaryotes. Histone deacetylation by the NuRD complex has been shown to specify recruitment of polycomb repressive complex 2 (PRC2) in ESCs. NuRD-mediated deacetylation of H3K27 provides a substrate for PRC2-mediated trimethylation of H3K27 at NuRD target promoters, leading to transcriptional silencing [Reynolds et al., 2012b]. NuRD/PRC2 targets in ESCs are bivalent genes and NuRD targets in ESCs include developmental genes as well as genes associated with pluripotency [Reynolds et al., 2012b]. Histone acetylation status of genes are regulated by antagonists HDACs and histone acetyltransferases (HATs) [Qiao et al., 2015]. HDACs and HATs are involved in multiple cellular pathways and are essential to pluripotency and differentiation of ESCs [Qiao et al., 2015]. Loss of HDAC1 leads to embryonic lethality in mouse and HDAC1-null ESCs have proliferation defects [Lagger et al., 2002].

The NuRD complex is also associated with LSD1, which demethylates H3K4 and impacts the chromatin configuration governing transcription regulation [Wang et al., 2009]. LSD1 is also associated with various co-repressor complexes including CoREST, CtBP, and a subset of HDAC complexes [Wang et al., 2009]. LSD1/NuRD complexes target promoters of genes that constitute signalling pathways, including TGFβ, which is relevant in cell growth, survival, migration, and invasion and has been shown to suppress breast cancer metastasis [Wang et al., 2009]. In ESCs, the LSD1/NuRD complex is shown to be required for silencing enhancers during differentiation and for the transition to new cell states [Whyte et al., 2014]. Therefore, LSD1 demethylates H3K4me1 at the active enhancers of ESC-specific genes during differentiation which leads to the silencing of the genes associated with these enhancers.

1.11 MBD and NuRD complex in methylation-free planarians

A single Smed-MBD2/3 protein without the full complement of conserved residues required to bind to methylated DNA was found in planarians [Jaber-Hijazi et al., 2013]. Knockdown of *Smed-mbd2/3* resulted in the loss of differentiated cell lineages (e.g. epidermis, gut and pharynx) without reducing stem cell numbers [Jaber-Hijazi et al., 2013].

Apart from MBD2/3, four other NuRD complex components have been studied in planarians. Planarian HDAC-1 was found to be expressed in neoblasts and knockdown of *hdac1* led to failure in stem cell maintenance [Eisenhoffer et al., 2008, Zhu and Pearson, 2013]. Smed-CHD4 is necessary for neoblast maintenance and differentiation during regeneration and homeostasis [Scimone et al., 2010]. The *Dugesia japonica RbAp48* homologue also plays a role in neoblast self-renewal and differentiation function [Bonuccelli et al., 2010]. It is known that Mi-2/CHD4 and *RbAp46* are involved in other chromatin-modifying complexes such as Sin3, CAFU1 and dMec [Kunert et al., 2009, Mcdonel et al., 2009], so their roles in planarian stem cell maintenance may also be associated with functions within these complexes [Aboobaker, 2011]. The involvement of the proteins in stem cell differentiation, however, appears to be related to their participation in the NuRD complex [Aboobaker, 2011]. This view is supported by the observation that the only gene exclusive to the NuRD complex, *Smed-mbd2/3*(RNAi), led to perturbed neoblast differentiation but intact neoblast self-renewal following knockdown [Jaber-Hijazi et al., 2013]. The GATA-type zinc-finger domain-containing gene *p66* acts to suppress photoreceptor neurons production in wild-type worms [Vásquez-Doorman and Petersen, 2016]. Knockdown of *Smed-p66* leads to an increase in photoreceptor neurons, but does not affect eye pigment cup cell production. A similar effect on the precursors of epidermal lineage is seen after the loss of *Smed-mbd2/3*. A detailed study of the *p66* and *mbd2/3* phenotype might reveal whether the two are have similar function and/or targets.

1.12 Was MBD2/3 originally meant to bind to methylated DNA?

DNA methylation has been lost several times in the course of invertebrate evolution, with concomitant loss of genes encoding for one or both DNA methyltransferases (*Dnmt1* and *Dnmt3*). Alongside the diversity of DNMT and MBD genes, the extent of 5-methylcytosine and its genomic distribution also differ among invertebrates. In addition to the MBD domain, the MBD family of proteins also contain several other domains, pertaining to their specific functional roles [Du et al., 2015]. Some proteins

contain a transcriptional repression domain that mediates interactions with protein partners [Wade et al., 1999, Boeke et al., 2000, Ng et al., 2000], while others confer enzymatic activities. While MBD family proteins have overlapping functions in the epigenetic modulation of the genome, they also perform different functions and exhibit distinct DNA binding specificities [Du et al., 2015].

In vertebrates, MBD2 and MBD3 interchangeably associate the NuRD complex. MBD2 has the capacity to selectively recognise 5-methyl cytosine (5mC), whilst MBD3 has lost the ability to bind to 5mC during vertebrate evolution [Hendrich and Bird, 1998, Zhang et al., 1999]. Although earlier studies suggested that MBD3/NuRD had a role independent of DNA methylation, MBD3 can bind to 5-hydroxymethylcytosine (5hmC), a product of the oxidation of carbon-5 methyl group in 5mC into a hydroxyl residue mediated by TET1 (ten–eleven translocation 1) [Tahiliani et al., 2009, Lu et al., 2015]. In a recent study, Hainer et al. [2016] showed that MBD2 and MBD3 have overlapping targets that are also regulated by DNA methylation/hydroxymethylation proteins. The authors propose a model in which the two MBD proteins are interdependent and, together with DNA modifications, form a regulatory loop to modulate gene expression (Fig. 1.4a) [Hainer et al., 2016]. Although the mechanistic details of this complex interplay are unclear, physical interactions between DNA methylation/hydroxymethylation factors and MBD3 have been reported [Yildirim et al., 2011, Cai et al., 2014, Hainer et al., 2016]. The preferential binding of MBD2 and MBD3 to methylated and hydroxymethylated DNA, respectively, has been a subject of dispute. Several reports have suggested that MBD2/NuRD and MBD3/NuRD play a role in the transcriptional activation of genes and enhancers, and may function independently of CpG methylation in mammals [Baubec et al., 2013, Shimbo et al., 2013, Günther et al., 2013, Menafra and Stunnenberg, 2014]. However, data from these studies were re-analysed and no evidence of the methylation-independent functions of MBD2 and MBD3 was found [Hainer et al., 2016].

A few MBD proteins also have RNA-binding activity, including MeCP2, MBD2 and BAZ2A. MeCP2 and MBD2 have been shown to form RNA–protein complexes in vitro and BAZ2A binds to noncoding RNA, directing the BAZ2A-NoRC complex to

target rDNA loci for heterochromatin formation [Jeffery and Nakielny, 2004, Mayer et al., 2006, Du et al., 2015].

Most invertebrates have an ancestral MBD2/3 gene that, upon 2R, resulted in one MBD2 gene and one or two copies of the MBD3 gene in vertebrates [Hendrich and Tweedie, 2003]. Apart from MBD3, all core mammalian MBDs have been shown to preferentially bind to methylated CpG in vitro with different affinities [Hendrich and Tweedie, 2003]. Human MBD3 has been shown to bind to HDAC1 and MTA2, two other components of the NuRD/Mi2 complex. It is possible that MBD3 has evolutionarily conserved its MBD domain because of the secondary role played by the domain in protein-protein interactions [Saito and Ishikawa, 2002]. In invertebrates like *D. melanogaster* and *C. elegans*, MBD2/3 has been shown not to bind methylated DNA and in other organisms, like *S. mediterranea*, it probably cannot due the absence of correct amino acid sequences required for binding methylated DNA. Yet MBD2/3 continues to associate with the NuRD complex [Jaber-Hijazi et al., 2013, Marhold et al., 2004, Cramer et al., 2017]. The existence of DNA methylation has been confirmed in red flour beetle *Tribolium castaneum* [Song et al., 2017] but recently it was shown that TcMBD2/3 does not possess binding activity to methylated DNA [Song et al., 2020]. The MBD2/3 protein of sponges, one of the earliest branching extant metazoans, has been shown to selectively bind to methylated DNA consistent with the presence of DNA methylation in the species [Cramer et al., 2017]. In addition to MBD2/3, MBD4 has been identified in some invertebrates like sponges, cnidarians, lophotrochozoans, parhyale, tunicates and Amphioxus [Albalat et al., 2012, Kao et al., 2016]. These are animals that also have DNA methylation and MBD4 is likely another ancestral methyl binding protein [Albalat et al., 2012, Kao et al., 2016].

Only five of the 13 MBD proteins identified in *A. thaliana* have been reported to bind methylated DNA [Scebba et al., 2003, Grafi et al., 2007]. Other mammalian proteins in the MBD protein family have not been shown to stably interact with methylated DNA in vitro or in vivo, even though they have the MBD domain. These include MBD5, MBD6, SETDB1 and BAZ2A [Strohner et al., 2001, Laget et al., 2010, Du et al., 2015].

Orthologues of these proteins exist in several methylation-free species and some have RNA-binding activity [Jeffery and Nakielny, 2004, Mayer et al., 2006].

Given the association of MBD2/3 with DNA methylation in basal metazoans, the ancestral MBD2/3 is likely to have had 5mC or 5hmC binding activity, which has since been lost in non-methylated invertebrates. Moreover, it is probable that the role of MBD2/3 in these species is independent of DNA methylation and may not use DNA methylation as a genomic reference to recruit the NuRD complex to target loci. It has been established that *S. mediterranea* has no detectable levels of endogenous cytosine methylation [Jaber-Hijazi et al., 2013] and it remains to be addressed whether *Smed-mbd2/3* forms part of the planarian NuRD complex. In order to resolve this long-standing question of whether ancestral MBD2/3 plays a part in coordinating NuRD activity independently of DNA methylation, it may be of value to investigate the function of MBD2/3 in *S. mediterranea* and whether the mechanism has been conserved in mammals [Yildirim et al., 2011, Baubec et al., 2013, Shimbo et al., 2013, Hainer et al., 2016]. In DNA methylation-free species, it is possible that MBD2/3 directly, or indirectly via an unknown DNA-binding protein partner, associates with target genes related to pluripotency or differentiation, thereby recruiting NuRD to regulate their transcription in stem cells and progeny (Fig. 1.4b-c).

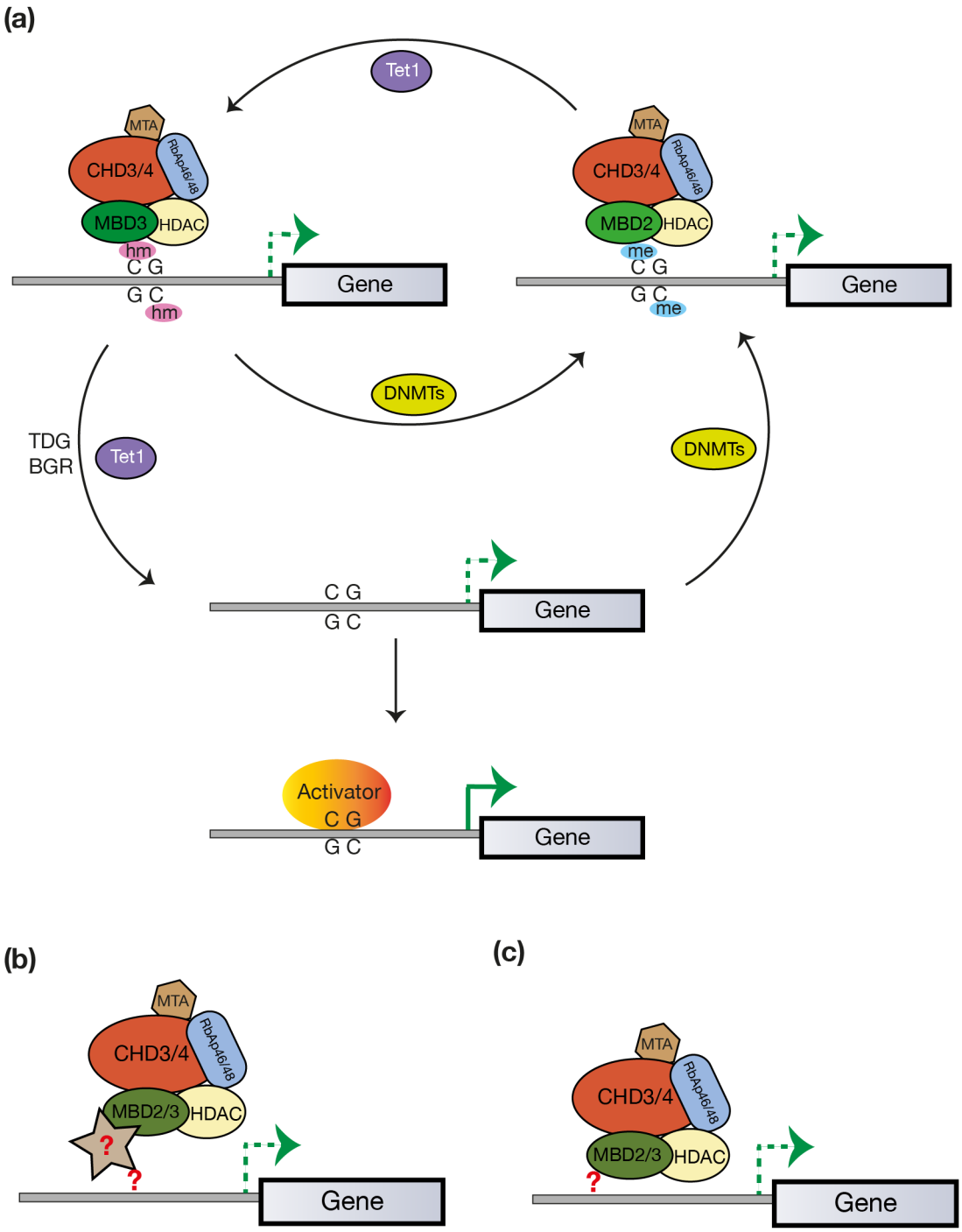

Figure 1.4: Proposed model for the regulatory mechanism of MBDs. (a) Role of MBDs and DNA methylation machinery in the establishment of a regulatory loop in vertebrates adapted from Hainer et al. [2016]. In methylation-free invertebrates, MBD2/3-NuRD could either bind to target genes (b) via an (unknown) DNA binding protein or (c) directly, leading to dampening of transcription or silencing of the gene.

2

Characterisation of *mbd2/3* in planarians

Contents

2.1 Introduction

Pluripotent stem cells undergo extensive self-renewal and have the ability to differentiate into all cells lineages of the body. Lineage progression of stem cells requires long-lasting changes in gene expression, including the repression of stem cell pluripotency maintenance genes. Epigenetic mechanisms, including DNA methylation, histone modifications and nucleosome remodelling, play a critical role in the dynamic regulation of stem cell renewal and differentiation.

The epigenetic regulation of stem cell pluripotency and differentiation has been explored primarily through research using embryonic stem cells (ESCs). ESCs are derived from the inner cell mass of the blastocyst stage of mammalian embryos. They are characterised by pluripotency, the ability to differentiate toward all cell types of the embryo and adult, and their potential for self-renewal. Research into the genes that are expressed in ESCs has led to the identification of cell pluripotency-associated genes and the discovery that transcription factors *Oct4*, *Sox2* and *Nanog* regulate pluripotency [Boyer et al., 2005, Young et al., 2011]. Nowadays, pluripotent stem cell lines are routinely established by the introduction and overexpression of these key transcription factors. The resulting induced pluripotent stem cells (iPSCs) have been studied extensively and have helped reveal the molecular interplay of key factors during de-differentiation [Takahashi and Yamanaka, 2006]. Alternatively, stem cells can be established by inducing the expression of certain transcription factors or micro RNAs that artificially convert differentiated cells from one committed lineage to another [Song et al., 2012, Nam et al., 2013]. Cell reprogramming involves redefining the identity of a cell by altering its transcriptional and epigenetic landscapes [Pereira et al., 2012]. Genome-wide transcriptomic and epigenetic profiling has been imperative in understanding the epigenetic regulation of gene expression, cellular identity, and nuclear reprogramming.

The molecular cues guiding the epigenetic control of stem cell pluripotency and cellular differentiation in planaria remain largely unknown. In this chapter, I aim to understand the biology of neoblasts in *S. mediterranea* by studying various epigenetic regulators required for stem cell proliferation and differentiation. My primary focus is on the NuRD complex, which has been the subject of previous research in *S. mediterranea*. In my work, I aim to characterise genes that have not been studied and independently confirm the findings of previous studies. The components of NuRD that are unique to the complex, MBD2/3 and p66 (GATA2D), are studied in greater detail in this chapter.

Jaber-Hijazi et al. [2013] have shown the absence of cytosine methylation in the genome of *S. mediterranea* based on the absence of antibody staining against 5-methylcytosine, undetectable levels of 5-methylcytosine in HPLC-MS and evidence from methylation-sensitive amplification polymorphism. Only two genes previously associated with catalytic addition or recognition of methylated cytosine in nucleic acids – *Smed-mbd2/3* and *Smed-dnmt2* – were identified in the planarian genome [Jaber-Hijazi et al., 2013]. The *Smed-mbd2/3* gene was shown to be required for neoblast differentiation during regeneration and tissue homeostasis but *Smed-dnmt2* was found to have no apparent role in regeneration [Jaber-Hijazi et al., 2013]. Additionally, it was found that MBD2/3 in *S. mediterranea* does not contain the highly conserved arginine (position 17) involved in forming hydrogen bonds with guanine in methylated CpG islands. Furthermore, neither DNMT1 nor DNMT3 have been found in the genome of *S. mediterranea* and it does not contain MBD4/MeCP2 like other closely related Platyhelminthes [Dattani et al., 2018b]. As a consequence of these multiple lines of evidence, it is likely that the function of *Smed-mbd2/3* is independent of DNA methylation and that DNA methylation is not involved in the epigenetic control of stem cells in *S. mediterranea*.

This hypothesis may resolve the long-standing question of whether the ancestral MBD2/3 plays a part in coordinating NuRD activity independent of DNA methylation; it will be interesting to investigate whether this mechanism has been conserved in mammals [Baubec et al., 2013, Hainer et al., 2016, Shimbo et al., 2013, Yildirim et al., 2011]. In DNA methylation-free species such as planarians, it is possible that MBD2/3

directly, or indirectly via an unknown DNA-binding protein partner, associates with target genes related to pluripotency or differentiation, thereby recruiting NuRD to regulate their transcription in stem cells and progeny. Investigating the role of MBD2/3 in planarians, a DNA methylation-free organism, has an important evolutionary significance and may help understand the DNA-methylation independent role of MBD2/3. In this chapter, I present a characterisation of *Smed-mbd2/3*, along with other proteins associated with the NuRD complex.

2.2 Results and discussion

The functions of the NuRD complex components were investigated in planarians using RNA interference during regeneration. The functions of *Smed-mbd2/3*, *Smed-p66* and *Smed-hdac1* were investigated in detail during tissue homeostasis and the molecular defects were characterised by labelling different neoblasts and epidermal progeny cells.

Tissue homeostasis and regeneration are impaired after the knockdown of *Smed-mbd2/3*

To assess the role of *Smed-mbd2/3* in tissue maintenance during physiological turnover, worms were left unamputated for phenotypic observation after RNAi injections. *Smed-mbd2/3*(RNAi) animals exhibited anterior regression, starting with the anterior tip, after two weeks. Typically, a notch appeared at the middle anterior tip of the head of the animals, often accompanied by a dark cross near the eyes. After the third week, more pronounced head regression was observed, where some worms had lost parts of the brain or the entire head structure, while the control animals looked normal (Fig. 2.1a). The control animals showed no defects in tissue homeostasis, decreased in size and all survived up to day 27 (4 weeks), while *Smed-mbd2/3* animals had ultimately lysed by this time point. A similar phenotype was also observed by Jaber-Hijazi et al. [2013], although the animals that had regressed to tail stumps started dying at 4 weeks of homeostasis. In contrast, I observed a much stronger phenotype where animal death started after 3 weeks. In situ hybridisation of known cell-specific markers was performed to determine the defects at a molecular level.

In order to ascertain whether *Smed-mbd2/3* is required for planarian regeneration, I performed knockdown of *Smed-mbd2/3*. Animals were injected 6 times with dsRNA corresponding to *Smed-mbd2/3* or *gfp* (as control). In each case, 10 animals were injected and then amputated into 3 pieces – head, trunk, and tail. In the regeneration assay, phenotypic defects due to knockdown started becoming apparent at later regeneration stages. Three regeneration time points (day 4, day 7 and day 10) are shown in Fig. 2.1b for comparison. During early regeneration, all control animals formed a blastema while *Smed-mbd2/3*(RNAi) animals formed smaller blastema or no posterior blastema at all. At regeneration day 7, however, the *Smed-mbd2/3* knockdown animals were unable to replace missing structures. The head pieces showed signs of curling, head regression and slower mobility. The trunk and middle pieces were unable to form eyes and other missing tissues. At day 10 of regeneration, defects in *Smed-mbd2/3*(RNAi) animals were more discernible. None of the trunk and tail pieces formed eyes. The head pieces were unable to form the pharynx or tail and showed head regression, including loss of brain and eye tissue. Generally, the slow motility phenotype increased in severity with the progression of regeneration – from low to complete immobility. In contrast, control animals had regenerated all missing structures including eyes, tail and pharynx without any visible defects. These results were broadly similar to those previously described in the literature, although the phenotypes I observed appear to be slightly stronger and pertain to the asexual strain of *S. mediterranea*. After 7 days, I observed head regression in most head pieces and most of these pieces had no posterior blastema/unpigmented regions. In Jaber-Hijazi et al. [2013], animals unable to regenerate missing structures had an unpigmented blastema after 7 days. I observed a stronger and consistent phenotype during both homeostasis and regeneration that is likely attributable to incremental improvements in RNAi over the years.

Consequently, I conclude that knockdown of *Smed-mbd2/3* by RNAi results in defects during both homeostasis and upon amputation, with loss of viability amongst all worms (Fig. 2.1c-d). The observation that *Smed-mbd2/3* knockdown regenerating animals were able to form a blastema that later did not increase in size indicates that the initial

wound healing response is intact, as is the production of early post-mitotic progeny. The problem might lie in the growth and differentiation of this blastema into missing tissues.

Defects caused by knockdown of *Smed-mbd2/3* are not due to the failure in neoblast maintenance

Head regression in planarians is typically observed due to impaired neoblast function [Reddien et al., 2005a,b, Guo et al., 2006]; however, this "no neoblast" or loss of neoblasts phenotype is usually accompanied by ventral curling. Although *Smed-mbd2/3*(RNAi) animals showed head regression, ventral curling was not common. In order to ascertain if neoblast loss is the underlying cause of defects during regeneration and homeostasis in *Smed-mbd2/3*(RNAi) worms, fluorescent whole mount in situ hybridisation was performed for the stem cell marker *smedwi-1*. There was no obvious difference in *smedwi-1*$^+$ cell numbers and localisation between control and *mbd2/3*(RNAi) animals. In order to confirm this, the number of *smedwi-1*$^+$ cells was counted in the pre-pharyngeal region and normalised to the area (Fig. 2.2a-b). The number of *smedwi-1*$^+$ cells did not differ significantly between *Smed-mbd2/3*(RNAi) and controls after 7 and 14 days (Fig. 2.2c).

Previously, Jaber-Hijazi et al. [2013] assessed neoblast markers *smedwi-2* and *Smed-H2B*, which are both expressed in proliferating stem cells. They observed no effect on the distribution, proliferation, or relative numbers of neoblasts even after prolonged knockdown (an extra round of RNAi injections). Together, the data suggest that *Smed-mbd2/3* is not essential for neoblast maintenance and proliferation.

As the total neoblast populations were unperturbed after the loss of *Smed-mbd2/3*, I sought to establish whether different lineage committed neoblasts were affected after knockdown. Neoblasts are a heterogeneous population in terms of the level of gene expression and the level of potency. Single-cell transcriptomic profiling revealed a stem cell population, called sigma, that gives rise to zeta and gamma stem cells. These three classes are enriched for expression for specific transcription factors [van Wolfswinkel et al., 2014]. During the previous work, markers of neoblast heterogeneity were unavailable. In order to ascertain whether different neoblast sub-populations

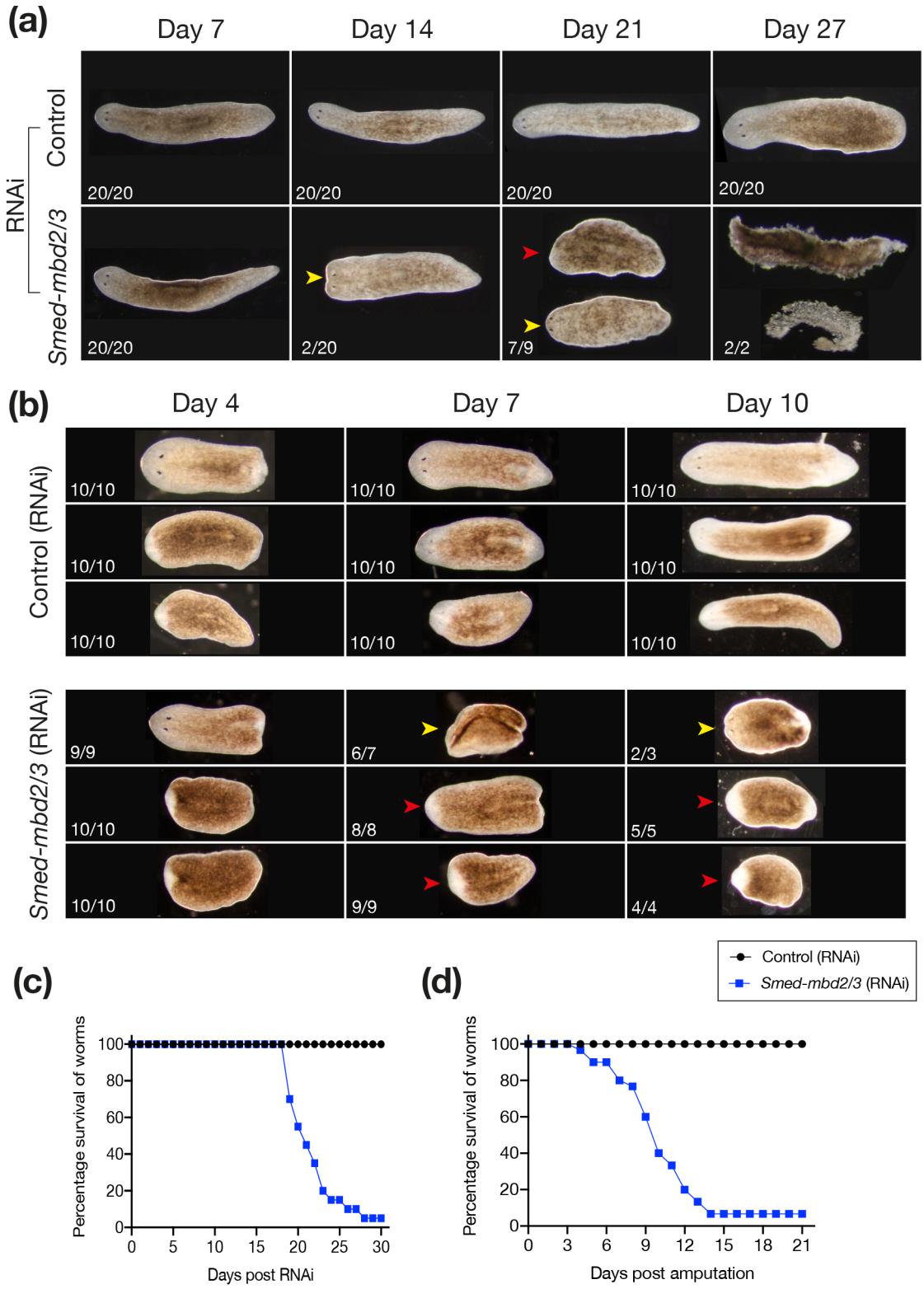

Figure 2.1: *Smed-mbd2/3* is essential for tissue homeostasis and regeneration. (a) Intact animals during homeostasis after *Smed-mbd2/3* and control RNAi. (b) Head, middle and tail pieces during regeneration after *Smed-mbd2/3* and control RNAi. Survival curves of knockdown animals during (c) tissue homeostasis and (d) regeneration. White numbers indicate proportion of animals exhibiting the same phenotype at this time point. Yellow and red arrowheads respectively indicate head regression and no eyes.

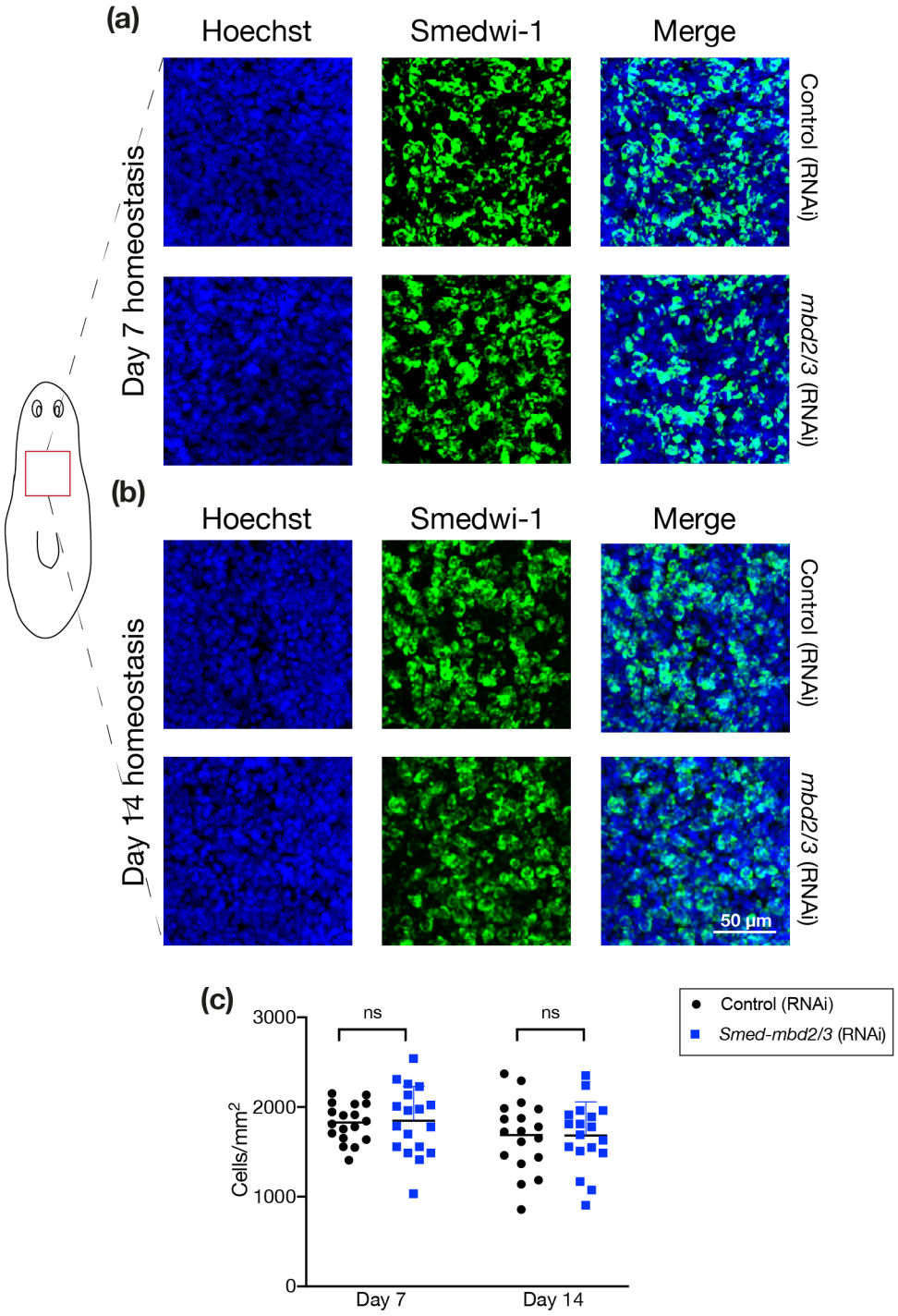

Figure 2.2: *Smed-mbd2/3* is not required for neoblast maintenance. *Smedwi-1* expression in unamputated animals (a) 7 days and (b) 14 days after knockdown. The distribution of *smedwi-1* labelled neoblasts was consistent between *Smed-mbd2/3*(RNAi) animals and controls during homeostasis. (c) Graph illustrating mean number of cells normalised to area. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. *Smedwi-1*$^+$ numbers do not differ significantly between *Smed-mbd2/3*(RNAi) and control animals during homeostasis.

are affected after the silencing of *Smed-mbd2/3*, I performed in situ hybridisation for sigma-, zeta-, and gamma-class neoblasts.

***Smed-mbd2/3*(RNAi) does not affect sigma-class neoblasts.** In order to check whether knockdown of *Smed-mbd2/3* affects pluripotent stem cells, I labelled sigma neoblasts in RNAi animals 7 and 14 days after knockdown during homeostasis. Sigma neoblasts are double positive for *smedwi-1* and the sigma pool (*soxP-1* and *soxP-2*) [van Wolfswinkel et al., 2014]. After knockdown of *Smed-mbd2/3*, animals were fixed at two different time points, one early, before any phenotypic changes could appear, and one late, just before visible changes occur in animals. These animals were stained with the two RNA probes and imaged. Figure 2.3a and b respectively show representative images of control and *Smed-mbd2/3*(RNAi) animals 7 and 14 days post knockdown. At both time points, control and *Smed-mbd2/3*(RNAi) animals show a similar distribution of sigma neoblasts.

In order to confirm this finding, a small pre-pharyngeal region of each worm was selected and the number of cells double positive for *smedwi-1* and the sigma pool probes was counted. These numbers were normalised to the area and then graphed. Figure 2.4a-b shows representative images of areas used for cell counting at day 7 and 14. The relative number of sigma neoblasts was not significantly different between *Smed-mbd2/3*(RNAi) and controls (Fig. 2.4c). Together, these results suggest that *Smed-mbd2/3* does not play a role in the regulation of sigma neoblasts during tissue turnover.

***Smed-mbd2/3*(RNAi) does not affect zeta-class neoblasts.** Functional experiments have revealed that zeta neoblasts are committed to the epidermal lineage, which is currently the best-described lineage in planarians [van Wolfswinkel et al., 2014]. Zeta neoblasts do not appear to have any self-renewing capacity. Inhibition of *Smed-mbd2/3* results in reduced numbers of late epidermal progeny, which may be explained by an accumulation of zeta neoblasts. The zeta neoblasts might be unable to differentiate to form epidermal progeny cells, leading to their accumulation. To test this, I labelled zeta neoblasts in RNAi animals 7 and 14 days after knockdown. Zeta neoblasts are double positive for *smedwi-1* and the zeta pool of RNA probes (*zfp-1* and *soxP-3*). After

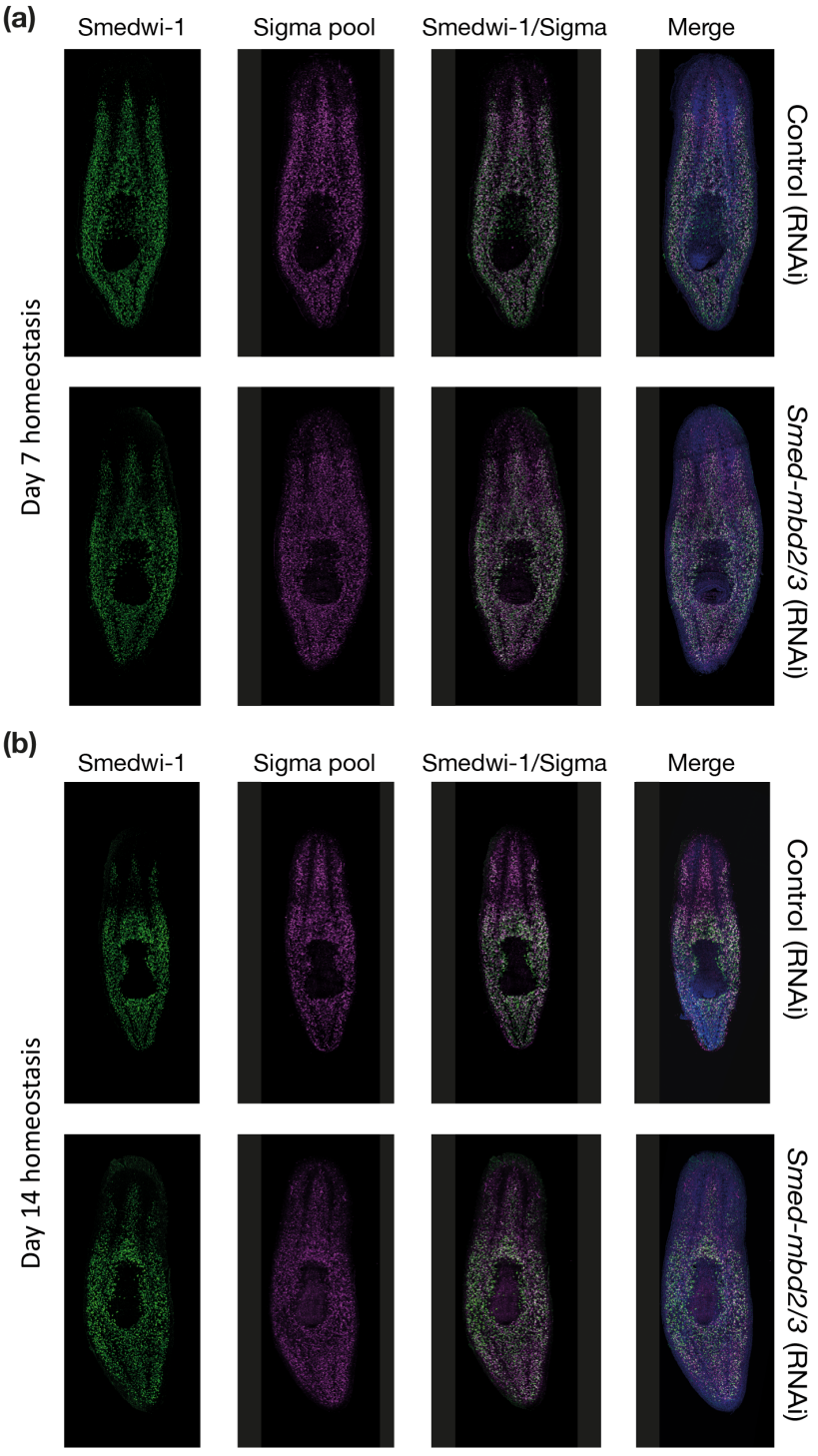

Figure 2.3: Sigma-class neoblast maintenance is not affected by *Smed-mbd2/3* knockdown. Lineage markers labelling sigma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distributions of sigma-neoblasts at the two time points appear similar.

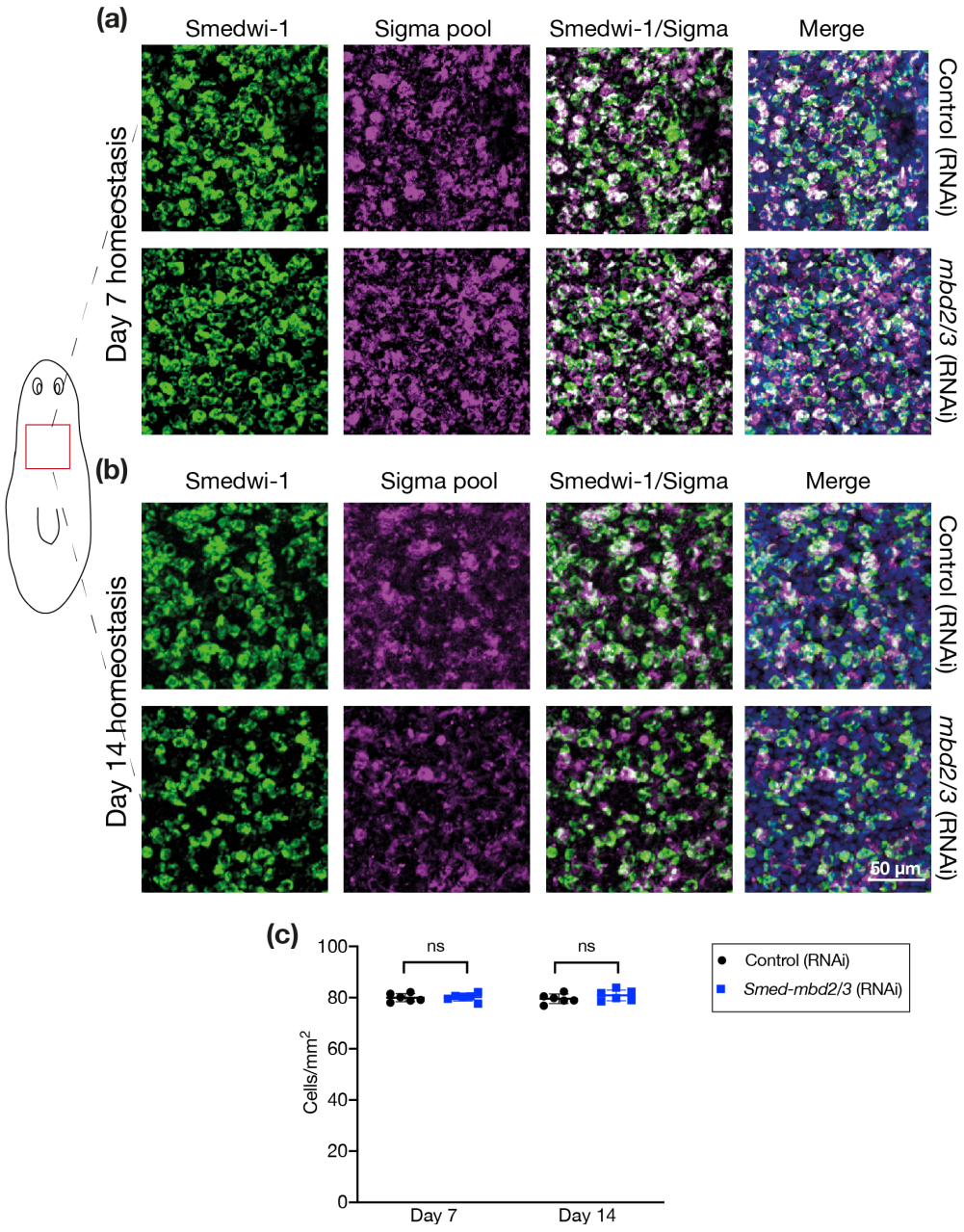

Figure 2.4: Maintenance of sigma-class neoblasts is not affected by knockdown of *Smed-mbd2/3*. Lineage markers labelling sigma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of sigma-neoblasts normalised to area 7 and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Sigma neoblast counts do not differ significantly between *Smed-mbd2/3*(RNAi) and control animals during homeostasis.

knockdown unamputated animals were fixed at the two time points, labelled for these markers and imaged. Figure 2.5a and b respectively show representative images of control and *Smed-mbd2/3*(RNAi) animals 7 and 14 days post knockdown. At both time points, control and *Smed-mbd2/3*(RNAi) animals show similar distributions of zeta neoblasts.

In order to confirm this finding, a small pre-pharyngeal region of each worm was selected and the number of cells double positive for *smedwi-1* and the zeta pool probes was counted. These numbers were normalised for area and then graphed. Figure 2.6a-b show representative images of the areas used for cell counting at days 7 and 14. The relative number of zeta neoblasts was not significantly different between *Smed-mbd2/3*(RNAi) and controls (Fig. 2.6c).

These results suggest that during tissue turnover, *Smed-mbd2/3* does not affect zeta neoblast maintenance or early differentiation, but prevents cells from differentiating completely. New tissue made after MBD2/3 knockdown is unpigmented or less pigmented, which suggests that the epidermis, although formed, is not quite the same as in wild-type worms. Although an insufficient number of *agat-1*$^{+}$ cells are produced, the animals appear to form an epidermis nonetheless.

***Smed-mbd2/3*(RNAi) is not required for the maintenance of gamma-class neoblasts.** Previously, Jaber-Hijazi et al. [2013] showed that *Smed-mbd2/3* (RNAi) animals are unable to remodel and regenerate the gut correctly (figures 3D and 3H in the paper). The expression patterns of the gamma sub-population of neoblasts suggests that these stem cells are very likely committed to making the gut and endodermal lineages [van Wolfswinkel et al., 2014].

In order to ascertain whether knockdown of *Smed-mbd2/3* affects gamma neoblasts, I labelled gamma neoblasts. Gamma-neoblasts are double positive for *smedwi-1* and the gamma pool (*gata4/5/6* and *hnf-4*) [van Wolfswinkel et al., 2014]. Figure 2.7a and b respectively show representative images of control and *Smed-mbd2/3*(RNAi) animals labelled with gamma neoblast markers 7 and 14 days post knockdown. At both time points, control and *Smed-mbd2/3*(RNAi) animals show similar distributions of gamma neoblasts and gut branches.

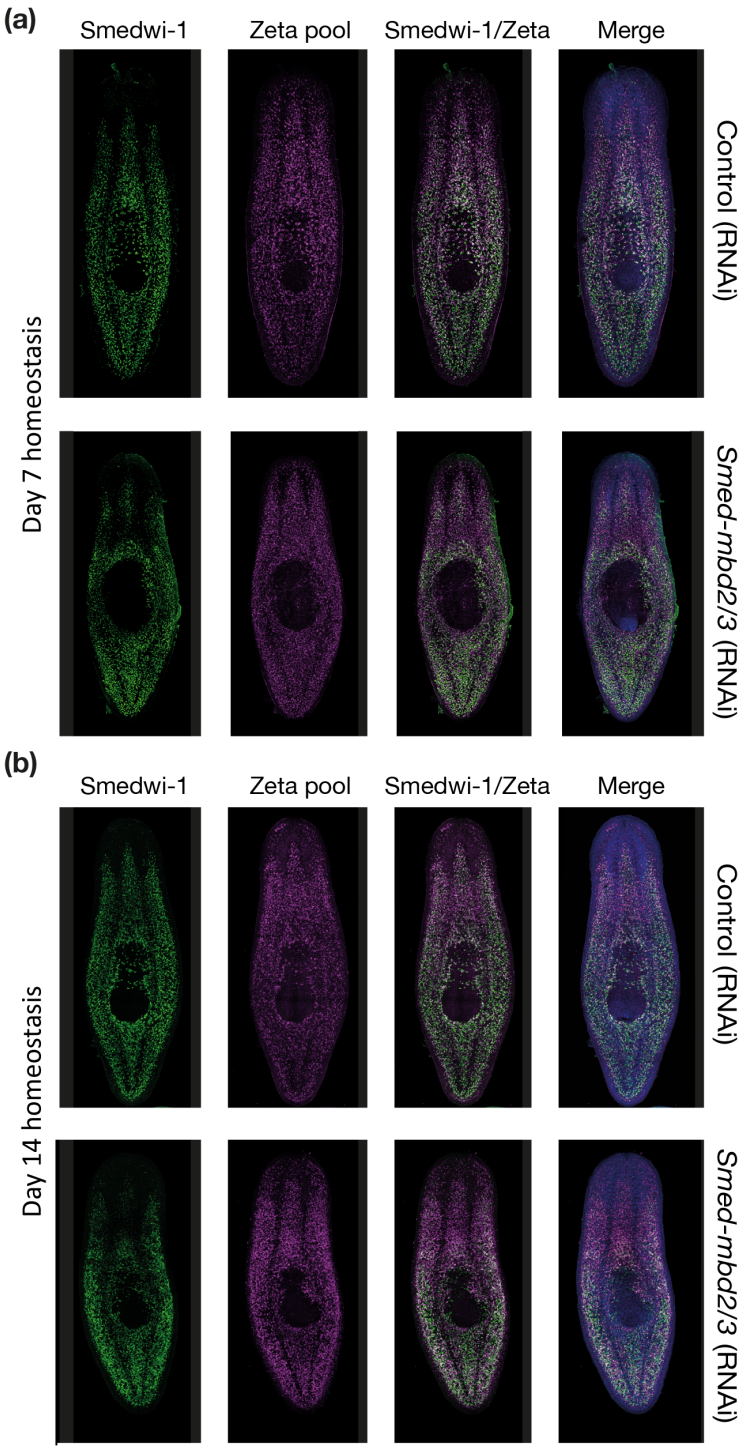

Figure 2.5: Zeta-class neoblast maintenance is not affected by knockdown of *Smed-mbd2/3*. Lineage markers labelling zeta-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distributions of zeta-neoblasts at the two time points appear similar.

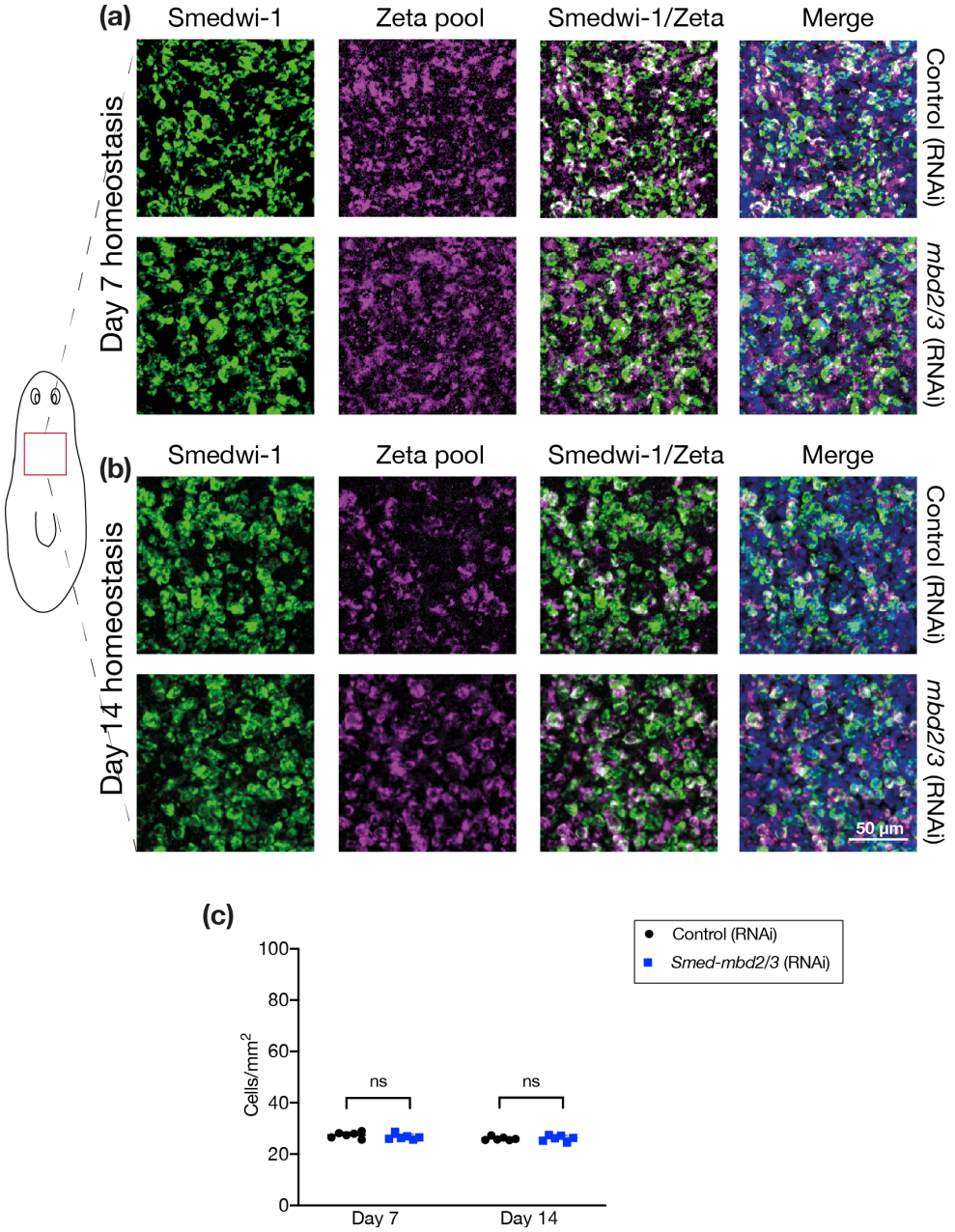

Figure 2.6: Maintenance of zeta-class neoblasts is not affected by knockdown of *Smed-mbd2/3*. Lineage markers labelling zeta-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating the relative number of zeta-neoblasts normalised to area 7 days and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Zeta neoblast counts do not differ significantly between *Smed-mbd2/3*(RNAi) and control animals during homeostasis.

In order to confirm this finding, a small pre-pharyngeal region of each worm was selected and the number of gamma neoblasts labelled was counted. These numbers were normalised to the area and then plotted on a graph. Figure 2.8a-b show representative images of areas used for cell counting at days 7 and 14. The relative number of gamma neoblasts was not significantly different between *Smed-mbd2/3*(RNAi) and controls (Fig. 2.8c). Together, these results suggest that *Smed-mbd2/3* does not play a role in the regulation gamma neoblasts during tissue turnover. Defects in gut branching were seen in the previous study after the inhibition of *Smed-mbd2/3* [Jaber-Hijazi et al., 2013]. As new intestinal cells arise from the progeny of gamma neoblasts, suggesting that *Smed-mbd2/3* might be essential for the differentiation of this neoblast class.

Knockdown of *Smed-mbd2/3* leads to the failure to progress through differentiation

In order to investigate whether neoblasts are able to differentiate, in situ hybridisation for early epidermal progeny cells (*prog-1*$^+$) and late epidermal progeny cells (*agat-1*$^+$) was performed. These genes label sub-epidermal cells that undergo rapid cell turnover. Although *Smed-mbd2/3*(RNAi) does not affect neoblast maintenance, it leads to an accumulation of early epidermal progenitors (*prog-1*$^+$). 7 days after knockdown, both control and *mbd2/3*(RNAi) animals exhibit similar distributions of *prog-1* and *agat-1* expressing cells (Fig. 2.9a). After 14 days, an accumulation of *prog-1* expressing cells was observed in *Smed-mbd2/3*(RNAi) worms compared to control worms (Fig. 2.9a). The control animals showed a uniform distribution of *agat-1* expressing cells at both time points. In order to confirm this, the number of each of the two cell types was counted and normalised to the area. The number of *prog-1*$^+$ and *agat-1*$^+$ cells was not significantly different between *Smed-mbd2/3*(RNAi) and controls (Fig. 2.9b) after 7 days while a significant difference was seen after 14 days (Fig. 2.9c). The results agree with the findings of the previous study [Jaber-Hijazi et al., 2013]. This suggests that *Smed-mbd2/3* knockdown worms are able to initiate differentiation along the epidermal lineage but not complete it. Any new tissue formed after *mbd2/3* knockdown

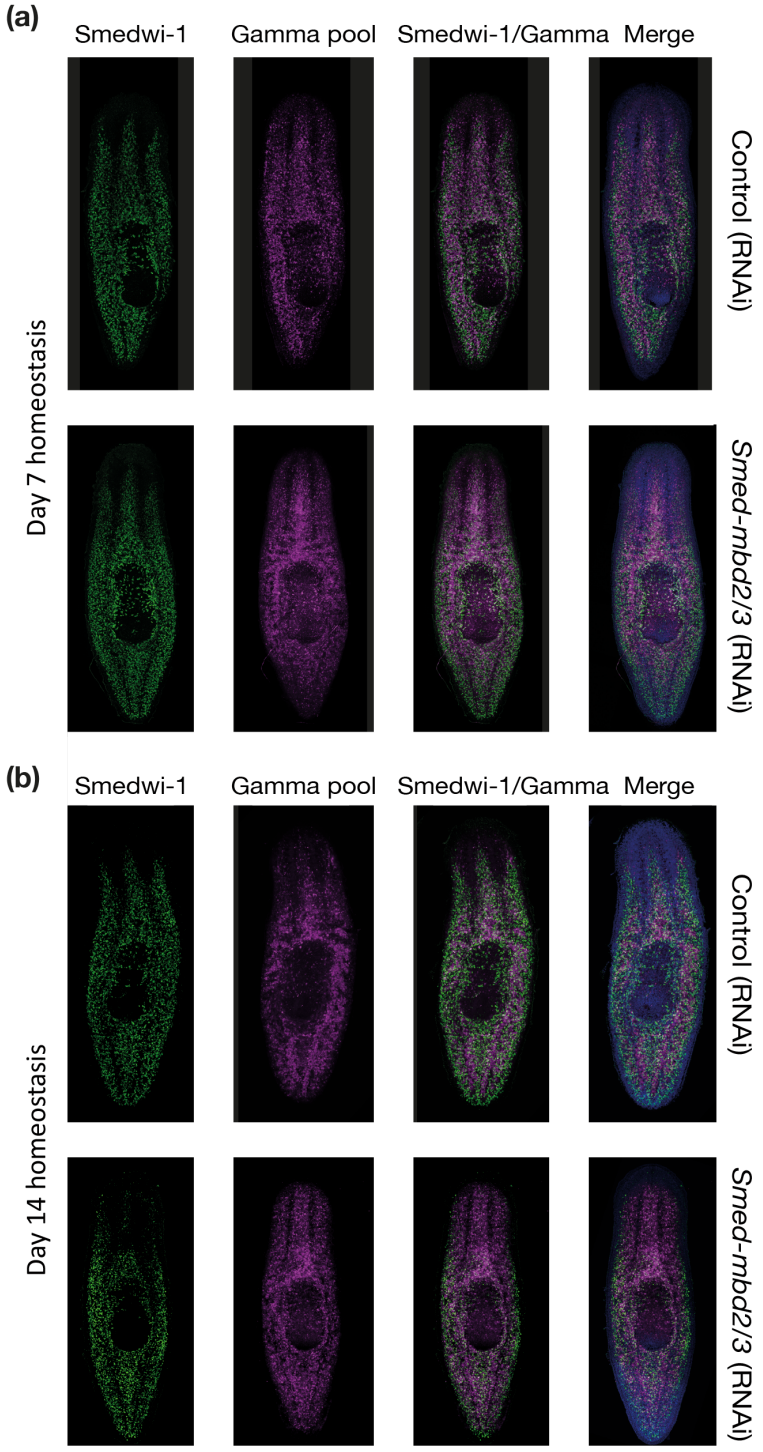

Figure 2.7: Gamma-class neoblast maintenance is not affected by knockdown of *Smed-mbd2/3*. Lineage markers labelling gamma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distributions of gamma-neoblasts at the two time points appear similar.

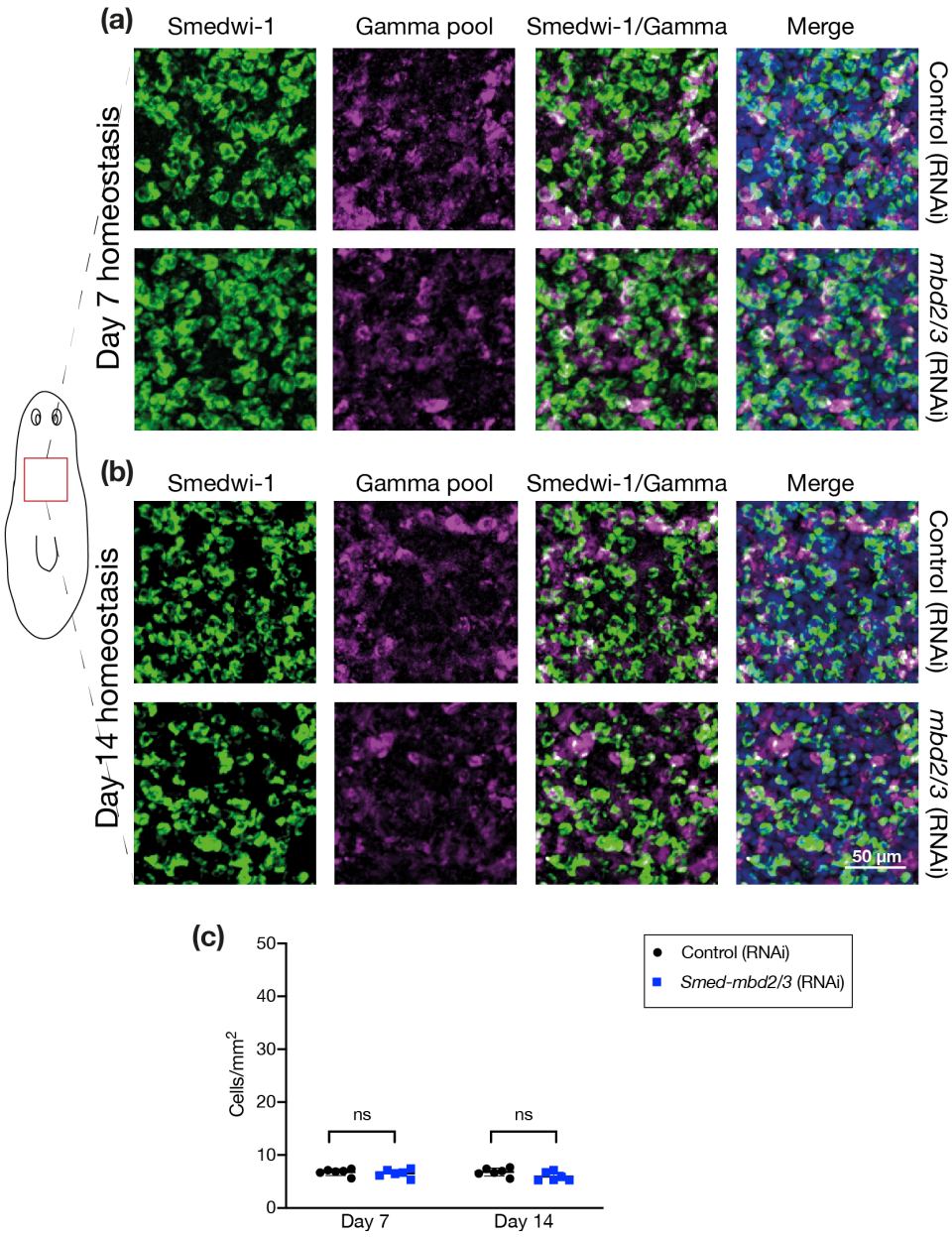

Figure 2.8: Knockdown of *Smed-mbd2/3* does not affect gamma-class neoblasts. Lineage markers labelling gamma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of gamma-neoblasts normalised to area 7 days and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Gamma neoblast counts are consistent between *Smed-mbd2/3*(RNAi) and control animals during homeostasis.

remains unpigmented, suggesting that although the epidermis is formed, it is not quite the same as wild-type worms.

Together, the results so far show that *Smed-mbd2/3* is not essential for stem cell maintenance or proliferation. However, *Smed-mbd2/3* is essential for the maintenance of differentiated cells. The *Smed-mbd2/3*(RNAi) phenotype disentangles the two main processes in stem cells—self-renewal and pluripotency. These functions are analogous to that of the mammalian MBD3. Mbd3 is required for the stable formation of the NuRD complex in embryos and has an essential role in mESC differentiation and cell fate decisions [Kaji et al., 2006]. *Mbd3*$^{-/-}$ ESCs are capable of self-renewal but are unable to differentiate properly, with ESCs defaulting to a trophectodermal fate [Kaji et al., 2006, 2007, Zhu et al., 2009]. Kaji et al. [2006] showed that *Mbd3*$^{-/-}$ ESCs lack the ability to differentiate within erythroid bodies and consistently express pluripotent cell markers. Furthermore, these cells are unable to downregulate expression of undifferentiated cell markers [Kaji et al., 2006]. *Mbd3*$^{-/-}$ ESCs were shown to be capable of initiating one early stage of differentiation in embryoid bodies but were unable to proceed beyond it and instead commit towards the typically inaccessible trophectoderm lineage [Kaji et al., 2006].

Smed-hdac1 is essential for tissue homeostasis and regeneration

Functional studies of histone deacetylase 1 (HDAC1) has been previously reported in planarians [Robb and Alvarado, 2014]. *Smed-hdac1* is highly expressed in neoblasts; knockdown of this gene mimics the effects of irradiation (head regression, ventral curling, and the inability to regenerate after amputation), indicating it plays an essential role in the regulation of neoblast function [Eisenhoffer et al., 2008, Reddien et al., 2005b, Robb and Alvarado, 2014]. I used *Smed-hdac1* as a positive control in all experiments, as it also forms a part of the NuRD complex. To assess the role of *Smed-hdac1* in tissue maintenance during homeostasis, worms were left unamputated for phenotypic observation after two rounds of injections. *Smed-hdac1*(RNAi) animals exhibited varying degrees of anterior regression, starting at the tip and margins, as early as after 8 days. A notch appeared at the middle anterior tip of the head of

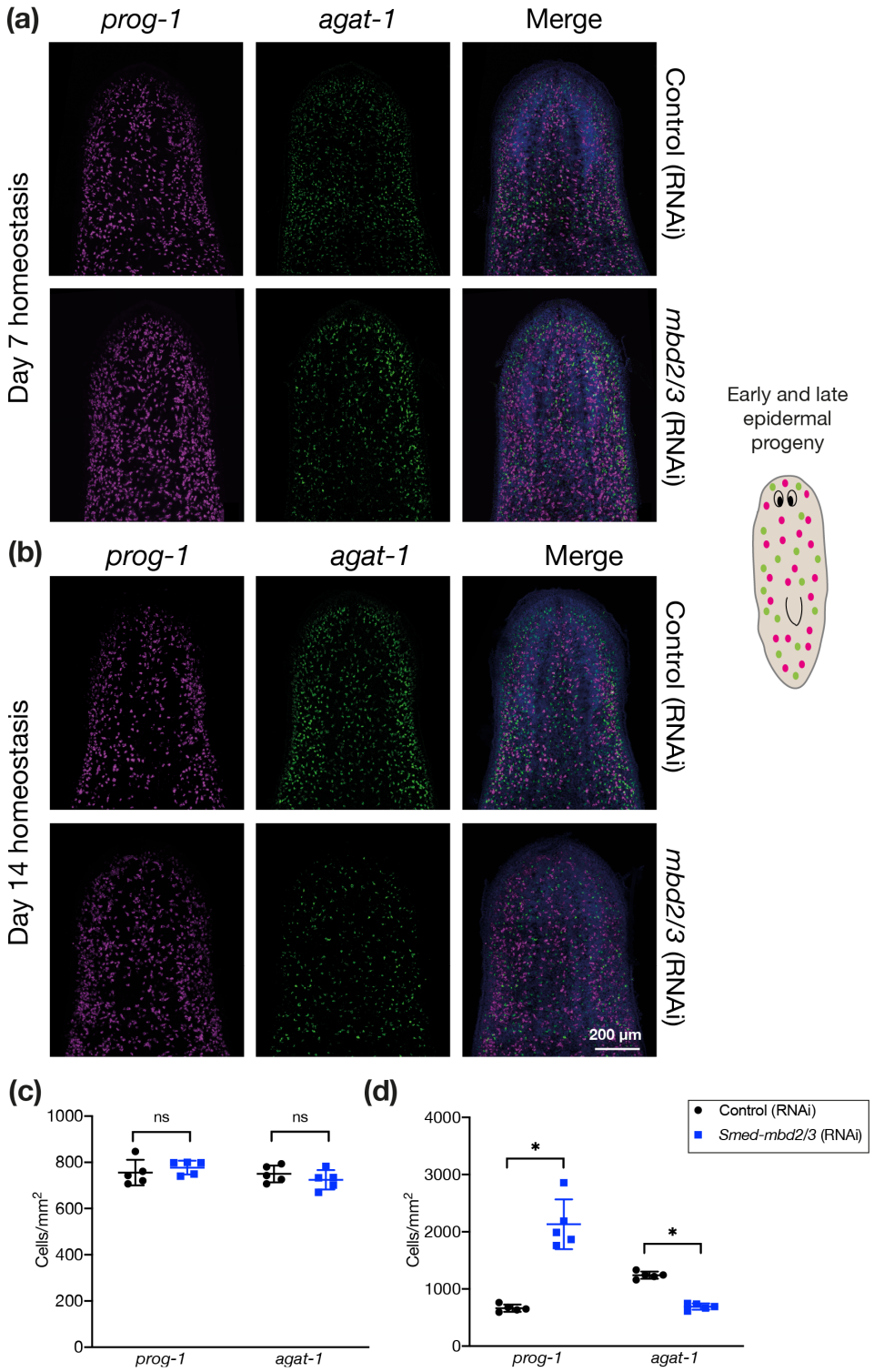

Figure 2.9: Differentiation of epidermal progenitor cells is perturbed following knockdown of *Smed-mbd2/3*. Expression of *prog-1* and *agat-1* in unamputated *Smed-mbd2/3*(RNAi) and control animals (a) 7 days and (b) 14 days after knockdown. Graph illustrating mean number of each cell type normalised to area (c) 7 days and (d) 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

the animals and typically a dark cross developed near the eyes (Fig. 2.10a). Some animals were missing the anterior portion of the head, including eyes. Silencing of *Smed-hdac1* elicits stronger phenotypic defects in animals than the silencing of *Smed-mbd2/3*. The control animals showed no defects in tissue turnover and all survived up to day 21 (3 weeks), while *Smed-hdac1*(RNAi) animals had ultimately lysed by this time point (Fig. 2.10a). An analysis of defects caused by RNAi was performed by assessing cell-specific markers using in situ hybridisation.

To confirm the role of HDAC1 during regeneration, animals were injected with dsRNA corresponding to *Smed-hdac1* and *gfp* (as control) 6 times. In each case, 10 animals were injected and then amputated into 3 pieces – head, trunk, and tail. In the regeneration assay, phenotypic defects due to knockdown were visible even at early regeneration stages. Three regeneration time points (day 4, day 7 and day 10) are shown in Fig. 2.10b for comparison. During early regeneration, all control animals formed a proper blastema while *Smed-hdac1*(RNAi) animals formed extremely small blastemas. At regeneration day 7, however, the *Smed-hdac1* knockdown animals show recognisable signs of inability to replace missing structures. The head pieces showed signs of ventral curling, head regression and slower mobility. The trunk and middle pieces were unable to form eyes and other missing tissues. At day 10 of regeneration, 100% of trunk and tail pieces had not formed eyes. The head pieces were unable to form the pharynx or tail and showed regression, including loss of brain and/eye tissue. In contrast, control animals had regenerated all missing structures including eyes, tail and pharynx without any visible defects. In conclusion, knockdown of *Smed-hdac1* by RNAi results in severe defects during homeostasis and regeneration, with loss of viability amongst all worms (Fig. 2.10c-d).

Knockdown of *Smed-hdac1* impairs neoblast maintenance

In order to confirm the underlying cause of defects during regeneration and homeostasis in *Smed-hdac1*(RNAi) worms, animals were labelled with the stem cell marker *smedwi-1*. The number of *smedwi-1*$^+$ cells was counted in the pre-pharyngeal region and normalised according to the area (Fig. 2.11a-b). *Smed-hdac1*(RNAi) animals show a

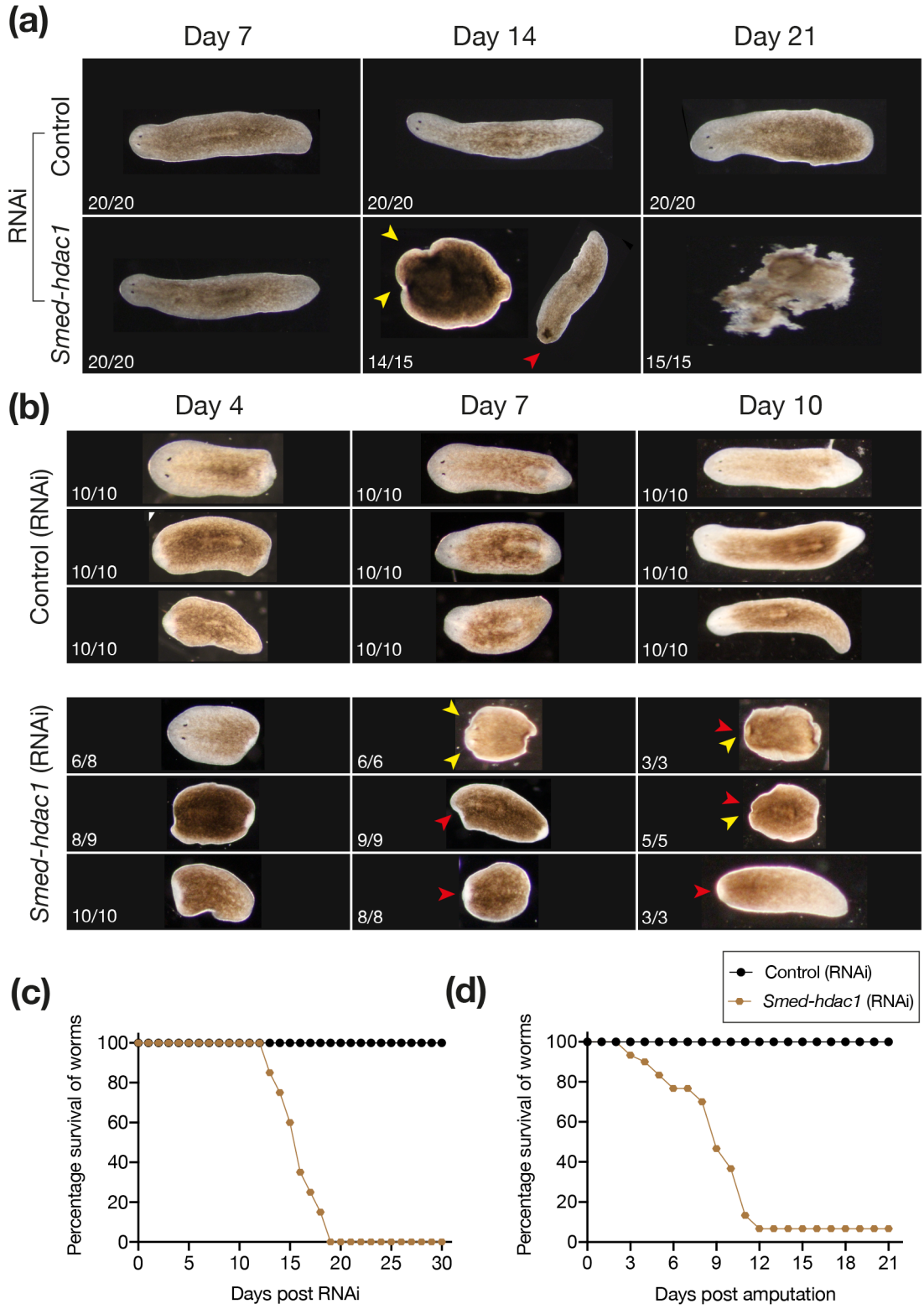

Figure 2.10: Knockdown of *Smed-hdac1* affects homeostasis and regeneration. (a) Intact animals during homeostasis after *Smed-hdac1* and control RNAi. (b) Head, middle and tail pieces during regeneration after *Smed-hdac1* and control RNAi. Survival curves of knockdown animals during (c) tissue homeostasis and (d) regeneration. White numbers indicate proportion of animals exhibiting the same phenotype at this time point. Yellow and red arrowheads respectively indicate head regression and no eyes.

drastic reduction in the number of neoblasts which consistent with the "no neoblast" phenotype seen after knockdown (head regression accompanied by a characteristic ventral curling). The number of *smedwi-1*$^+$ cells was significantly different between *Smed-hdac1*(RNAi) and controls both 7 and 14 days post RNAi (Fig. 2.11c). These results corroborate those of previous studies [Zhu and Pearson, 2013, Robb and Alvarado, 2014].

As the neoblast populations are depleted after *Smed-hdac1*(RNAi), it is likely that the different lineage committed stem cells are equally affected. This was confirmed using in situ hybridisation to label sigma, zeta, and gamma neoblasts.

Neoblasts in *Smed-hdac1*(RNAi) animals fail to form lineage committed stem cells. After the knockdown of *Smed-hdac1*, unamputated animals were fixed at two different time points, 7 and 14 days, after knockdown. In order to check whether all stem cells population were affected, I labelled sigma neoblasts in RNAi animals. Sigma neoblasts, double positive for *smedwi-1* and the sigma pool (*soxP-1* and *soxP-2*), were counted in the pre-pharyngeal region and normalised to the area. Representative images of control and *Smed-hdac1*(RNAi) animals after 7 days (Fig. 2.12a) and 14 days (Fig. 2.12b) of knockdown are shown. At both time points, *Smed-hdac1*(RNAi) animals exhibited significantly lower numbers of sigma neoblasts (Fig. 2.12c).

As the number of sigma neoblasts are drastically reduced *Smed-hdac1*(RNAi), it is most likely that zeta and gamma neoblasts are also affected. To confirm this, I labelled zeta neoblasts using zeta pool (*zfp-1*, *soxP-3*) and *smedwi-1*. The number of double positive zeta neoblasts was counted and normalised to the area. Representative images of control and *Smed-hdac1*(RNAi) animals stained for zeta neoblasts after 7 days (Fig. 2.13a) and 14 days (Fig. 2.13b) of knockdown are shown. At both time points, *Smed-hdac1*(RNAi) animals exhibited significantly lower numbers of zeta neoblasts (Fig. 2.13c).

To confirm the loss of gamma neoblasts, I labelled them using gamma pool (*gata4/5/6* and *hnf-4*) and *smedwi-1*. The number of double positive cells was counted and normalised to area. Representative images of control and *Smed-hdac1*(RNAi) animals with labelled gamma neoblasts after 7 days (Fig. 2.14a) and 14 days (Fig. 2.14b)

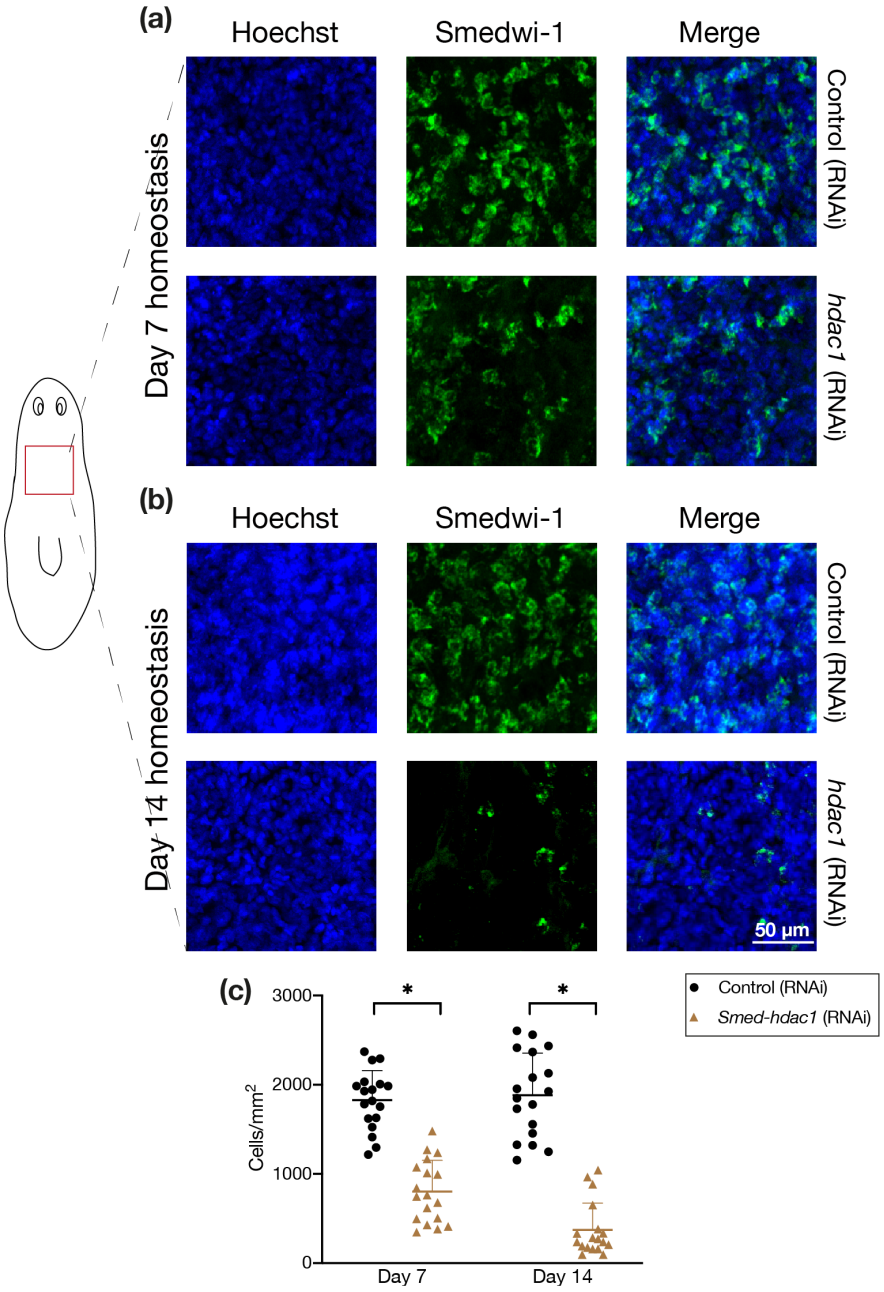

Figure 2.11: Maintenance of neoblasts is severely affected by the inhibition of *Smed-hdac1*. *Smedwi-1* expression in unamputated animals (a) 7 days and (b) 14 days after knockdown. The distribution of neoblasts assessed by *smedwi-1* expression is consistent between *Smed-hdac1*(RNAi) animals and controls during homeostasis. (b) Graph illustrating mean number of cells normalised per imaged animal area 7 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

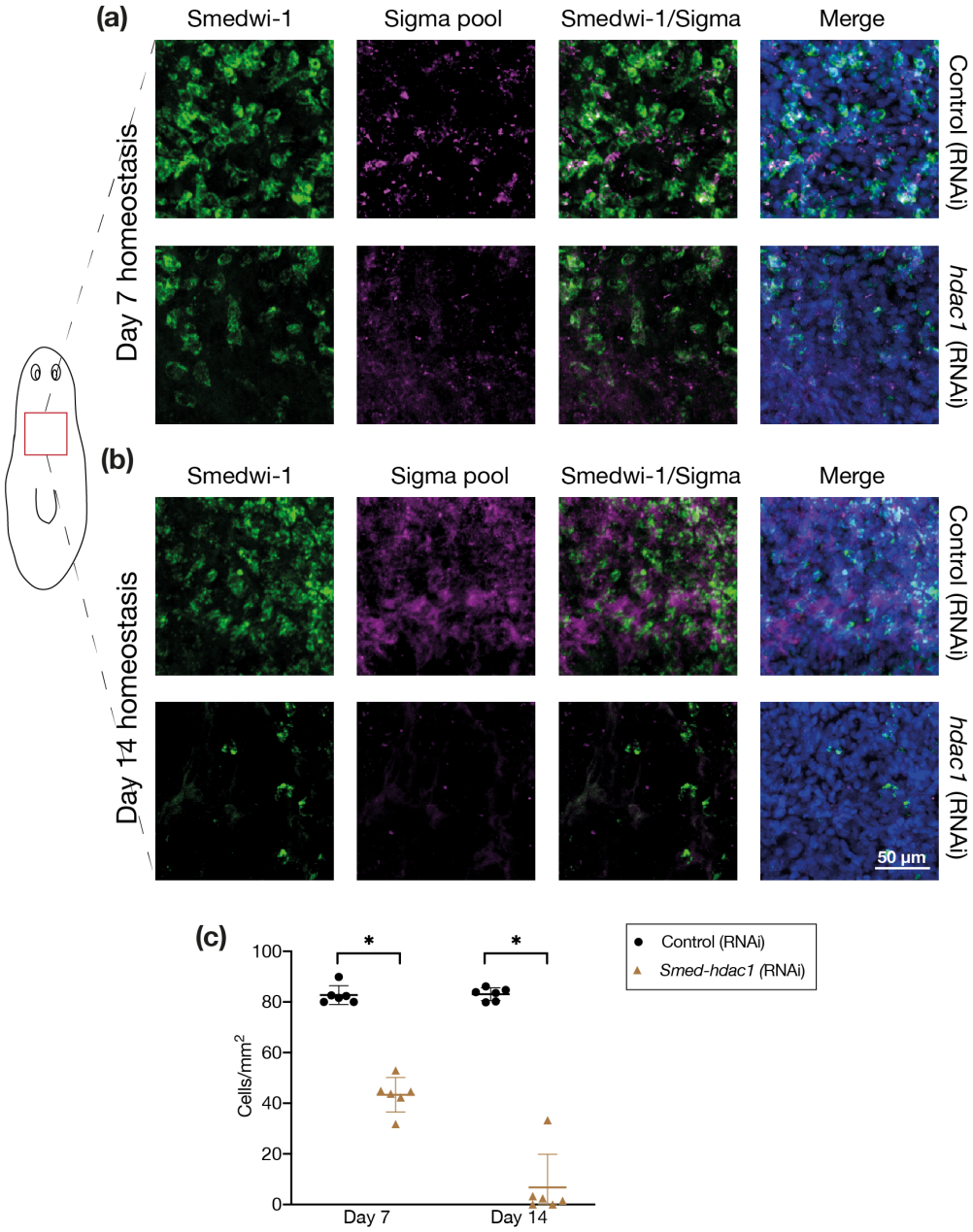

Figure 2.12: Knockdown of *Smed-hdac1* leads to a dramatic loss in sigma neoblasts. Lineage markers labelling sigma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of sigma-neoblasts normalised to area 7 days and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

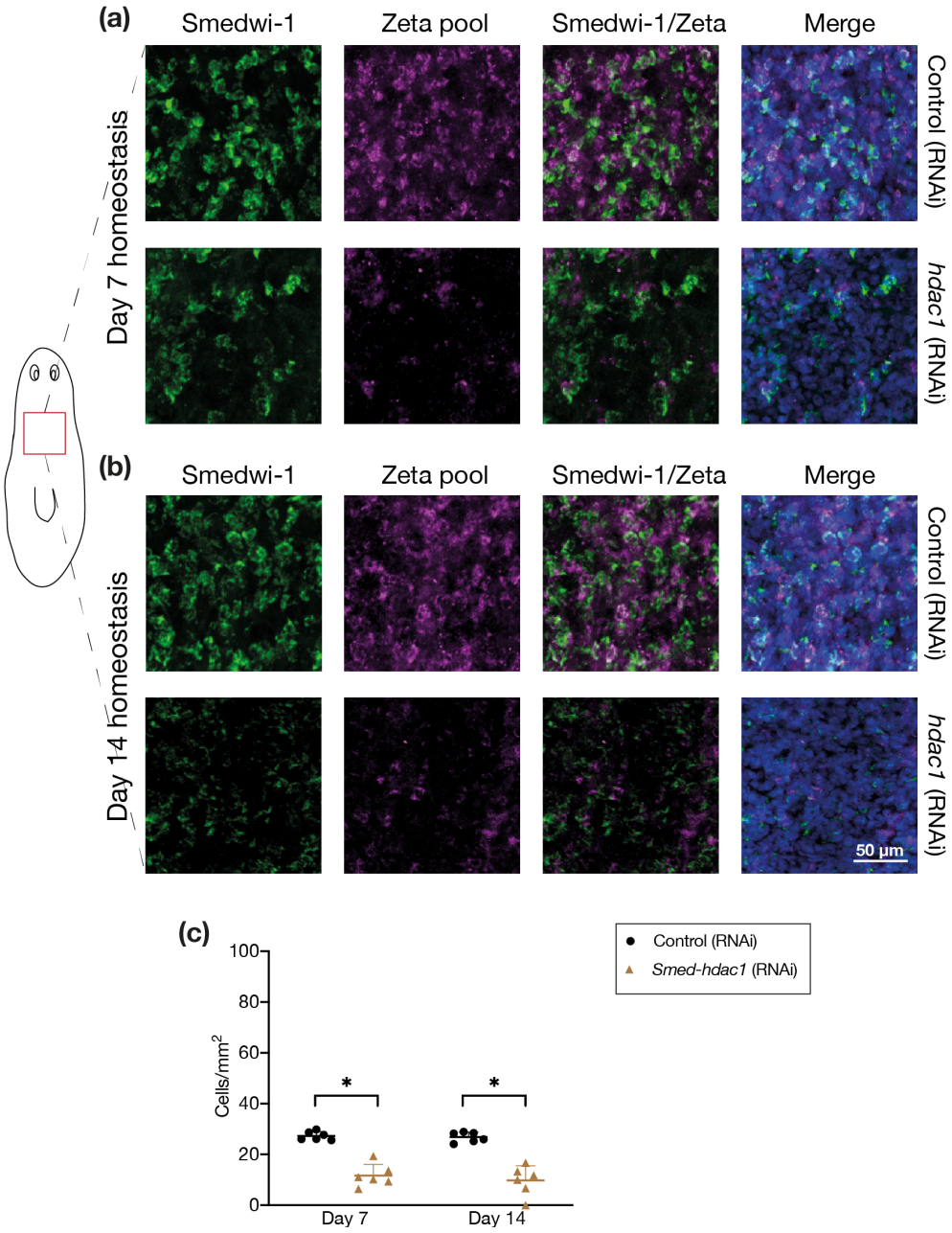

Figure 2.13: Maintenance of zeta-class neoblasts is affected by the knockdown of *Smed-hdac1*. Lineage markers labelling zeta-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of zeta-neoblasts normalised to area 7 days and 14 days after knockdown during homeostasis. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

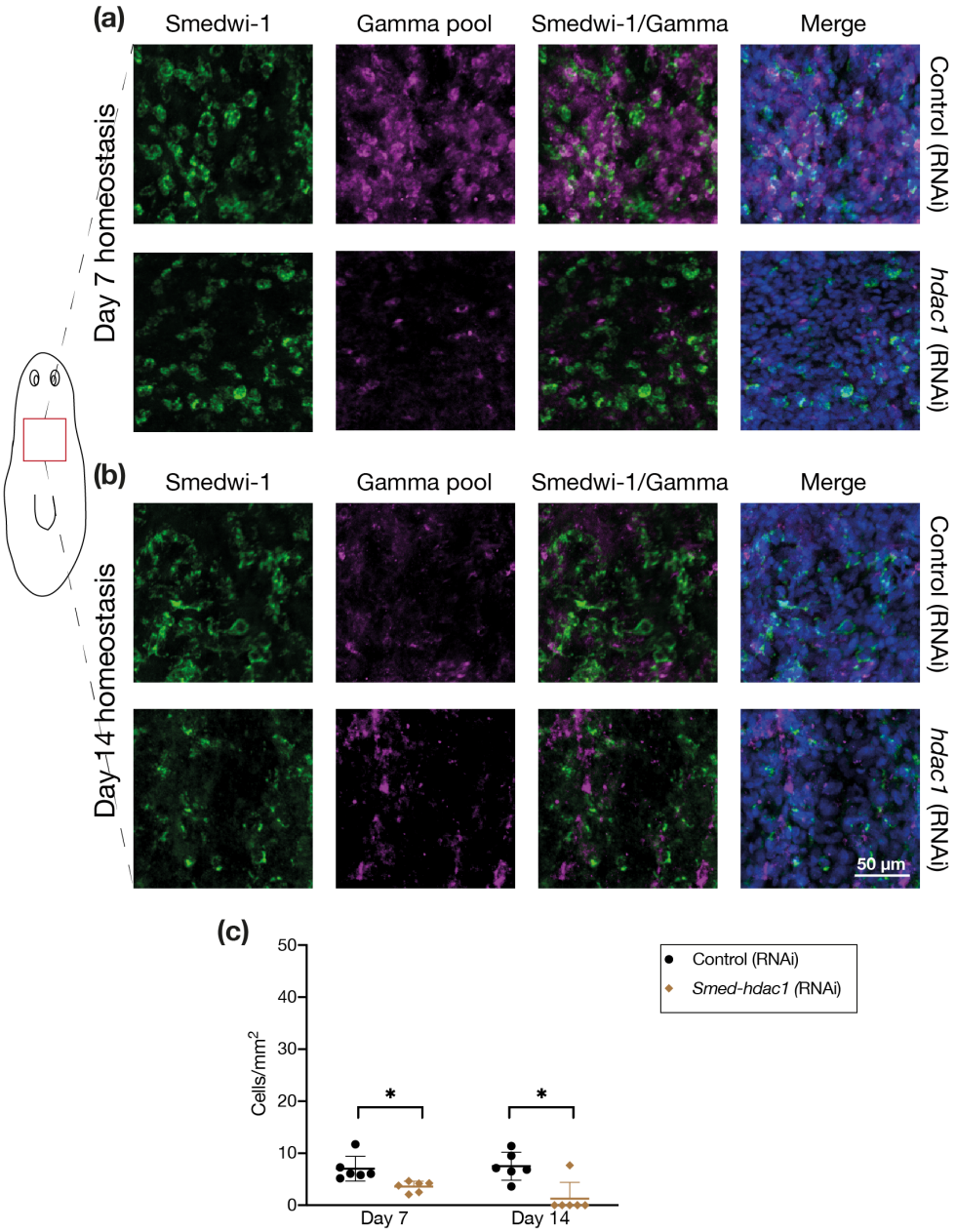

Figure 2.14: Silencing of *Smed-hdac1* affects gamma-class neoblasts. Lineage markers labelling gamma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of gamma-neoblasts normalised per imaged animal area 7 days and 14 days after knockdown during homeostasis. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

of knockdown are shown. At both time points, *Smed-hdac1*(RNAi) animals exhibited significantly lower numbers of gamma neoblasts (Fig. 2.14c). As expected, a drastic decrease in all three stem cell populations was seen after knockdown of *Smed-hdac1*.

Knockdown of *Smed-hdac1* leads to decreased production of epidermal progenies

Considering the dramatic loss of neoblasts after the knockdown of *hdac1*, it is likely that all differentiated progenies are equally affected. I labelled animals with *prog-1*$^+$ and *agat-1*$^+$ markers. After knockdown, *Smed-hdac1*(RNAi) animals exhibited a drastic reduction in the number of *prog-1* and *agat-1* expressing cells (Fig. 2.15a-b) compared to control animals.

The numbers of *prog*$^+$ and *agat*$^+$ cells were counted and normalised to the area. The numbers of *prog-1*$^+$ and *agat-1*$^+$ cells were significantly lower in *Smed-hdac1*(RNAi) after 7 days (Fig. 2.15c) and 14 days (Fig. 2.15d) compared to the controls. After knockdown of *hdac1* in planarians, the levels of stem cells *smedwi1*$^+$ as well as different neoblast classes were significantly reduced. Given that zeta neoblasts give rise to *prog* and *agat* progeny cells, this result is expected. This observation is more drastic than that previously observed [Robb and Alvarado, 2014], where an obvious reduction in progeny cells is seen only after 12 days.

Overall, these results show that the knockdown of *Smed-hdac1* impairs the maintenance of stem cell pluripotency and differentiation. In contrast, inhibition of *Smed-mbd2/3* only affects differentiation. This explains the severity in phenotypic defects seen after inhibition of *hdac1* in planarians.

Smed-p66 is required for tissue homeostasis and regeneration

A novel function for GATA-type zinc finger domain protein (GATA2D) or p66 in suppressing photoreceptor neuron formation has been recently reported in planarians [Vásquez-Doorman and Petersen, 2016]. Knockdown of *Smed-p66* results in the regeneration of unpigmented eyes with normal numbers of eye pigment cup cells but with elevated numbers of photoreceptor neurons [Vásquez-Doorman and Petersen, 2016]. Animals also show a reduction in epidermal progenitor cells. This phenotype is a parallel to the phenotype I have observed after knockdown of *Smed-mbd2/3*, where an accumulation of early epidermal progeny is seen along with a reduction in late progeny (Fig. 2.9). Moreover, p66 and MBD2/3 are the only two components

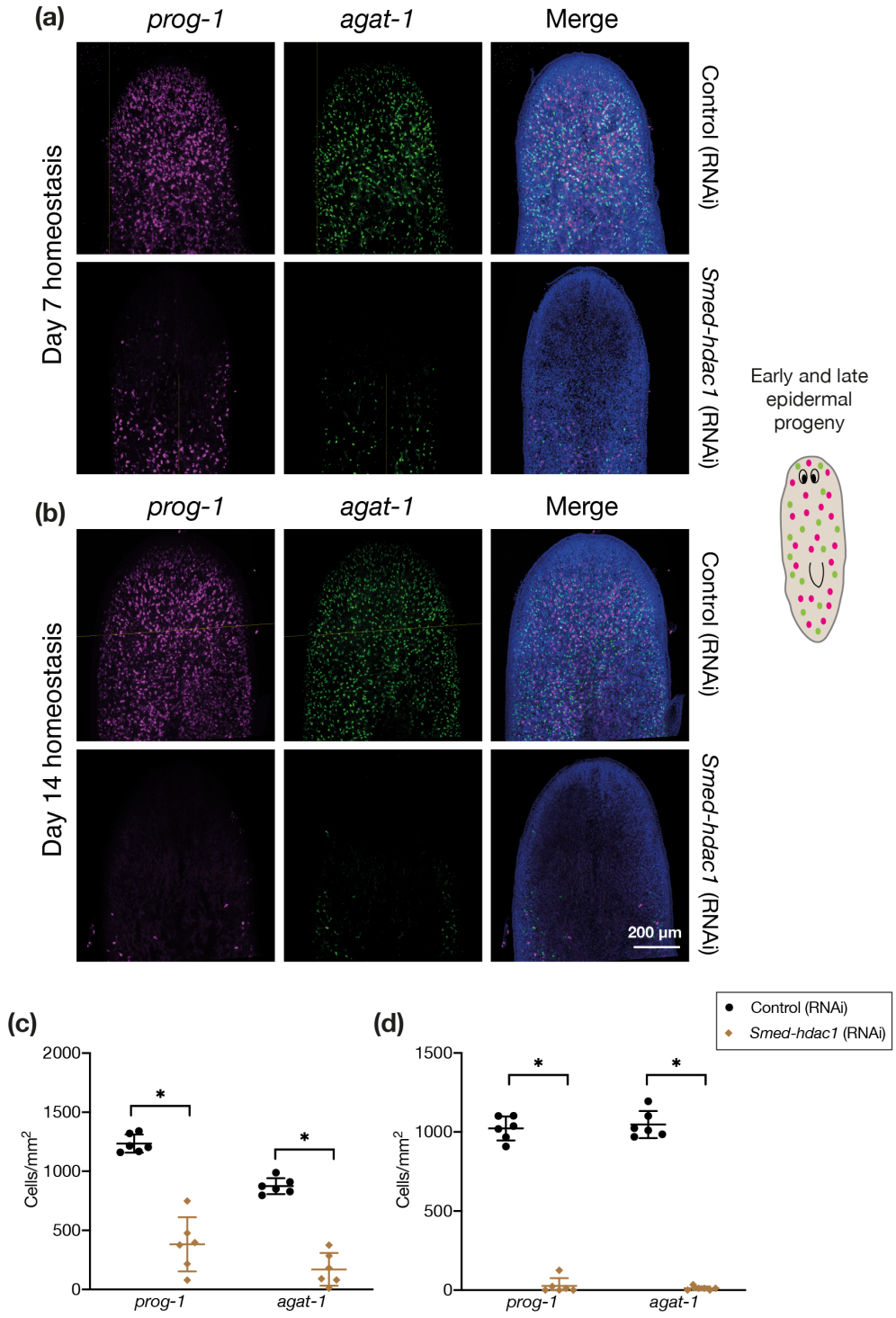

Figure 2.15: Differentiation of epidermal progenitor cells is perturbed following knockdown of *Smed-hdac1*. Expression of *prog-1* and *agat-1* in unamputated *Smed-hdac1*(RNAi) and control animals (a) 7 days and (b) 14 days after knockdown. Graph illustrating mean number of each cell type normalised to area (c) 7 days and (d) 14 days after knockdown during homeostasis. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

unique to the NuRD complex and can provide direct insights into the functions of the NuRD complex in planarians.

To assess the role of *Smed-p66* for proper tissue maintenance during homeostasis, RNAi worms were left unamputated for phenotypic observation. *Smed-p66*(RNAi) animals showed varying degrees of anterior regression, starting with the anterior tip. After 10 days, a notch appeared at the middle anterior tip of the head of the animals and typically a dark cross appeared near the eyes. After two weeks, more pronounced head regression was seen, with some worms missing parts of the brain or the entire head structure while the control animals were normal. The control animals showed no defects in tissue homeostasis and all survived up to day 27 while *Smed-p66* animals were dead well before this point.

To re-ascertain the essential role of p66 in planarian regeneration, I performed knockdown of *Smed-p66*. After the second round of injections, animals were amputated into 3 pieces – head, trunk, and tail. In the regeneration assay, phenotypic defects due to knockdown became apparent in the early stages. Three regeneration time points (day 4, day 7 and day 10) are shown in Fig. 2.16b for comparison. During early regeneration, all control animals formed a blastema and *Smed-p66*(RNAi) animals were unable to form proper blastema. 6 out of 9 head pieces at day 4 showed tissue damage and head regression. At regeneration day 7, all *p66*(RNAi) animals were unable to replace missing structures. The head pieces showed signs of curling, head regressions and slower mobility. The trunk and middle pieces were unable to form eyes and other missing tissues. At day 10 of regeneration defects in *Smed-p66*(RNAi) animals were more marked. 100% of trunk and tail pieces failed to form eyes. The head pieces were unable to form the pharynx or tail and showed head regression, including loss of brain and visible eyes. Generally, the slow motility phenotype increased in severity with the progression of regeneration – from low mobility to complete immobility. The defects are similar to those seen after inhibition of *Smed-mbd2/3*, albeit more pronounced. In contrast, control animals had regenerated all missing structures including eyes, tails and pharynx without any visible defects. In conclusion, loss of gene function of *Smed-p66* by RNAi results in defects during homeostasis and regeneration, with loss of viability amongst all worms (Fig. 2.16c-d).

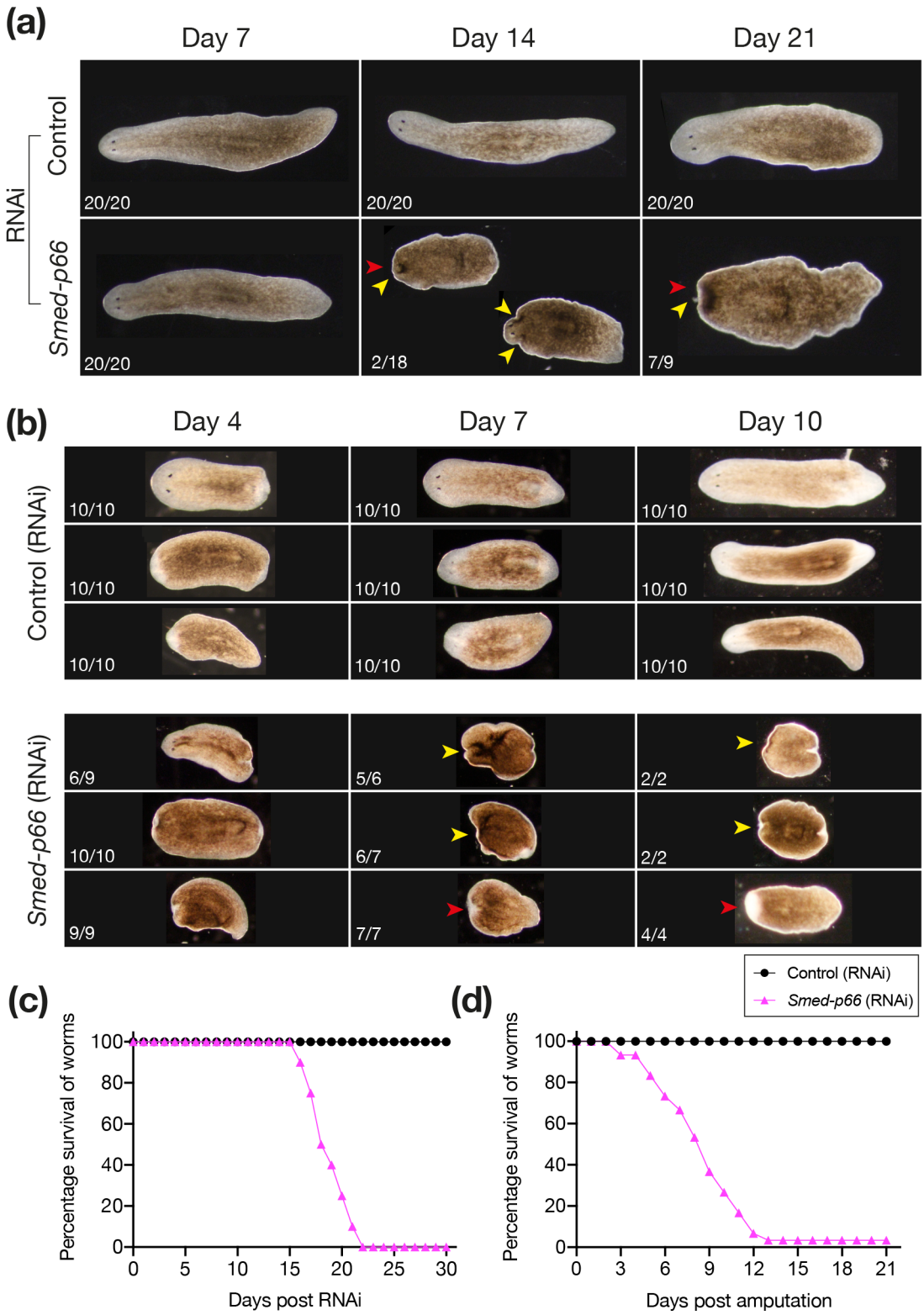

Figure 2.16: Inhibition of *Smed-p66* impairs tissue homeostasis and regeneration. (a) Intact animals during homeostasis after *Smed-p66* and control RNAi. (b) Head, middle and tail pieces during regeneration after *Smed-p66* and control RNAi. Survival curves of knockdown animals during (c) tissue homeostasis and (d) regeneration. White numbers indicate proportion of animals exhibiting the same phenotype at this time point. Yellow and red arrowheads respectively indicate head regression and no eyes.

Smed-p66 is not required for neoblast maintenance

As *Smed-p66*(RNAi) animals show a similar phenotype as the *Smed-mbd2/3* animals, I sought to determine if the molecular changes also follow. To determine whether impaired neoblast maintenance is the underlying cause of defects in *Smed-p66*(RNAi) worms, I labelled the animals with stem cell marker *smedwi-1*. No obvious difference was seen in *Smed-p66*(RNAi) animals. In order to confirm this, the number of *smedwi-1*$^+$ cells was counted and normalised to the area. The number of *smedwi-1*$^+$ cells were not significantly different between *Smed-p66*(RNAi) and controls after 7 days (Fig. 2.17a) and after 14 days (Fig. 2.17b). These results contradict those found by Vásquez-Doorman and Petersen [2016], where an apparent increase in expression of *smedwi-1* was seen. Further analysis of RNAi effects was performed to resolve this and to confirm any similarities with the *Smed-mbd2/3* phenotype.

***Smed-p66*(RNAi) does not affect sigma-class neoblasts.** As the total neoblast populations were unperturbed after the silencing of *Smed-p66*, I next ascertained if the different neoblast sub-populations were affected by the knockdown. In order to check whether pluripotent stem cells are affected, I labelled sigma neoblasts in unamputated animals 7 and 14 days after knockdown. Sigma-neoblasts are double positive for *smedwi-1* and the sigma pool of RNA probes (*soxP-1* and *soxP-2*) [van Wolfswinkel et al., 2014]. After knockdown of *Smed-p66*, animals were fixed at two different time points, one early, before any phenotypic changes could appear, and one late, just before visible changes occur in tissue. These animals were labelled with sigma neoblast probes and imaged. Figure 2.18a shows representative images of control and *Smed-p66*(RNAi) animals after 7 days of knockdown. Figure 2.18b shows representative images of control and *Smed-p66*(RNAi) animals 14 days post knockdown. At both time points, control and *Smed-p66*(RNAi) animals showed similar distributions of sigma neoblasts.

In order to confirm this finding, a small pre-pharyngeal region of each worm was selected and the number of cells double positive for *smedwi-1* and the sigma pool probes was counted. These numbers were normalised to the area and then graphed. Figure 2.19a shows representative images used for cell counting at day 7 time point and Fig. 2.19b at

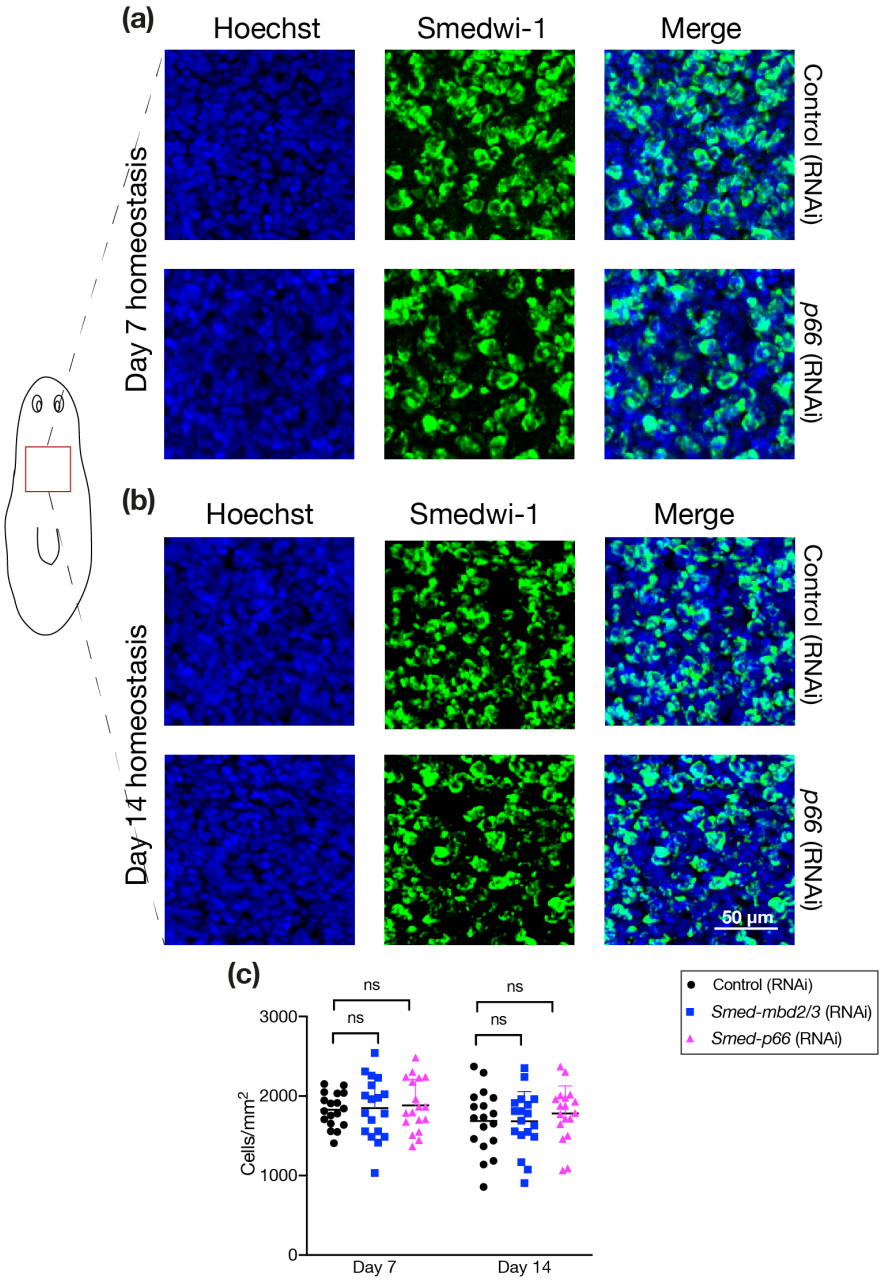

Figure 2.17: Knockdown of *Smed-p66* does not affect neoblast maintenance. *Smedwi-1* expression in unamputated animals (a) 7 days and (b) 14 days after knockdown. The distribution of neoblasts assessed by *smedwi-1* expression is consistent between *Smed-p66*(RNAi) and control animals during homeostasis. (b) Graph illustrating mean number of cells normalised to area after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

day 14 time point. The relative number of sigma neoblasts was not significantly different between *Smed-p66*(RNAi) and controls (Fig. 2.19c). Together, these results suggest that *Smed-p66* does not affect the maintenance sigma neoblasts during tissue turnover.

***Smed-p66*(RNAi) does not affect zeta-class neoblasts.** To test the changes in levels of zeta neoblasts, I probed RNAi animals for zeta (*zfp-1* and *soxP-3*) and *smedwi-1*. Zeta neoblasts are double positive for *smedwi-1* and the zeta pool of RNA probes. After knockdown of *Smed-p66*, animals were fixed and labelled using RNA probes and imaged. Figure 2.20a and b respectively show representative images of control and *Smed-p66*(RNAi) animals after 7 and 14 days of knockdown. At both time points, control and *Smed-p66* animals show similar distribution of zeta neoblasts.

To confirm this, a small pre-pharyngeal region of each worm was selected and the number of cells double positive for *smedwi-1* and the zeta pool probes was counted. These numbers were normalised for area and then graphed. Figure 2.21a shows representative images of areas used for cell counting at day 7 time point and Fig. 2.21b at day 14 time point. The relative numbers of zeta neoblasts were not significantly different between *Smed-p66*(RNAi) and controls (Fig. 2.21c).

***Smed-p66*(RNAi) does not affect gamma-class neoblasts.** In order to ascertain whether knockdown of *Smed-p66* causes any changes in gamma neoblasts, I labelled them in RNAi animals after 7 and 14 days. Gamma-neoblasts are double positive for *smedwi-1* and the gamma pool (*gata4/5/6* and *hnf-4*) [van Wolfswinkel et al., 2014]. After knockdown of *Smed-p66*, animals were fixed, labelled for gamma neoblasts and imaged. Figure 2.22a and b show representative images of control and *Smed-p66*(RNAi) animals after 7 and 14 days of knockdown, respectively. At both time points, control and *Smed-p66*(RNAi) animals show similar distributions of gamma neoblasts and gut branches.

In order to confirm this finding, a small pre-pharyngeal region of each worm was selected and the number of cells double positive for *smedwi-1* and the gamma pool probes was counted. These numbers were normalised to the area and then plotted on a graph. Figure 2.23a shows representative images of areas used for cell counting at the day 7 time point and Fig. 2.23b at the day 14 time point. The relative numbers of gamma neoblasts

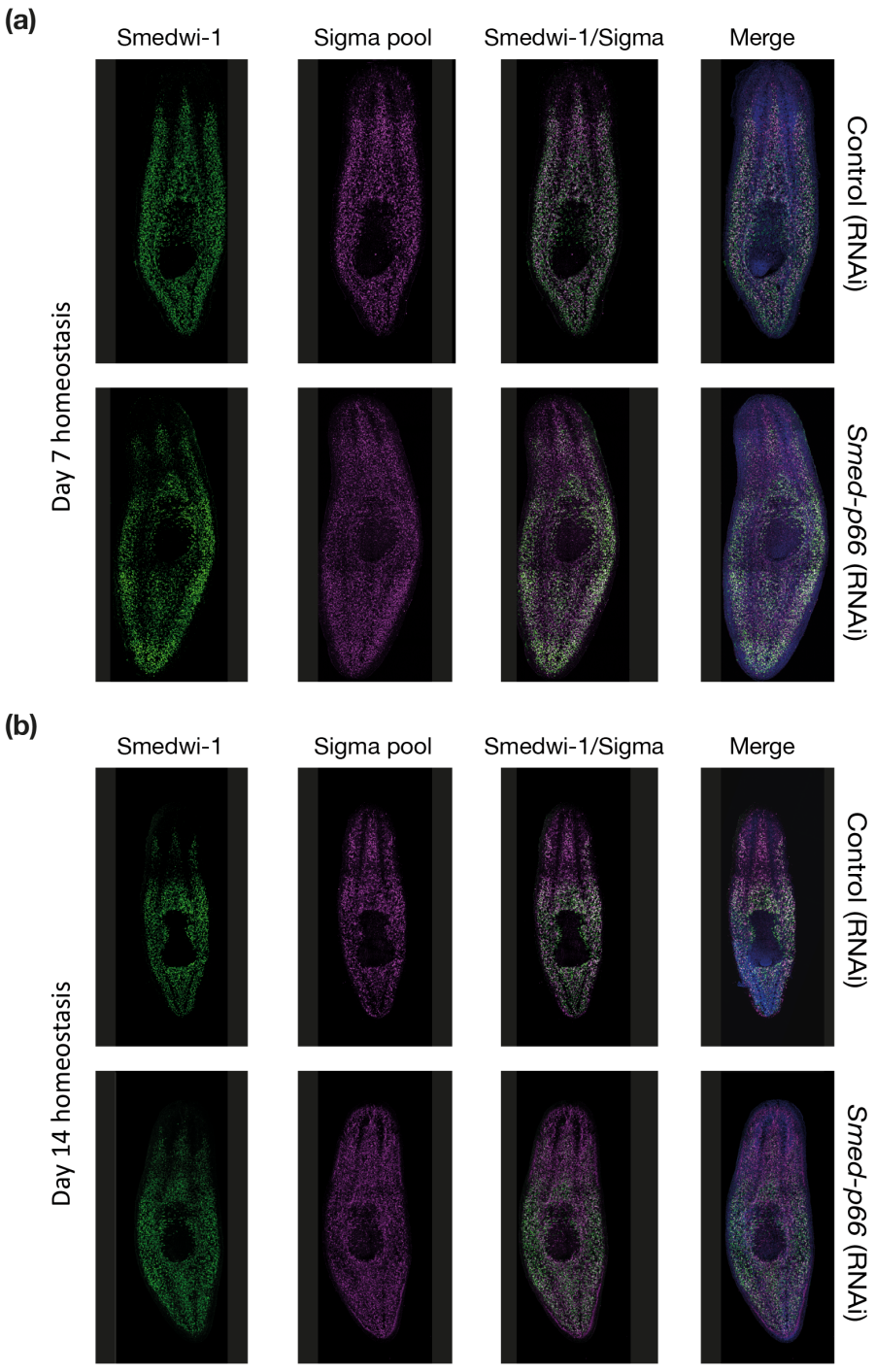

Figure 2.18: Sigma-class neoblast maintenance is not affected by knockdown of *Smed-p66*. Lineage markers labelling sigma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distributions of sigma-neoblasts at the two time points appear similar.

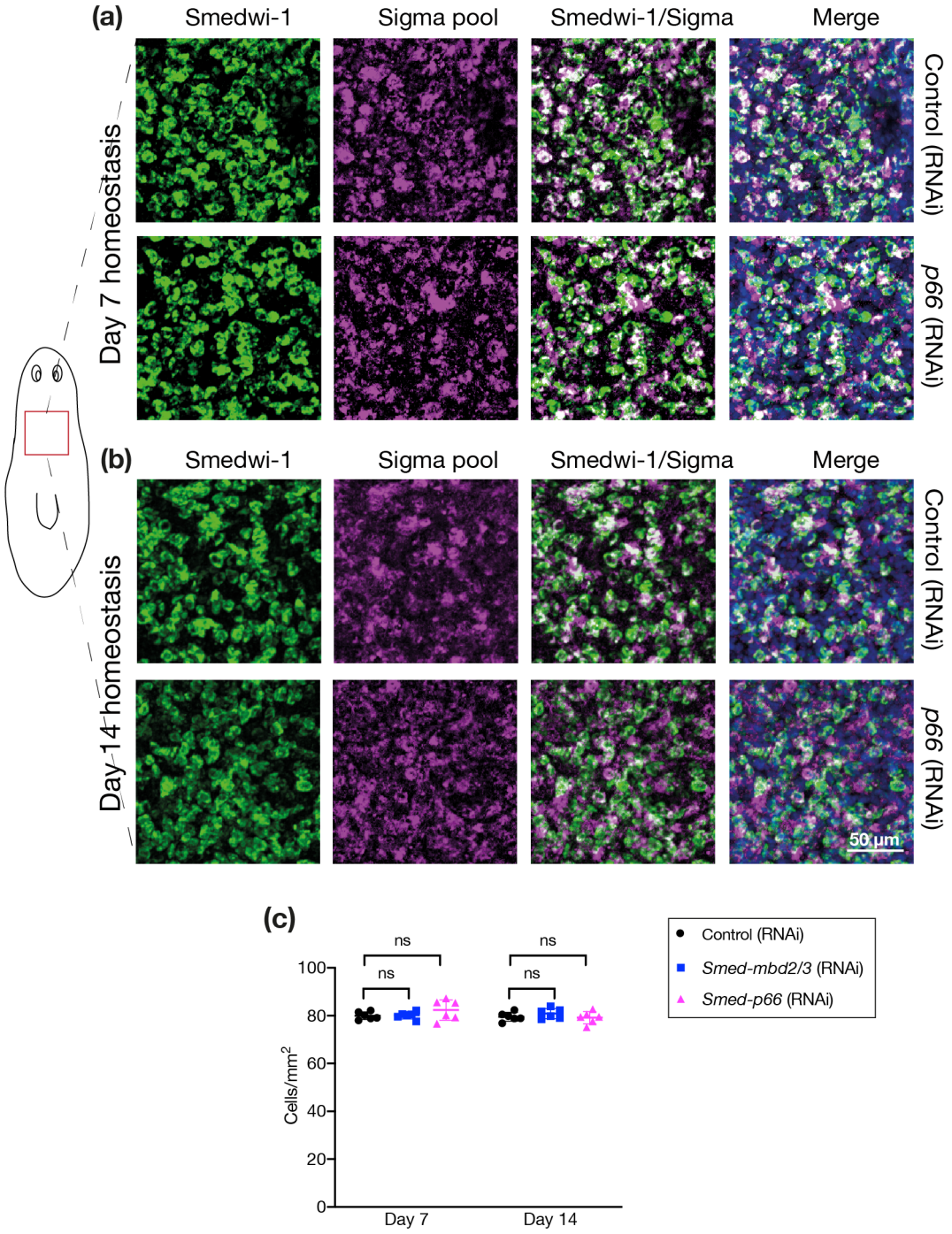

Figure 2.19: Maintenance of sigma-class neoblasts is not affected by knockdown of *Smed-p66*. Lineage markers labelling sigma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of sigma-neoblasts normalised to area 7 days and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Sigma neoblast count appear similar between *Smed-p66*(RNAi) and control animals during homeostasis.

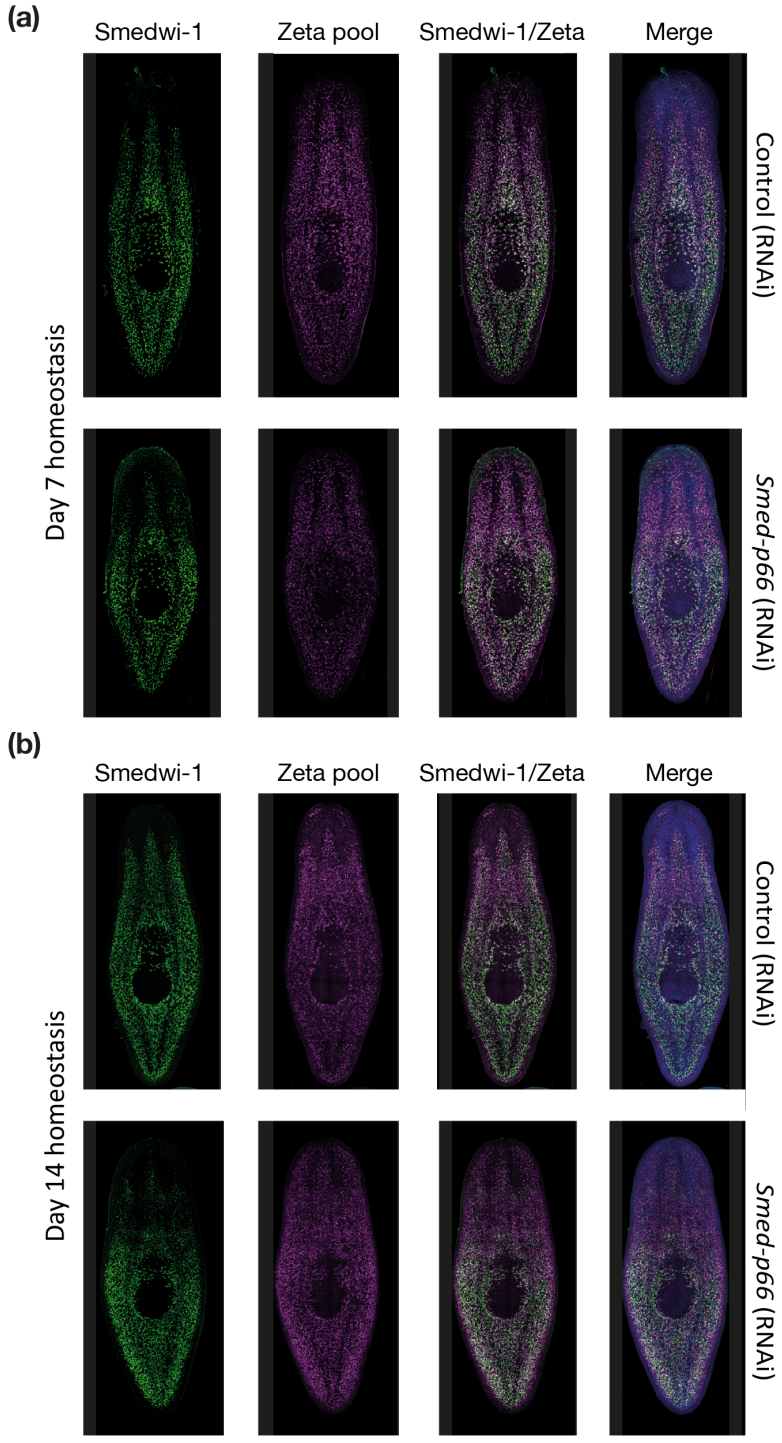

Figure 2.20: Zeta-class neoblast maintenance is not affected by knockdown of *Smed-p66*. Lineage markers labelling zeta-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distribution of zeta-neoblasts at the two time points appear similar.

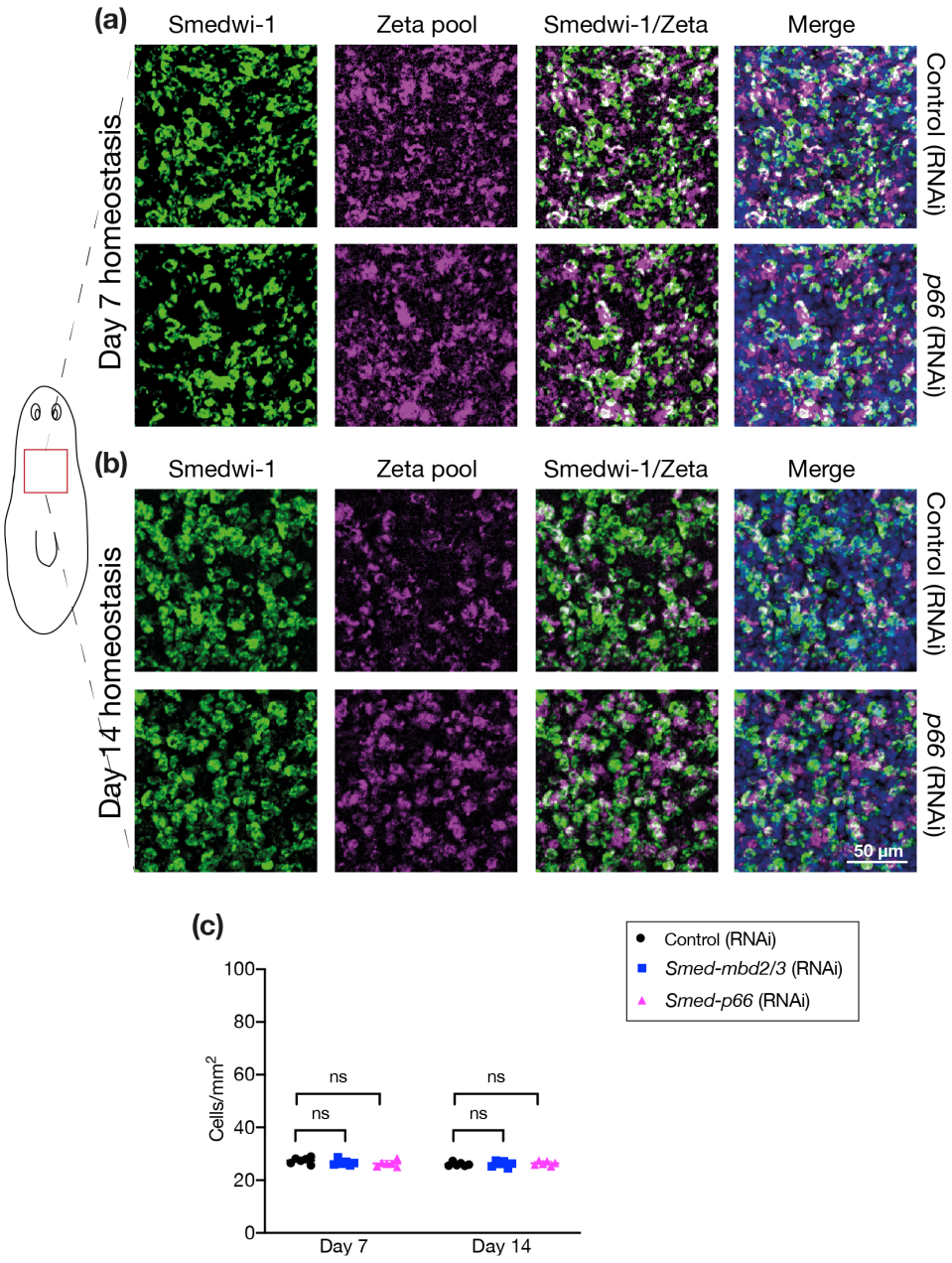

Figure 2.21: Maintenance of zeta-class neoblasts is not affected by knockdown of *Smed-p66*. Lineage markers labelling zeta-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of zeta-neoblasts normalised to area 7 days and 14 days after knockdown during homeostasis. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Zeta neoblast counts do not differ significantly between *Smed-p66*(RNAi) and control animals during homeostasis.

were not significantly different between *Smed-p66*(RNAi) and controls (Fig. 2.23c). Together, these results suggest that *Smed-p66* does not play a role in the regulation gamma neoblasts during tissue turnover. Further experiments are required to determine whether other intestinal cells arising from the progeny of gamma neoblasts are affected.

Knockdown of *Smed-p66* causes a failure to progress through differentiation

Silencing of *Smed-mbd2/3* results in an increase in early epidermal progeny and a decrease in late epidermal progeny. A similar effect is seen after the inhibition of *Smed-p66* in eye progeny [Vásquez-Doorman and Petersen, 2016]. In order to investigate whether neoblasts are able to differentiate along the epidermal lineage, early epidermal progeny cells (*prog-1*$^+$) and late epidermal progeny cells (*agat-1*$^+$) were labelled. Although *Smed-p66*(RNAi) does not affect neoblast maintenance, it lead to an accumulation of early epidermal progenitors (*prog-1*$^+$). After 7 days of homeostasis, *Smed-p66*(RNAi) animals showed significant changes in the amounts of *prog-1* and *agat-1* expressing cells (Fig. 2.24). After 14 days of homeostasis, these differences were more pronounced (Fig. 2.24b). The control animals exhibited a uniform distribution of *prog-1* and *agat-1* expressing cells at both time points.

In order to quantify this, the numbers of each of the two cell types were counted in the pre-pharyngeal region and normalised to the area. The numbers of *prog-1*$^+$ and *agat-1*$^+$ cells were significantly different between *Smed-p66*(RNAi) and controls (Fig. 2.24c) after 7 and 14 days (Fig. 2.24d). This results differ from those found in Vásquez-Doorman and Petersen [2016], where no detectable differences in the *prog-1* was seen.

These results indicate that *Smed-p66* does not affect zeta neoblast maintenance or early differentiation during tissue turnover, but prevents cells from progressing through differentiation. New tissue made after p66 knockdown are unpigmented or less pigmented, suggesting that the epidermis, although formed, is not quite the same as in wild-type worms. Although a smaller number of *agat-1*$^+$ cells is produced, other cells retain the ability to make an epidermis. Together, this suggests that like the

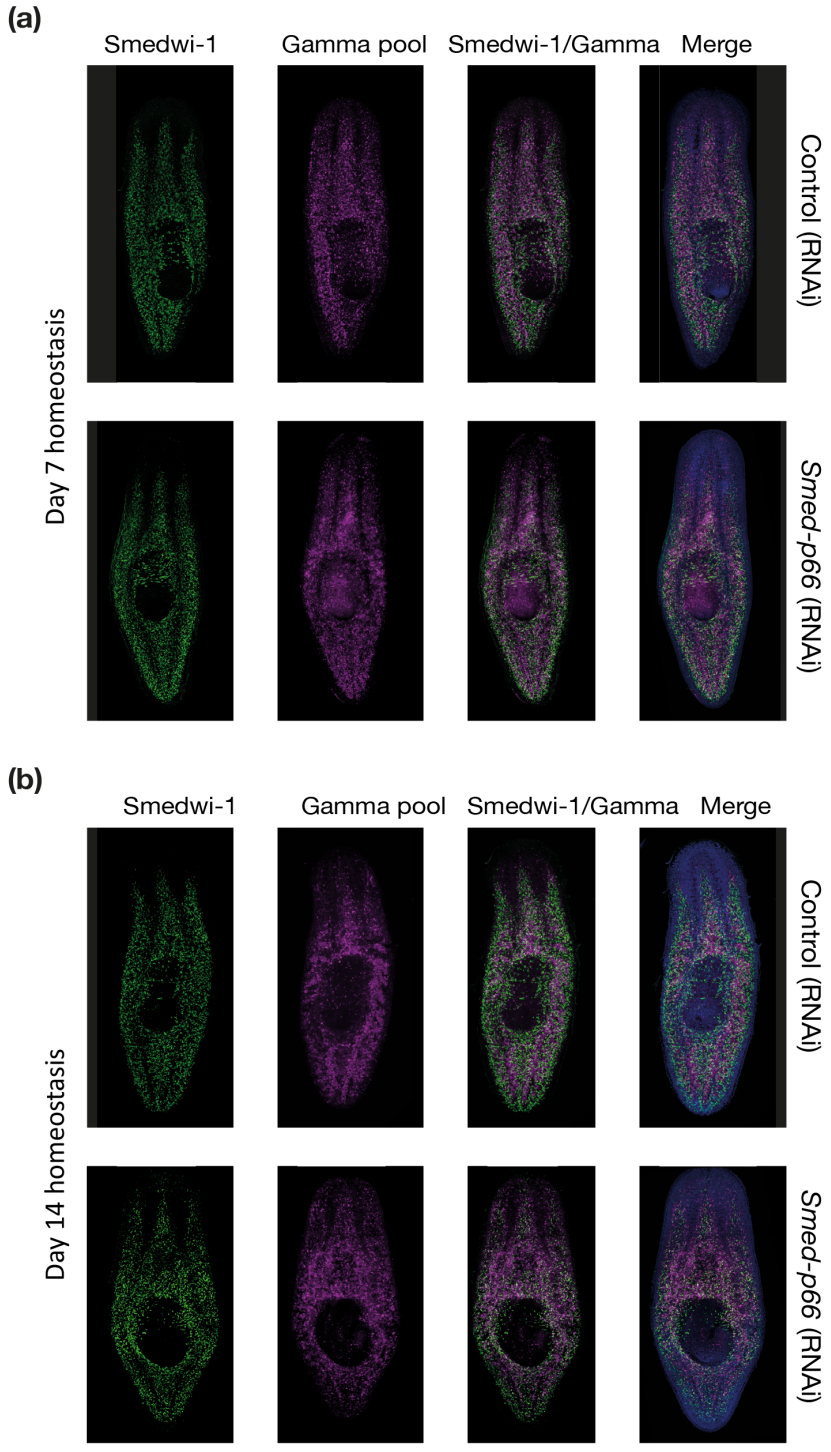

Figure 2.22: Gamma-class neoblast maintenance is not affected by knockdown of *Smed-p66*. Lineage markers labelling gamma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. Gross numbers and distributions of gamma-neoblasts at the two time points appear similar.

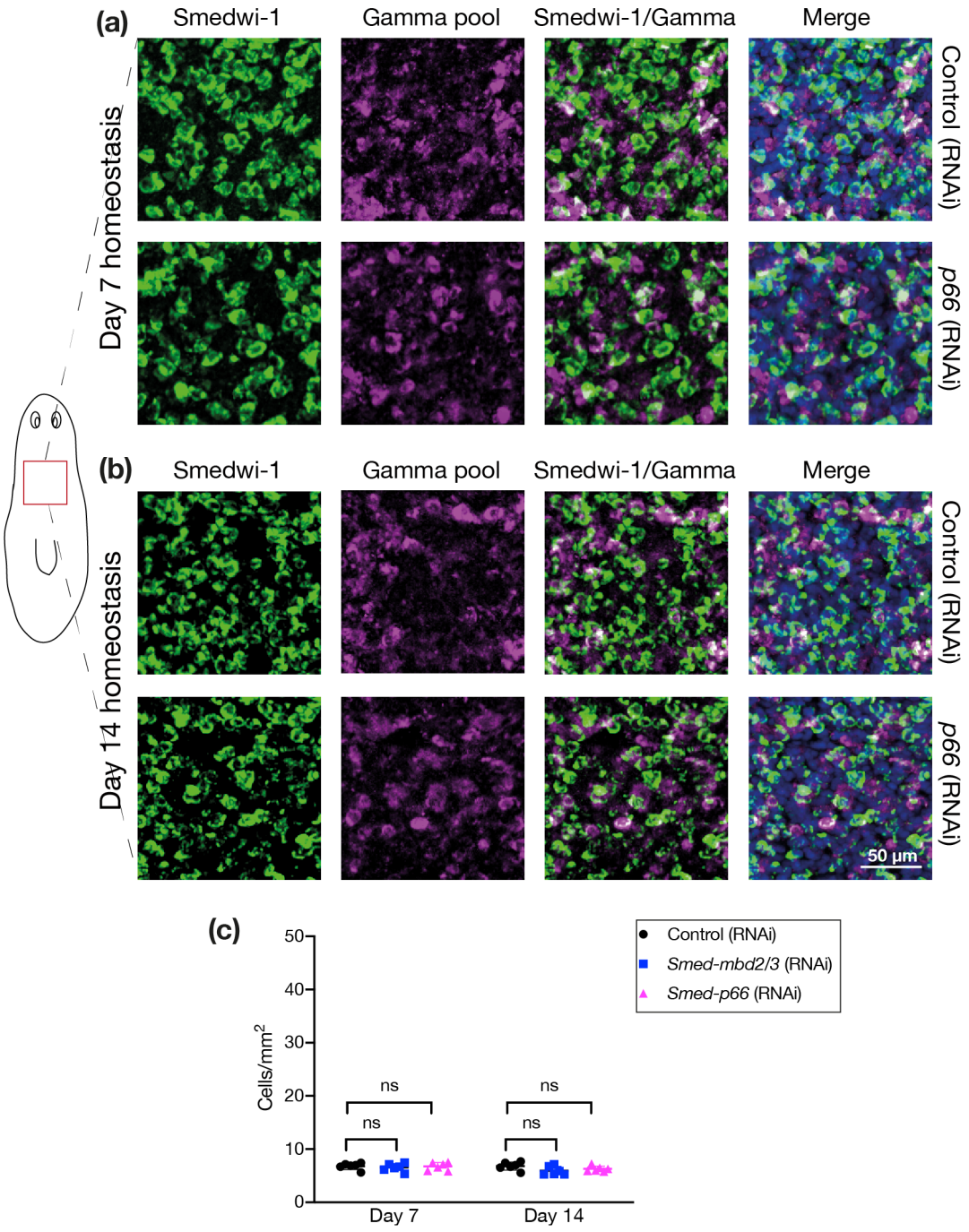

Figure 2.23: Knockdown of *Smed-p66* does not affect gamma-class neoblasts. Lineage markers labelling gamma-neoblasts in unamputated animals (a) 7 days and (b) 14 days after knockdown. (c) Graph illustrating relative number of gamma-neoblasts normalised to area 7 days and 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'. Gamma neoblast counts do not differ significantly between *Smed-p66*(RNAi) and control animals during homeostasis.

Smed-mbd2/3 knockdown worms, *Smed-p66*(RNAi) worms are also able to initiate differentiation but not complete it.

Role of metastasis-associated proteins in planarian regeneration

S. mediterranea has two MTA1 like genes – MTA1-like-1 (MTA1-1) and MTA1-like-2 (MTA1-2) [Dattani et al., 2018b]. Along with their function as an epigenetic regulator associated with the NuRD complex, MTA proteins play an important role in the development of the brain and eye, haematopoiesis, and differentiation [Sen et al., 2014]. In order to ascertain whether *Smed-mta1-1* and *Smed-mta1-2* are essential to regeneration, I performed knockdown using RNAi. Animals were injected with dsRNA corresponding to *Smed-mta1-1*, *Smed-mta1-2* and *gfp* (as control) 6 times. In each case, 10 animals were injected and then amputated into 3 pieces – head, trunk, and tail. After knockdown of *Smed-mta1-1*, the regenerating pieces showed no phenotypic changes in the three regeneration time points at day 4, day 7 and day 10 (shown in Fig. 2.25a). The viability of the animals was also unaffected by knockdown (Fig. 2.25b). In contrast, *Smed-mta1-2*(RNAi) animals were unable to regenerate and replace missing tissue (Fig. 2.25a).

At day 4, *Smed-mta1-2*(RNAi) regenerating pieces showed very small blastema. The differences were more apparent at day 7, where 27% of pieces had undergone lysis. The surviving head pieces showed head regression, and pieces were unable to form missing anterior or posterior structures. After 10 days, surviving pieces were curled up and had slowed mobility. Knockdown of *Smed-mta1-2* led to a complete loss of viability amongst all worms (Fig. 2.25b). The control animals showed no defects in regeneration and all survived up to day 21 while most of the *Smed-mta1-2*(RNAi) animals were dead by this point. The lack of apparent defects after the knockdown of *Smed-mta1-1* could be due to a functional redundancy in the MTA genes. Further experiments, including a molecular characterisation using different markers for neoblasts and their progeny after knockdown of both mta1-like genes, will help determine changes, if any, at the cellular level.

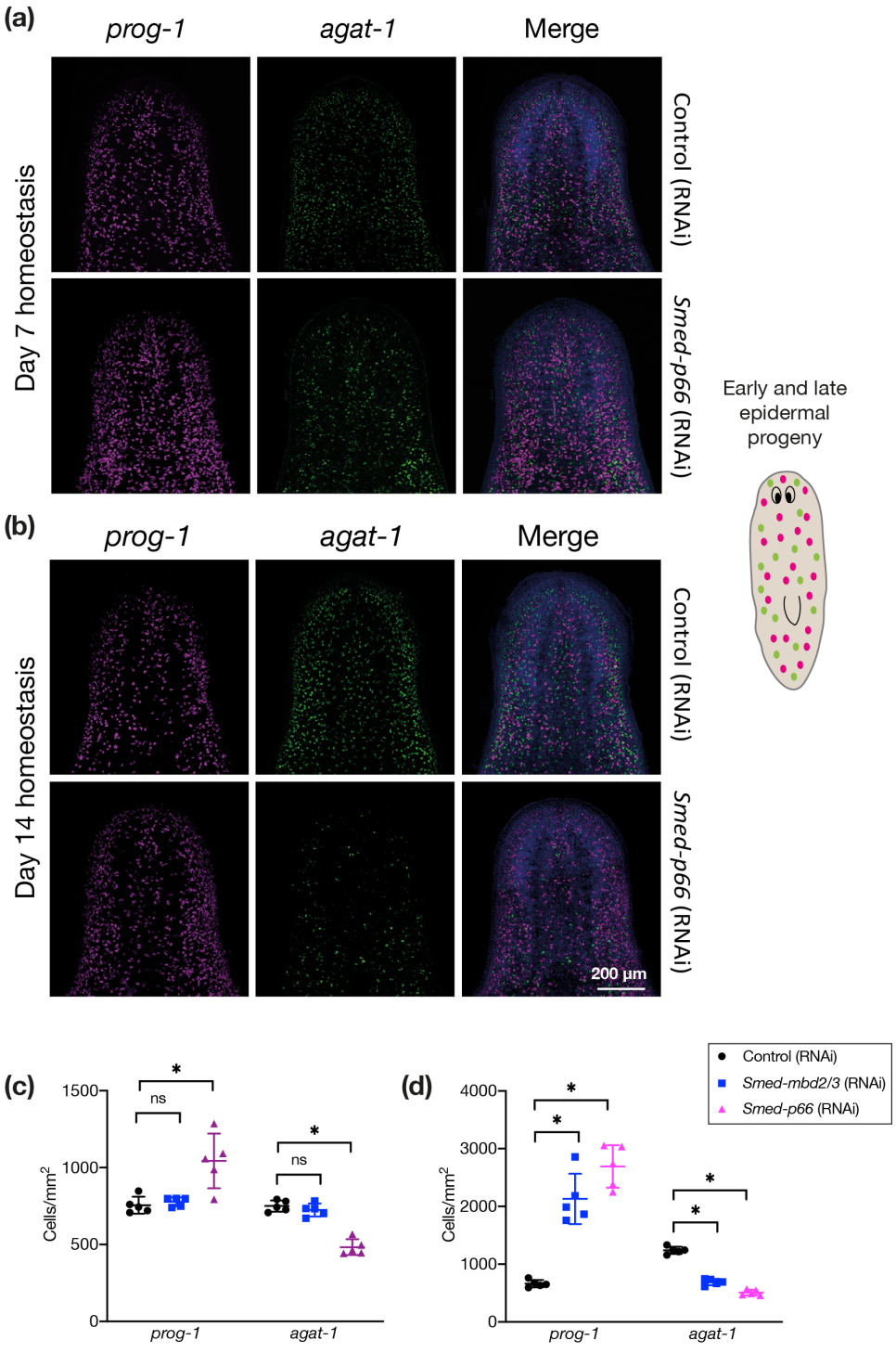

Figure 2.24: Differentiation of epidermal progenitor cells is perturbed following knockdown of *Smed-p66*. Expression of *prog-1* and *agat-1* in unamputated *Smed-p66*(RNAi) and control animals (a) 7 days and (b) 14 days after knockdown. Graph illustrating mean number of each cell type normalised to area (c) 7 days and (d) 14 days after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

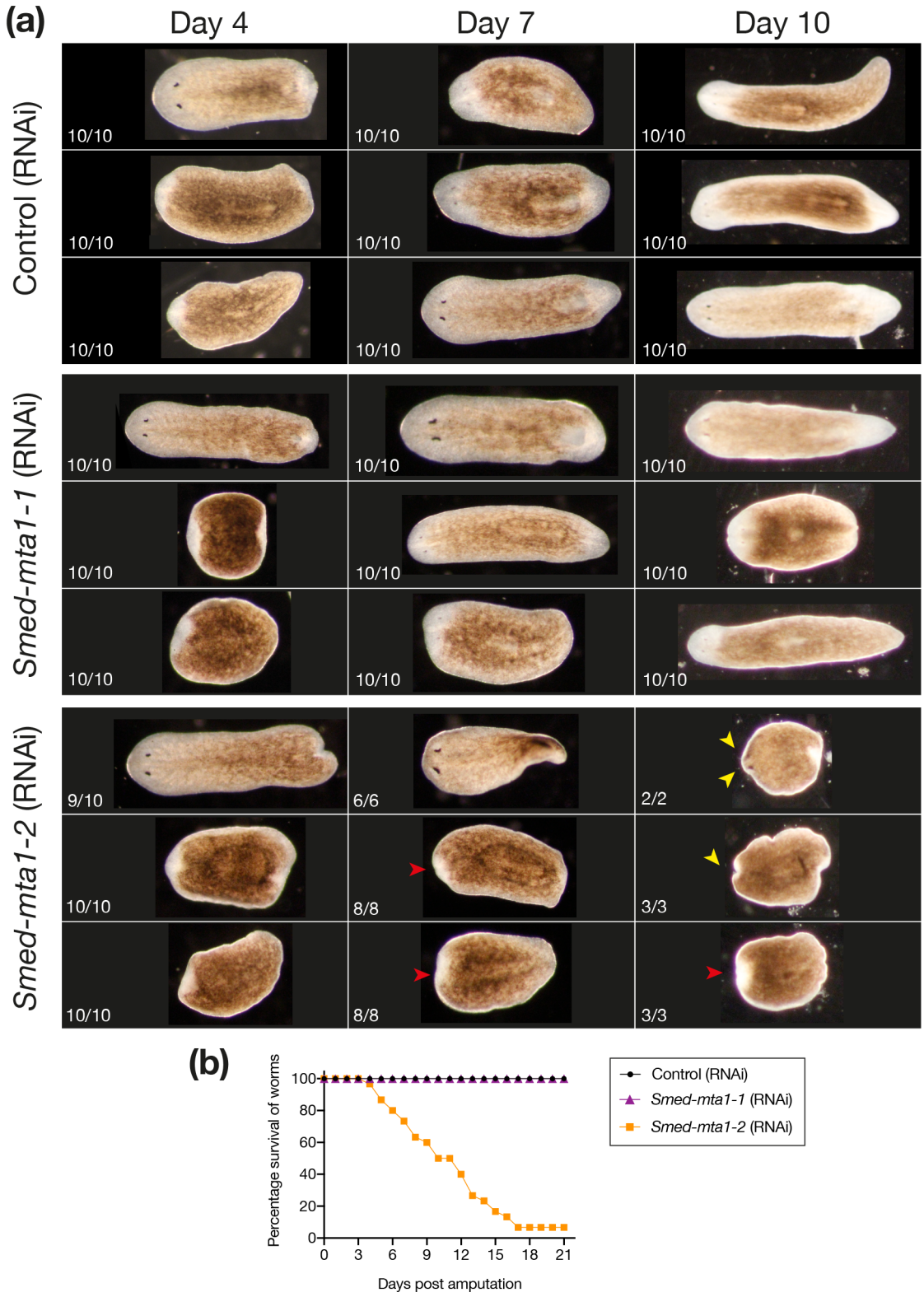

Figure 2.25: The effects of silencing of MTA genes on planarian regeneration. (a) Head, middle and tail pieces during regeneration after *Smed-mta1-1*, *Smed-mta1-2* and control knockdown. White numbers next to each piece indicate proportion of animals exhibiting the same phenotype at this time point. (b) Survival curves of knockdown animals during regeneration. Yellow and red arrowheads respectively indicate head regression and no eyes.

Chromodomain helicase DNA binding proteins in planarians

The NuRD complex is implicated in ATP-dependent chromatin remodelling as a result of the mutually-exclusive chromodomain helicase-DNA-binding paralogous subunits CHD3 (Mi-2α), CHD4 (Mi-2β) and CHD5. The *Smed-chd4* gene was identified and found to be required for planarian regeneration and tissue homeostasis [Scimone et al., 2010]. The number of neoblasts were normal following the knockdown of *Smed-chd4*, despite an inability of the animals to regenerate, mirroring the *Smed-mbd2/3* phenotype. However, *Smed-chd4*(RNAi) animals had reduced numbers of both *prog-1* and *agat-1* cells. Together, these results indicated that *Smed-chd4* is required for the maintenance and differentiation of neoblasts; this is crucial for controlling tissue turnover and regeneration, and suggests that CHD4 has a vital role in stem cell differentiation [Scimone et al., 2010]. The role of CHD3/CHD5 in planarians is yet to be elucidated.

To determine the role of *Smed-chd3*, I knocked down the gene using RNAi. Animals were injected with dsRNA corresponding to *Smed-chd3*, *Smed-chd4* and *gfp* (as control) 6 times. In each case, 10 animals were injected and then amputated into 3 pieces – head, trunk, and tail. In the regeneration assay, no visible defects were visible after knockdown of *Smed-chd3*, while phenotypic defects due to knockdown of *Smed-chd4* were apparent even at early regeneration stages. Three regeneration time points (day 4, day 7 and day 10) are shown in Fig. 2.26a for comparison, and the viability of animals is shown in Fig. 2.26b. *Smed-chd4*(RNAi) animals were unable to form any visible blastema at day 4. Head pieces showed signs of tissue damage and oedema. At day 7, the differences between control and *chd4*(RNAi) worms was more obvious with oedemas observed on head, trunk, and tail pieces. After 10 days, surviving pieces showed damage of the anterior tip and lateral margins, had slowed mobility, and exhibited reduced oedema formation. Control worms had formed unpigmented structures replacing lost tissue including eyes, while *Smed-chd4*(RNAi) animals show more tissue damage and, in some cases, lysis. Only 30% of worms survived to day 10 (Fig. 2.26b) and none of the pieces were able to complete regeneration.

The lack of apparent defects after the knockdown of *Smed-chd3* could be due to a functional redundancy in the different chromodomain helicase DNA binding proteins.

Fluid-filled oedemas, observed after *Smed-chd4* knockdown, are generally indicative of a defect in maintenance of protonephridia [Lin and Pearson, 2014, Rink et al., 2011]. Further experiments, including labelling for different neoblasts and their progeny after knockdown of both genes, will help determine changes at the cellular level as well as establish the cause for the formation of oedema.

2.3 Conclusion

The NuRD complex plays an important role in the control of transcription, stem cell self-renewal, and lineage commitment. The functions of the NuRD complex components *Smed-mbd2/3*, *Smed-hdac1*, *Smed-p66*, *Smed-chd3*, *Smed-chd4* and *Smed-mta1-like1/2* were investigated in planarians using RNA interference. *Smed-mbd2/3*, *Smed-hdac1*, *Smed-p66*, *Smed-chd4* and *Smed-mta1-like 2* were found to be essential for planarian regeneration. *Smed-mbd2/3*, *Smed-hdac1*, *Smed-p66* were found to be essential for normal tissue turnover during homeostasis as well. No visible defects were seen after the inhibition of *Smed-chd3* and *Smed-mta1-like1*, which could be due to functional redundancies with other closely related chromo-helicase domain and metastasis-associated genes.

During regeneration, *Smed-hdac1* and *Smed-chd4* animals did not form blastema, while *Smed-mbd2/3* and *Smed-p66* animals formed small blastemas. Silencing of *Smed-mbd2/3* and *Smed-p66* led to the abrogation of stem cell differentiation without affecting stem cell maintenance, while knockdown of *Smed-hdac1* and *Smed-chd4* led to defects in both stem cell maintenance and differentiation [Scimone et al., 2010]. Knockdown of *Smed-mbd2/3* and *Smed-p66* did not affect sigma or lineage-committed zeta and gamma neoblasts. As expected, a drastic reduction in sigma, zeta and gamma neoblasts was seen after the inhibition of *Smed-hdac1*. Animals had reduced numbers of both early (*prog-1*) and late epidermal progeny (*agat-1*) after the knockdown of *Smed-hdac1*, while knockdown of *Smed-mbd2/3* and *Smed-p66* led to the accumulation of *prog-1* and reduction of *agat-1* cells. Additionally, there was an increase in photoreceptor neurons after the inhibition of *p66* without effecting the production of eye pigment cup cells. This suggests that *p66* is involved in the suppression of photoreceptor neuron in

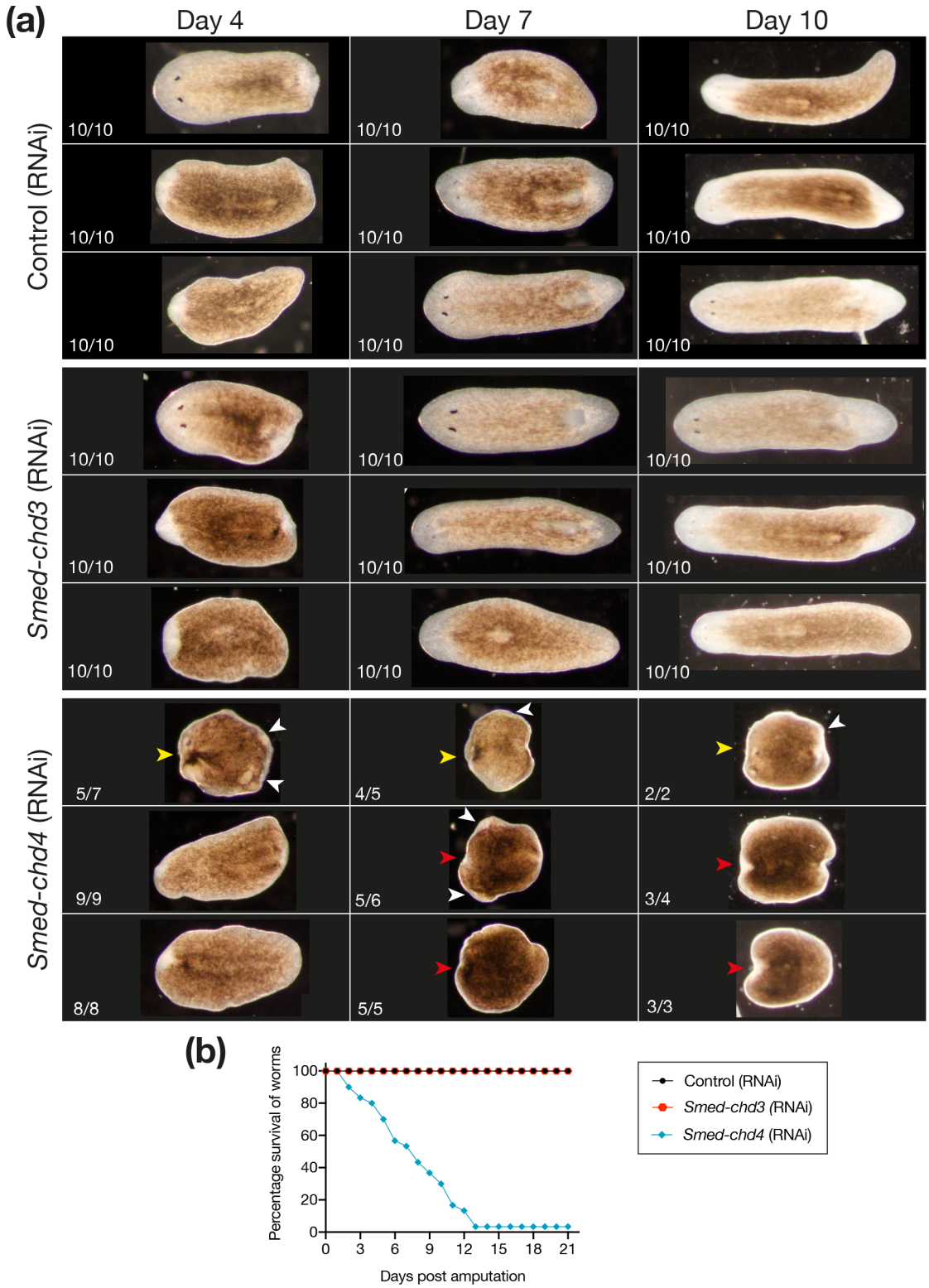

Figure 2.26: The effect of inhibition of CHD3 and CHD4 genes on planarian regeneration. (a) Head, middle and tail pieces during regeneration after *Smed-chd3*, *Smed-chd4* and control knockdown. White numbers next to each piece indicate proportion of animals exhibiting the same phenotype at this time point. (b) Survival curves of knockdown animals during regeneration. Yellow, red and white arrowheads respectively indicate head regression, no eyes and oedema.

planarians [Vásquez-Doorman and Petersen, 2016]. These results indicate that these genes have distinct functions in lineage commitment – *Smed-chd4* and *Smed-hdac1* are involved early in the process while *Smed-mbd2/3* and *Smed-p66* function at a later point in differentiation [Dattani et al., 2018b].

The conflicting phenotypes observed after the inhibition of various components of the NuRD complex is due to the fact that most of these proteins are also involved in other complexes. Silencing of components with roles in other complexes leads to a stronger phenotype. On the contrary, components that are present only in the NuRD complex, mainly *Smed-mbd2/3* and *Smed-p66*, elicit comparatively weaker phenotypes. Further investigations of all the phenotypes using newly identified lineage markers will help identify whether these NuRD components are involved in different stages of cellular differentiation.

Therefore, I chose Smed-mbd2/3 for further experiments to investigate the targets of the NuRD complex. Studying the role of MBD2/3 in a DNA methylation-free organism like planaria is of significance as it might help clarify a DNA-methylation-independent role of MBD2/3.

2.4 Methods

Planarian culture

An asexual strain of *Schmidtea mediterranea* was used in this study and maintained in 0.5 % instant ocean salts. Planarians are fed organic calf liver twice a week and cultured at 20 °C in the dark. Planarians were starved 7 days prior to all experiments.

Cloning of planarian genes

Gene specific primers with overhangs homologous to pPR-T4P vector (gift from Dr. Jochen Rink, Max Planck Institute) were used for PCR amplification from cDNA. The pPR-T4P vector is based on pPR242 and contains two T7 promoters and two T7 terminators (flanking the gatewayT4P cassette). The vector is linearised by digesting with Sma I restriction enzyme and then treated with T4 DNA polymerase. PCR products were treated with T4 DNA polymerase, mixed with linearised vector, and incubated

for 15 min at room temperature. Ligated products were transformed directly into *Escherichia coli* strain DH5-Alpha, positive colonies were then verified by PCR and sequencing. Planarian *mbd2/3*, *p66*, *hdac1*, *mta1-like-1*, *mta1-like-2*, *chd3* and *chd4* genes were cloned into the pPR-T4P vector with a product size of $600 - 800$ bp.

Generation of double-stranded RNA

DNA fragments of interest were cloned into pPR-T4P vector and amplified by PCR using M13 primers. The purified PCR product was used to set up an in vitro transcription reaction, with T7 RNA polymerase (37 °C for 6 hours). Samples were treated with DNase to degrade the PCR template and precipitated using ethanol-sodium acetate. Re-suspended pellets were incubated at 68 °C for 15 minutes, and at 37 °C for a further 45 minutes. The synthesised double-stranded RNA (dsRNA) was checked on a 1 % agarose gel and stored at −20 °C.

Riboprobe synthesis

DNA fragments of interest were cloned into either pPR-T4P or pCR II vectors and amplified by PCR. Digoxigenin (DIG) and Fluorescein (FITC) labelled anti-sense RNA probes were synthesised using an in vitro transcription reaction. The transcription reaction was performed using SP6/T7 polymerase (Roche) depending on the orientation of the insert as identified by sequencing. The probes were precipitated with LiCl/ethanol and re-suspended in 25 µl RNase free water. Probe quality was assessed on a 1 % agarose gel and probes were stored at −20 °C.

RNA interference

The dsRNA was administered to the animals via microinjections using a Nanoject II (Drummond Scientific) with 3.5" Drummond Scientific (Harvard Apparatus) glass capillaries pulled into fine needles on a micropipette puller. Animals were placed on a piece of dark cardboard in a petri dish of flattened ice to reduce movement. The animals were turned with their ventral side pointing upward and injected in pre-pharyngeal region. Depending on the size of the worm, 96.6 nl to 128.8 nl of dsRNA was injected into

each worm on each injection day. Each animal was injected with gene-specific dsRNA. Injections were repeated for 6 days over 2 weeks. dsRNA of GFP gene was used as control. All animals were injected under the same conditions.

In situ hybridisation

Whole mount fluorescent in situ hybridization (FISH) was performed as described previously [King and Newmark, 2013] and Fast Blue development followed a protocol adapted from [Lauter et al., 2011]. In brief, experimental animals were fixed using formaldehyde and treated with the riboprobe mix. Anti-DIG/anti-FITC antibodies conjugated with alkaline phosphatase or horse radish peroxidase was diluted in blocking solution ($1 : 2000$) and incubated at 4 °C overnight. The reaction was developed using either Fast Blue (colourimetric and fluorescent) or TSA (fluorescent) reaction. Nuclei were stained with Hoechst 33342 (Sigma).

Imaging

Bright-field images of worms were captured on a Zeiss SteREO Discovery V8 microscope. Fluorescent animals were imaged on Inverted Olympus Fluoview FV1000 or FV1200 confocal systems. Images were processed or cropped in Fiji [Schindelin et al., 2012] and representative images were used in figures.

Cell counting

Samples were imaged on a confocal microscope. A z-projection of the resultant stacks were created using Fiji [Schindelin et al., 2012]. Fiji's cell counter was used to count the number of cells with staining. The total number of cells was then normalised to the imaged area. This was done as follows: Area in μm^2 was measured on Fiji and was converted to mm^2. The number of cells was divided by the area in mm^2, multiplied by the number of slices in a confocal stack. The thickness of the distance between slices was kept constant, at 4 µm for all imaged stacks.

Statistical analysis

Prism 8.0 (`https://www.graphpad.com/`) was used to calculate statistical significance and plot curves. Results are expressed as mean $\pm$ standard deviation (SD). Statistical analyses were performed using Student's t-test using a statistical significance of $p < 0.05$. Wherever cell numbers were compared between experimental conditions, a two-tailed t-test assuming unequal variance was used.

6 animals were used for counting analysis of sigma, zeta and gamma as well as *prog-1* and *agat-1* cells. 21 animals were used for *smedwi-1* counting analysis.

3

MBD2/3 mediated transcriptional regulation in planarian cells

Contents

3.1 Introduction

Experiments in planarians have shown that knockdown of the orthologues of mammalian epigenetic regulators can lead to diverse stem cell defects and errors in lineage commitment of stem cells, resulting in a loss of regenerative capacity. In Chapter 2, the functions of three NuRD complex components in planarians, MBD2/3, HDAC1 and GATA2D/p66, were investigated in detail. Knockdown of *Smed-mbd2/3*, *Smed-hdac1*, *Smed-p66*, *Smed-chd4* and *Smed-RbAp48* leads to the abrogation of stem cell differentiation but affected stem cell pluripotency as well as differentiation differently [Scimone et al., 2010, Jaber-Hijazi et al., 2013, Robb and Alvarado, 2014, Hubert et al., 2015, Vásquez-Doorman and Petersen, 2016]. Inhibition of *Smed-mbd2/3* and *Smed-p66* led to the accumulation of early epidermal progeny (*prog*-1^+ cells) but a reduction in late epidermal progeny (*agat*-1^+ cells), whereas *Smed-chd4* and *Smed-hdac1* knockdown animals lost both *prog*-1^+ and *agat*-1^+ cells. These proteins, therefore, have distinct functions in the determination of stem cell fate, implying that MBD2/3 and p66 are essential later in differentiation, whereas CHD4 and HDAC1 are required much earlier. A reduction in mitotic activity is seen after *Smed-chd4*, *Smed-RbAp48* and *Smed-p66* knockdown whereas *Smed-hdac1* RNAi animals show a complete loss. Interestingly, *Smed-mbd2/3* (RNAi) animals show no reduction in mitotic activity [Jaber-Hijazi et al., 2013]. A reduction in number of stem cells is observed after knockdown of *Smed-hdac1*, *Smed-RbAp48* and *Smed-chd4* whereas no change was seen after the knockdown of *Smed-mbd2/3* or *Smed-p66* [Eisenhoffer et al., 2008, Scimone et al., 2010, Jaber-Hijazi et al., 2013, Robb and Alvarado, 2014, Vásquez-Doorman and Petersen, 2016]. *Smed-hdac1*, *Smed-RbAp48* and *Smed-chd4* knockdown animals were unable to form proper blastema, while *Smed-mbd2/3* and *Smed-p66* RNAi animals formed small blastema. These varying RNAi phenotypes of different NuRD complex components can be attributed to their roles in other complexes. Proteins that are involved in other complexes have a severe phenotype, whereas those unique to NuRD, mainly MBD2/3 and p66, only show impaired differentiation. Mammalian RbAp48 and HDAC1 are members of the Sin3 deacetylase complex which, along with *Nanog*, is involved in the activation of pluripotency factors and suppression of differentiation

genes. HDAC1 is also involved with CoREST/REST, NCoR/SMRT and SHIP1 [Choi et al., 2008, Hayakawa and Nakayama, 2011]. It is known that Mi-2/CHD4 and RbAp46 are involved in other chromatin-modifying complexes such CAF-1 and dMec [Kunert et al., 2009, Mcdonel et al., 2009], so their roles in planarian stem cell maintenance may also be associated with functions within these complexes [Aboobaker, 2011]. The role of these proteins in stem cell differentiation, however, appears related to their participation in the NuRD complex. This view is supported by the observation that the knockdown of *Smed-mbd2/3*, the only gene exclusive to the NuRD complex, led to impaired neoblast differentiation but intact neoblast self-renewal.

An understanding of how these epigenetic regulators affect gene regulation will help identify targetable mechanisms, contributing to a better understanding of stem cell fate. To develop a comprehensive understanding of *mbd2/3*-mediated transcriptional regulation, gene expression profiles of control and *Smed-mbd2/3*(RNAi) animals will be compared. These experiments will define the transcriptional differences caused by the inhibition of *Smed-mbd2/3* and help broadly identify differentiation factors and assess the level of underlying conservation across animals.

Transcriptional profiling of planarian stem cells

Cells morph from one type into another during development, giving rise to different lineages and differentiated cell types. An example of the importance of regulation is the reprogramming of differentiated somatic cells to induced pluripotent stem cells (iPSCs). This can be achieved by the forced expression of transcription factors (OKSM i.e. *Oct4*, *Klf4*, *Sox2*, *c-Myc* or other combinations) [Takahashi and Yamanaka, 2006], which initiates a cascade of transcriptional and epigenetic changes (such as demethylation of OKSM promoters) leading to cells becoming pluripotent [Yamanaka and Blau, 2010, Orkin and Hochedlinger, 2016].

Planarians have an indefinite capacity to replace any cell type during homeostasis or following injury owing to self-renewing pluripotent adult stem cells (neoblasts) [Aboobaker, 2011, Rink, 2013, Adler and Sánchez Alvarado, 2015]. There is evidence that planarian stem cells have an underlying pluripotency program that is conserved

with mESCs as well as stem cells of other animals. Independent studies have uncovered genes enriched in neoblasts that are also known to be expressed in ESCs and involved in the balance between self-renewal and cell fate, such as regulators and targets of *Oct4*, RNA splicing factors, epigenetic modifiers, and RNA binding proteins [Labbé et al., 2009, Önal et al., 2012, Solana et al., 2012]. The planarian neoblast transcriptome also largely mirrors that of totipotent archaeocytes of fresh-water sponge *Ephydatia fluviatilis* and multipotent stem cells of Hydra, suggesting the existence of an ancestral stem cell expression repertoire rich in RNA regulatory actors and poor in transcription factors [Solana, 2013, Alié et al., 2015].

van Wolfswinkel et al. [2014] combined single-cell profiling with functional assays to provide evidence for the existence of three distinct neoblast populations that constitute the population of cells remaining within the cell cycle. All three groups express the pan-stem cell marker *smedwi-1* transcript, an orthologue of the PIWI family proteins associated with germ line maintenance. Sigma neoblasts were found to proliferate in response to injury and have broad differentiation potential consistent with the possibility that they are pluripotent. Sigma neoblasts give rise to lineage-committed subsets – zeta and gamma [van Wolfswinkel et al., 2014]. The three neoblast classes are characterised by the expression of a distinct set of transcription factors. Recent studies using single cell technologies have lead to the identification of new cell types and a better understanding of the planarian cell types [Plass et al., 2018, Fincher et al., 2018, Zeng et al., 2018].

MBD, NuRD complex and transcriptional repression

Commitment to a specific cell fate requires a delicate balance of gene expression that represses lineage inappropriate and pluripotency genes while promoting the expression of genes required for specific cell fates. In this balance, the NuRD complex mainly acts as a transcriptional repressor, but has also been observed to play a more complex role in gene transcription, including upregulation of genes [Basta and Rauchman, 2015]. Moreover, NuRD also appears to perform other significant biological functions (such as DNA repair) due to its ability to modify the chromatin and post-translational modifications of transcription factors such as p53 acetylation status [Basta and Rauchman, 2015].

The NuRD complex is integral for the lineage commitment and differentiation of pluripotent cells. The histone deacetylation activity of NuRD complex specifically recruits polycomb repressive complex 2 (PRC2) in ESCs. NuRD-mediated deacetylation of histone H3K27 facilitates PRC2 recruitment, leading to subsequent H3K27 trimethylation at NuRD target promoters [Reynolds et al., 2012b]. In early mouse embryos, NuRD-mediated transcriptional regulation aides developmental transitions, lineage commitment in ESCs, and developmental decisions in haematopoietic and epithelial stem cells [Kaji et al., 2006, 2007, Kashiwagi et al., 2007, Yoshida et al., 2008, Mcdonel et al., 2009, Reynolds et al., 2012b]. In murine lymphocytes, Mi-2β interacts with *Ikaros* and *Aiolos*, which are zinc finger DNA-binding factors and potent repressors essential for lymphoid cell development. The NuRD complex does not form after the loss of *Mbd3*, embryonic development stops at the implantation stage, and the mutant embryos are unable to form differentiated cells [Zhang et al., 1999, Hendrich et al., 2001, Kaji et al., 2006, 2007, Reynolds et al., 2012b]. ESCs lacking Mbd3 are viable but are unable to exit self-renewal and commit to differentiation [Kaji et al., 2006]. The NuRD complex has also been shown to be essential for homeostasis of haematopoietic and epithelial stem cells [Williams et al., 2004, Kashiwagi et al., 2007]. A global misregulation in gene expression has been observed cells in the absence of MBD3-NuRD [Kaji et al., 2006, 2007, Yoshida et al., 2008, Reynolds et al., 2012b].

About 200 genes were found to be mis-regulated in *Mbd3*$^{-/-}$ ESCs [Kaji et al., 2007]. Analysis of a handful of genes mis-expressed in *Mbd3*-null ESCs identified at least 11 genes that are mis-regulated in pre-implantation *Mbd3*-null embryos. This indicates a role for NuRD-mediated transcriptional repression in the transition of pluripotent cells from a pre-implantation, potentially differentiation-resistant state, to that of the post-implantation embryo where cells are competent for lineage commitment. This role in silencing transcription of gene expression in pluripotent cells in vivo prior to developmental transitions is strikingly similar to the functions ascribed to the *Arabidopsis pickle* and *C. elegans let-418* genes, which were both shown to facilitate silencing of germline transcripts upon the initiation of embryogenesis [Mcdonel et al., 2009].

3.2 Results and discussion

MBD-NuRD is involved in the molecular mechanisms determining self-renewal versus differentiation in stem cells. To identify the transcripts that are regulated by *mbd2/3*, I performed transcriptional profiling of unamputated animals after knockdown of *Smed-mbd2/3*.

In order to ascertain transcriptional changes caused by the silencing of *Smed-mbd2/3* before the effect of the phenotype, one early and one late timepoint was chosen. The reasoning for choosing these time points is that the transcription of genes required for differentiation must have been altered by this point. From Fig. 2.3 in Chapter 2 we know that *Smed-mbd2/3* (RNAi) worms survive past 21 days. None of the worms show any visible phenotypic changes before day 18. Although no visible defects are seen before this time, it is important to check for changes at the molecular level. In Fig. 2.9, we found no significant changes in *prog-1*- and *agat-1*-expressing cells at day 7 but significant differences at day 14. In situ hybridisation for these two epidermal markers was performed at day 5 and, as expected, no changes were observed. After 5 days of homeostasis, both control and *mbd2/3* (RNAi) appear to have similar distributions of *prog-1*- and *agat-1*-expressing cells (Fig. 3.1a). In order to confirm this, the number of each of the two cell types was counted and normalised to the area. The number of *prog-1*$^+$ and *agat*-1$^+$ cells was not significantly different between *Smed-mbd2/3* (RNAi) and controls (Fig. 3.1b) after 5 days. Using these data points as a guide, I chose day 5 and day 10 as time points for RNA-seq in order to understand the effect of *mbd2/3* on global transcriptional responses of different stem cell populations.

Transcriptional response of planarian cells to loss of *mbd2/3*

Planarian cells were sorted into X1 (containing stem cells in the S/G2/M phase of the cell cycle) and X2 (containing stem cells in the G1 phase of the cell cycle and post-mitotic progenies) sub-populations after 5 and 10 days post knockdown (Fig. 3.2a). A total of 24 RNA libraries were prepared and analysed in this study – X1 and X2 cells from control and *Smed-mbd2/3*(RNAi) animals at days 5 and 10 in triplicates. Total

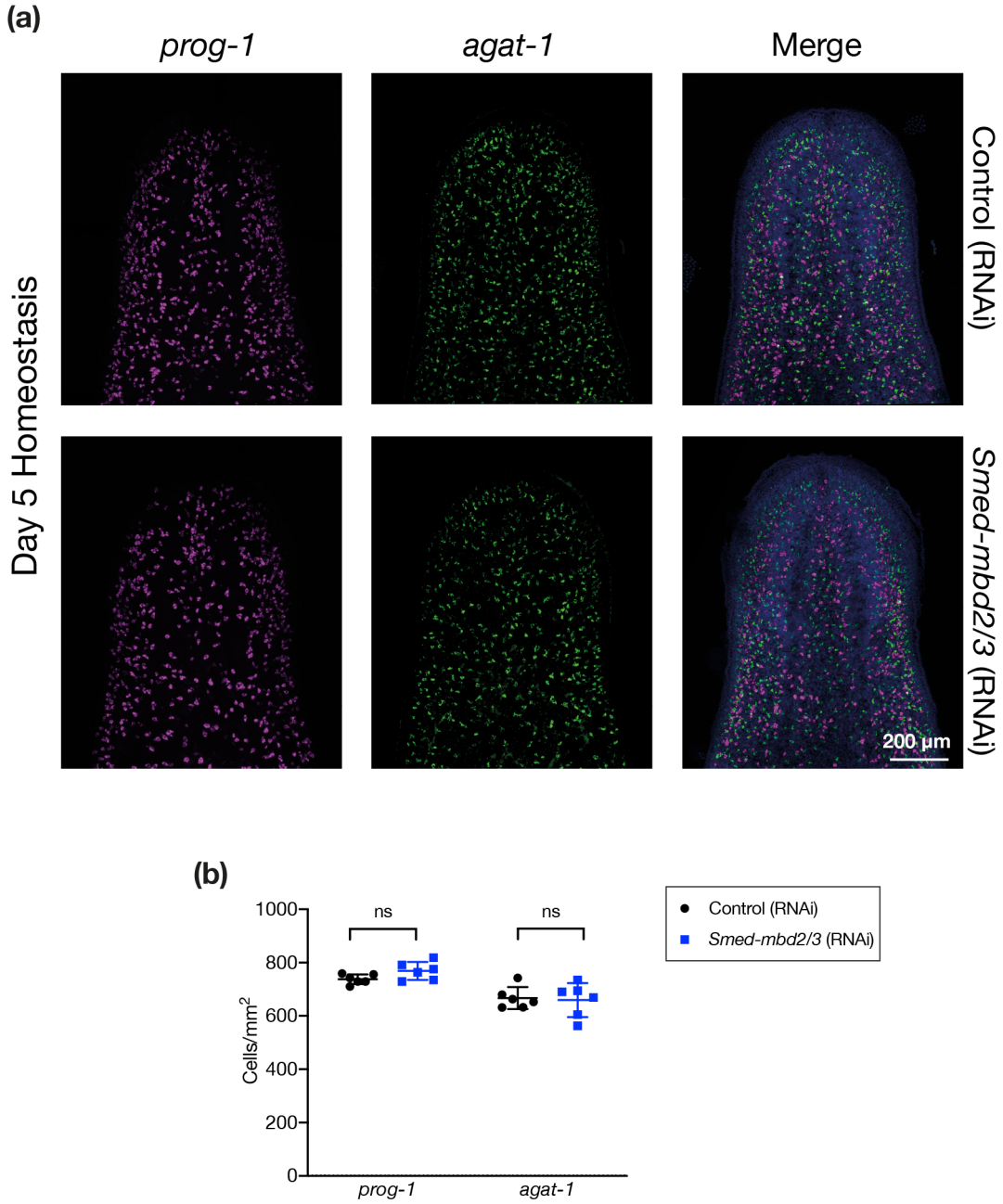

Figure 3.1: Differentiation of epidermal progenitor cells is not perturbed following knockdown of *Smed-mbd2/3* 5 days post knockdown. Expression of *prog-1* and *agat-1* in unamputated *Smed-mbd2/3* (RNAi) and control animals 5 days after knockdown. (b) Graph illustrating mean number of each cell type normalised to area after knockdown. Statistical comparisons performed via unpaired 2-tailed t tests, where * signifies $p < 0.05$ and 'ns' stands for 'not significant'.

RNA was isolated from X1 fraction and X2 fraction followed by library preparation for sequencing (Fig. 3.2a-c).

Table 3.1 shows mapping statistics, generated using Sleuth [Pimentel et al., 2017], of RNA-seq data generated in this thesis. Sleuth provides various summary tables and plots that can help identify problems with our data. The processed data shows how many reads have been sequenced (total reads), and how many were mapped to the planarian transcriptome (reads mapped), for each sample. All of the samples sequenced have > 28 million reads and thus provide good abundance estimates. For most samples, the fraction of fragments mapped exceeds 0.7, which is considered desirable. Fractions for five of the samples are below 0.7, suggesting that the samples may have become contaminated with ribosomal RNA or that the reference transcriptome is incomplete. In this case it is most likely the latter, as the *S. mediterranea* transcriptome has many gaps. Figure 3.3 shows a PCA plot of all 24 datasets that were sequenced. X2 libraries do not cluster as well as X1 libraries, which could be explained by the fact that the X2 cell compartment has a larger diversity of cell types than X1. Moreover, X1 and X2 datasets of *Smed-mbd2/3* knockdown cluster more closely than that of the control samples, which could be due to the misregulation of differentiation genes in X2. Overall, two replicates of each dataset appear to cluster together, which could be due to the variability in RNAi levels.

Kallisto was used to quantify abundances of transcripts sequenced and pseudo-aligned to the *S. mediterranea* transcriptome [Bray et al., 2016]. Publicly available *S. mediterranea* transcriptome, dd_smed_v6 transcriptome assembly [Brandl et al., 2016], was used to map the RNA-seq datasets. The Wald test (part of Sleuth) was used for differential gene expression analysis of RNA-seq data, for which transcript abundances have been quantified using Kallisto [Pimentel et al., 2017].

Differential analysis of the gene expression in day 5 datasets revealed 319 genes in X1 and 54 genes in X2. After 5 days of knockdown, 12 genes were downregulated (adjusted p-value < 0.05) and 307 genes were upregulated (adjusted p-value < 0.05) in stem cell population (X1 cells), while 15 genes were downregulated (adjusted p-value < 0.05) and 39 genes were upregulated (adjusted p-value < 0.05) in stem cell progeny population (X2 cells) (Fig. 3.4a-d).

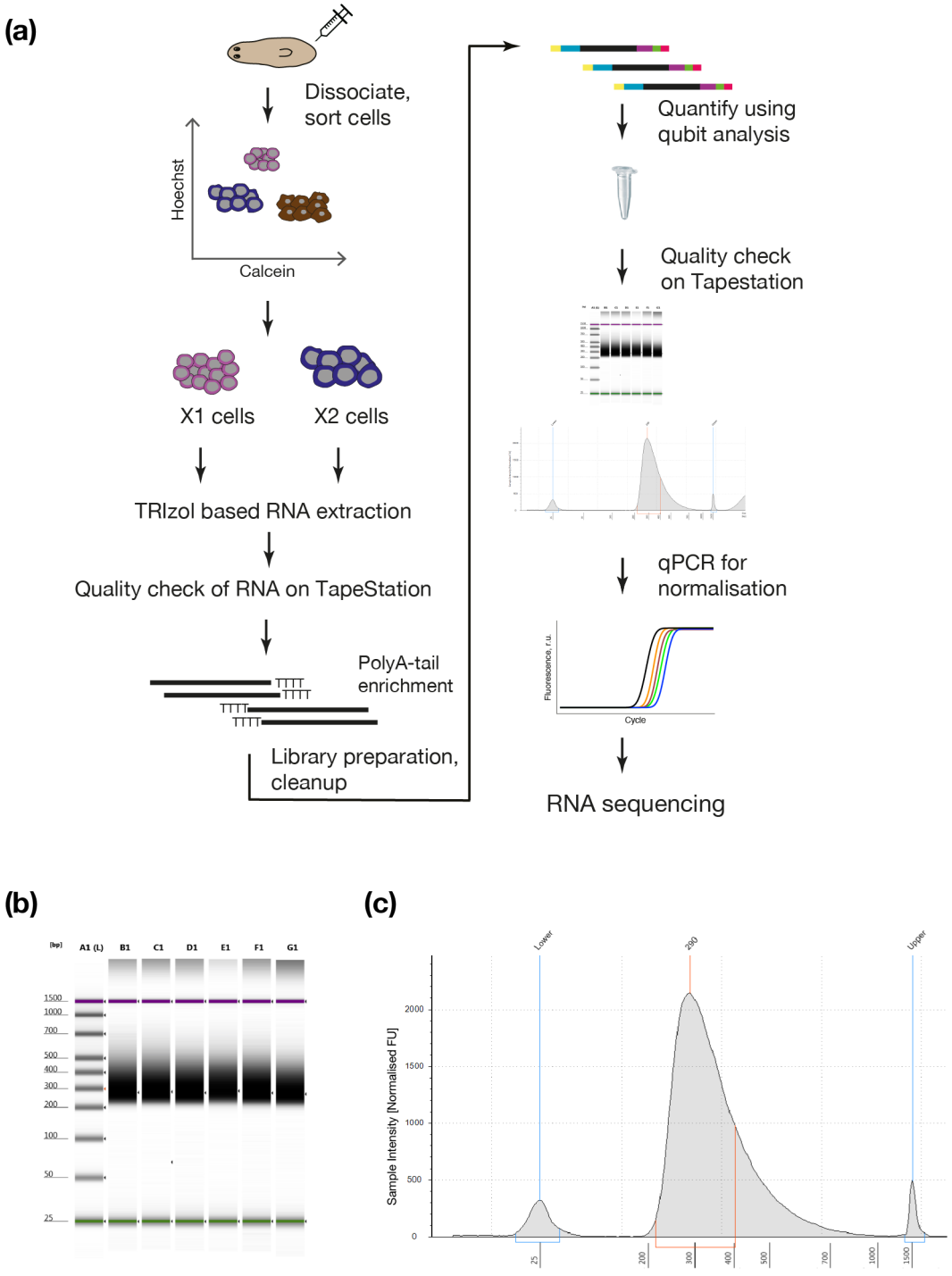

Figure 3.2: Experimental strategy to understand the transcriptional response of planarian stem cells after the inhibition of *Smed-mbd2/3*. (a) Stem cells were isolated based on nuclear to cytoplasmic content. X1 (cells from S/G2/M phase) and X2 (G1 stem cells, post mitotic progenies) population were separately collected from control and *Smed-mbd2/3* (RNAi) worms. RNA was isolated, libraries were prepared and sequenced on a NextSeq 550 machine. (b) and (c) Quality check of libraries using Agilent TapeStation 4200. An electropherogram showing a clean band $\sim$ 280 bp with no adapter dimers from each representative experimental replicate.

			Reads		
	Condition	Replicate	mapped	total	[%] mapped
Day 5					
X1	Control (RNAi)	1	32210140	40031867	80.46
		2	36903129	48896981	75.47
		3	34696684	50479082	68.73
	MBD (RNAi)	1	29835933	45654233	65.35
		2	35846565	51125582	70.11
		3	28131249	43382627	64.84
X2	Control (RNAi)	1	32070356	38530038	83.23
		2	40026029	51285870	78.04
		3	33516965	48993565	68.41
	MBD (RNAi)	1	35351855	47452727	74.50
		2	40625271	49954572	81.32
		3	37379377	53553490	69.80
Day 10					
X1	Control (RNAi)	1	39247754	47055772	83.41
		2	35927980	42796403	83.95
		3	36995218	46443852	79.66
	MBD (RNAi)	1	34116181	48276789	70.67
		2	31778950	38452819	82.64
		3	35052100	52905935	66.25
X2	Control (RNAi)	1	47045645	59106525	79.59
		2	34153706	40003102	85.38
		3	37019411	46100605	80.30
	MBD (RNAi)	1	44220659	56777668	77.88
		2	35279283	41865412	84.27
		3	35338122	51494357	68.63

Table 3.1: Mapping statistics of RNA-sequencing data used in this study. X1 sub-population includes planarian stem cells in S/G2/M phase of cell cycle and X2 sub-population includes stem cells at G1 phase and post mitotic progeny.

Differential analysis of the gene expression of day 10 datasets revealed 408 genes in X1 and 662 genes in X2. 10 days after knockdown, 153 genes were downregulated (adjusted p-value < 0.05) and 255 genes were upregulated (adjusted p-value < 0.05) in the stem cell population (X1 cells) while 179 genes were downregulated (adjusted p-value < 0.05) and 237 genes were upregulated (adjusted p-value < 0.05) in the stem cell progeny population (X2 cells) (Fig. 3.5a-d).

The planarian dd_Smed_v6 transcriptome assemble has two transcripts that belong

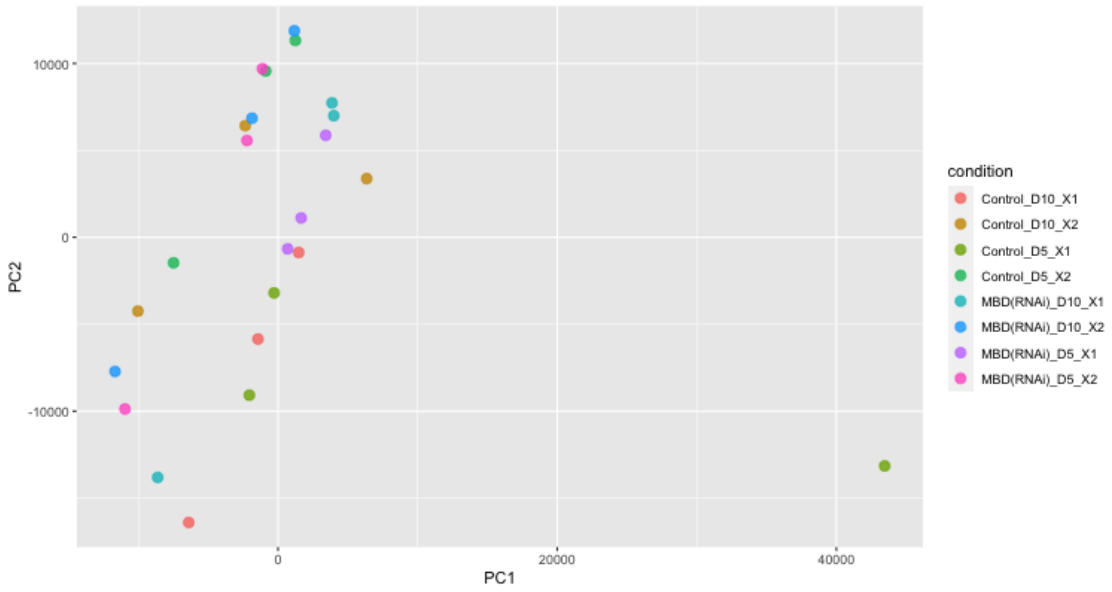

Figure 3.3: PCA plot showing the clustering of the different replicates of the RNA-seq libraries.

to *Smed-mbd2/3* (dd_Smed_v6_3054_0_1 and dd_Smed_v6_3054_0_2). dd_Smed_v6_3054_0_1 is 1999 bp while dd_Smed_v6_3054_0_2 is 1952 bp long; the two sequences are 99.8 % identical. *Smed-mbd2/3* was found to be downregulated (> 3 log-fold change) in X1 and X2 cells at both day 5 and day 10 time points. This demonstrates that the knockdown of *mbd2/3* was successful. No changes were seen in the levels in *smedwi-1* levels in any datasets, which corroborates in situ hybridisation results from Chapter 2 and confirms that levels of *smedwi-1*$^+$ stem cells are not affected by the inhibition of *mbd2/3* in planarians. As expected, a downregulation of *Smed-agat-1* transcripts was seen in X2 cells.

Interestingly, an upregulation of an uncharacterised zinc finger protein (*zfp*) transcript was seen in X2 cells in both day 5 (2.17-fold) and day 10 (2.87-fold) datasets. The zeta-class neoblasts exhibit a high expression of *piwi-1*, *zfp1* and *soxP-3*. In situ hybridisations in Chapter 2 show no changes in total zeta neoblasts numbers after loss of *Smed-mbd2/3*. Unlike *zfp1*, which has high expression in neoblasts, single cell transcriptomics shows that this *zfp* transcript is expressed in differentiated cell types, and mainly in parenchymal cells (`https://shiny.mdc-berlin.de/psca/`).

Dachshund (dd_Smed_v6_14598_0_1) was downregulated 1.5-fold in stem cells

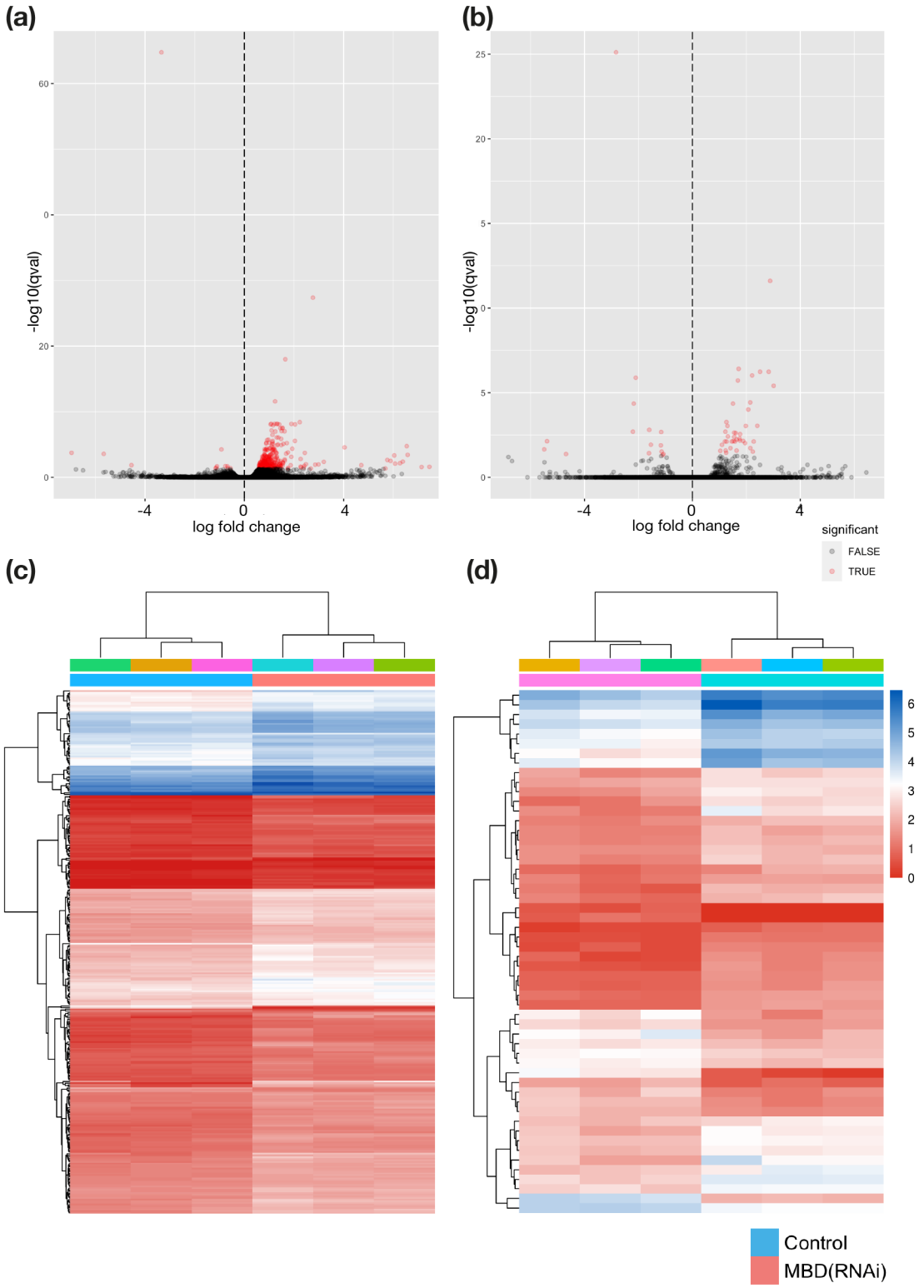

Figure 3.4: Transcriptional response of planarian stem cells 5 days after knockdown of *Smed-mbd2/3*. (a) Volcano plots showing the differentially expressed transcripts in X1 and (b) X2 cells 5 days post-RNAi. Statistical significance was calculated using the Wald test with an adjusted p-value < 0.05 (red dots). The hierarchical clustering of differentially expressed genes in (c) X1 and (d) X2 cells 5 days post-RNAi.

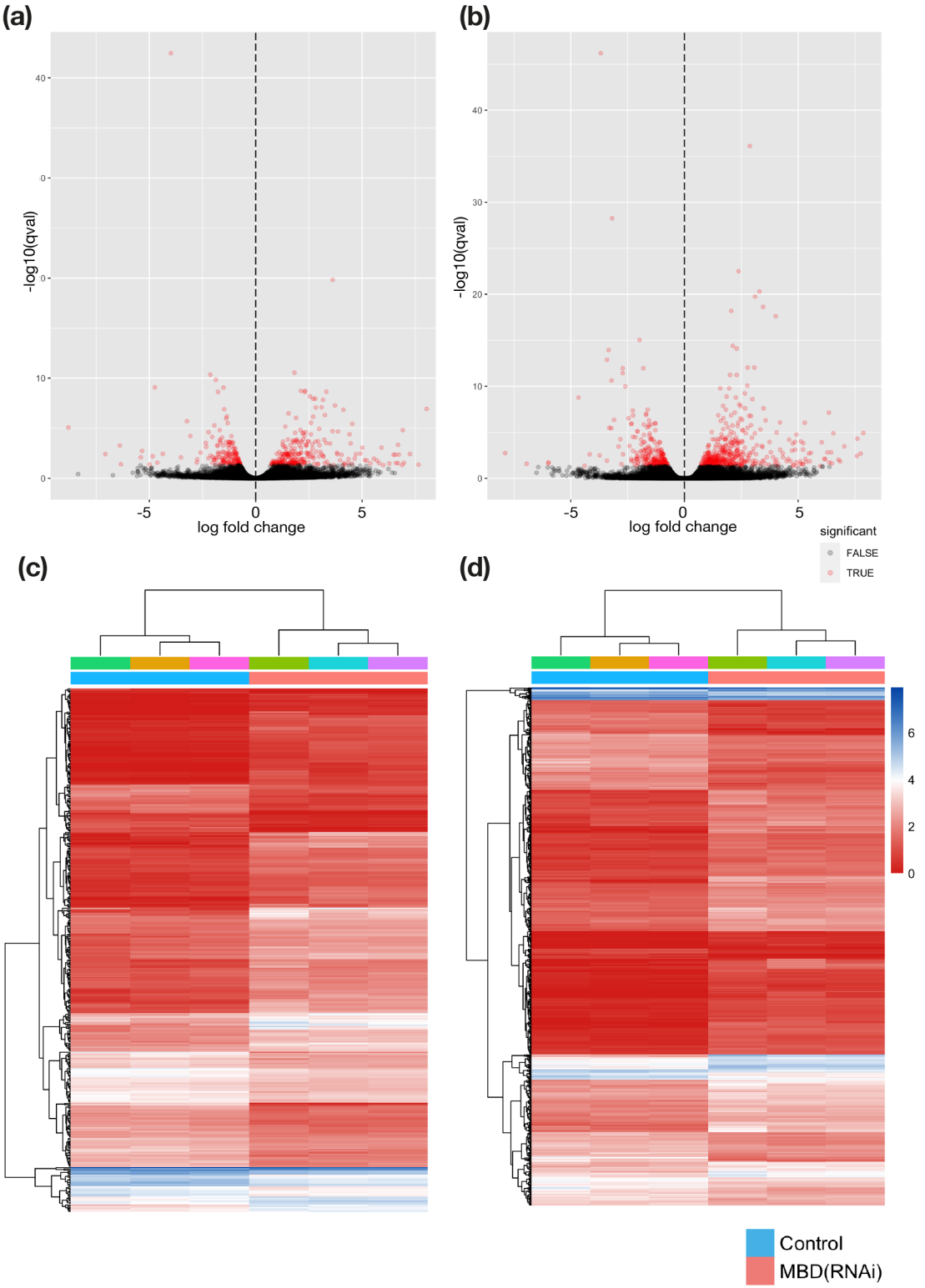

Figure 3.5: Transcriptional response of planarian stem cells 10 days after knockdown of *Smed-mbd2/3*. (a) Volcano plots showing the differentially expressed transcripts in X1 and (b) X2 cells 10 days post-RNAi. Statistical significance was calculated using the Wald test with an adjusted p-value < 0.05 (red dots). (c) The hierarchical clustering of differentially expressed genes in X1 and (d) X2 cells 10 days post-RNAi.

at day 10 post knockdown. Several transcripts of mitofusins (including MFN1 and MFN2) and mitochondrial assembly regulatory factor were found to be upregulated in all datasets. Some mitofusin transcripts were found to be > 6-fold upregulated in X1 cells in both time points. An upregulation of these mitofusins was also observed after the knockdown of *Smed-chd4* [Tu et al., 2015].

A Bicaudal-like transcription factor (dd_Smed_v6_86976_0_1) with basic transcription factor 3 (BTF3) and NAC domains was upregulated in day 5 X1, day 10 X1 and X2 datasets. Bicaudal is essential for differentiation and the planarian *Bic-C* is necessary for the maintenance of *gH4* and *nanos* expression in presumptive germ cells in asexual planarians [Wang et al., 2010]. *Bic*-C, together with *eIF3c*, plays an essential role in the proper development and maintenance of planarian male germ cells. *BTF3* is a general transcription factor that is able to forms a complex with RNA polymerase II and has been shown to be required for the initiation of transcription [Zheng et al., 1987]. Overexpression of *BTF3* has been implicated in a variety of tumours, including adenocarcinoma, colorectal cancer, and gastric cancer [Yang and Kang, 2019]. *BTF3* has a role in proliferation, cell cycle and apoptosis, and significantly promotes cancer progression in vivo [Yang and Kang, 2019]. An upregulation of this transcription factor upon knockdown of *mbd2/3* suggests its involvement in the dysregulation of neoblast genes, which leads to over-expression of targets that are typically silenced in wild-type conditions.

65 genes were common to X1 cells in day 5 and day 10 datasets, while 43 genes were common to X2 cells in day 5 and day 10 datasets (Fig. 3.6a-b). 15 genes were common in day 5 datasets while 181 genes were common in the day 10 datasets (Fig. 3.6c-d). Common genes that were upregulated in X1 datasets after knockdown of *Smed-mbd2/3* include bicaudal-like, mitofusins, serine protease inhibitor-1, low density lipoprotein receptor related-2, *slc17a-10* and *slc30a-8*. All others are unknown genes and over 60% of these genes were also seen to be upregulated after knockdown *chd4*. Additionally, more than half the genes have proportionally higher expression in Xins or differentiated cell populations. Plexin B-5, CDC25-2, synaptotagmin XV-like protein, *slc17a-8*, *zfp1*-like and histones including H2A were upregulated in X2 datasets after the

knockdown of *mbd2/3*. Most unknown genes that are upregulated after the knockdown of *mbd2/3* are uncharacterised genes that have higher expression in differentiated cells and could be potential markers for epidermal, gut and protonephridial lineages. Fewer genes are downregulated after knockdown of *mbd2/3* (than upregulated) and overall, a majority include uncharacterised proteins including those with transcription factor, DNA, and chromatin binding domains. This suggests that many targets of MBD2/3 are transcription factors which, together with chromatin modifiers, modulate the expression of various genes essential for lineage commitment.

MBD2/3 plays a role in the determination of neuronal lineage

Dachshund (dd_Smed_v6_14598_0_1) was downregulated 1.5-fold in stem cells at day 10 post knockdown. Dachshund is a transcription factor that, together with other DNA-binding transcription factors, modulates gene expression and cell fate determination during development. Dachshund (*Dac*) plays a key role in regulating the development of the embryonic brain, optic lobes, and central nervous system as well as the adult brain, mushroom bodies, gonads, antennae, and legs [Mardon et al., 1994, Martini et al., 2000, Noveen et al., 2000, Kurusu et al., 2000, Rauskolb, 2001, Keisman and Baker, 2001]. *Dac* forms a crucial part of the retinal determination gene network in the Drosophila eye. Moreover, the *Dac* protein is necessary for normal cell-fate determination of imaginal disc cells [Mardon et al., 1994]. Defective eye development occurs upon loss of function mutations of the *Dac* gene. Moreover, ectopic *Dac* expression suffices to redirect the development of non-retinal tissues into an eye fate [Mardon et al., 1994, Anderson et al., 2006]. *Smed-dachshund* expression was not detected in eye regeneration [Lapan and Reddien, 2011] but was observed in *S. polychroa* embryos [Martín-Durán et al., 2012]. Moreover, Dachshund is not required for eye embryonic development and adult regeneration in planarians [Lapan and Reddien, 2011, Martín-Durán et al., 2012]. These results suggest that the Dachshund network, as it has been shown in other organisms, is not constructed in planarians – at least not during the embryonic formation of the eye. Dachshund could be involved in the development of nervous structures in planarians and this is supported by its expression in neuronal cell types (Fig. 3.7a-b). I hypothesise that

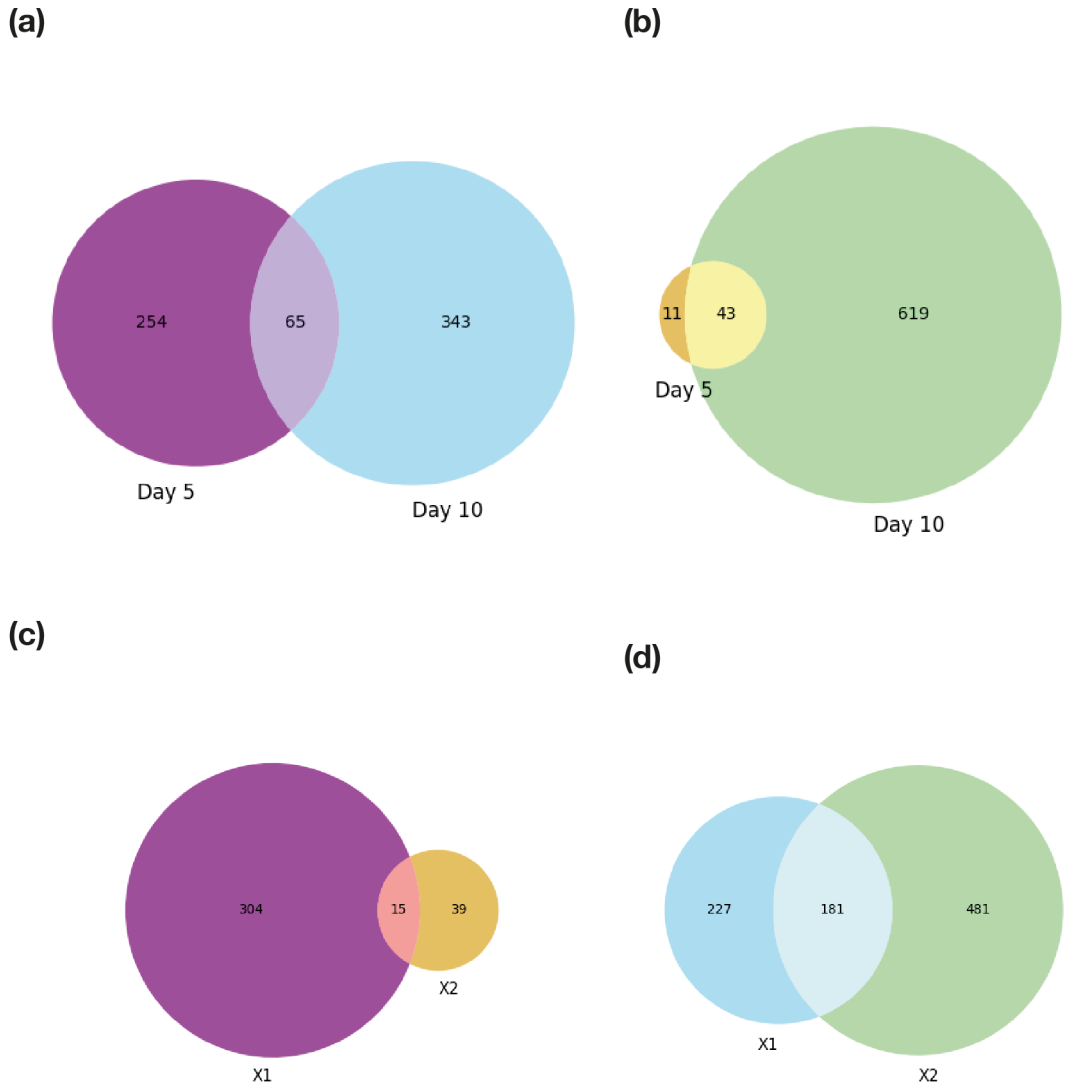

Figure 3.6: Venn diagrams showing the number of differentially expressed genes common in (a) X1 datasets and (b) X2 datasets between day 5 and day 10 datasets. Venn diagrams showing number of genes common to X1 and X2 cell populations after (c) day 5 and (d) day 10 post RNAi.

MBD expression in planarians is essential for maintaining the levels of Dachshund, which is required for the lineage specification of various neuronal cell types. In the absence of *mbd2/3*, Dachshund is downregulated, leading to defects in the differentiation of neuronal cells. MBD2/3 could be either directly, or indirectly through another regulator, be responsible for maintaining the levels of Dachshund in planarians (Fig. 3.7c).

In mammals, recent research has shown that the MBD3 and NuRD complex play a key role in regulating stem cell biology, including self-renewal and neurogenesis, by spatially and temporally regulating gene expression [Knock et al., 2015]. In rodents, depletion of the NuRD complex impaired the establishment of granule neuron parallel fiber/Purkinje cell synapses in vivo [Yamada et al., 2014]. A set of NuRD regulated genes that are critical regulators of presynaptic differentiation was also identified in the cerebellar cortex [Yamada et al., 2014]. Deletion of MBD3 in the developing mouse led to a wide range of defects including reduced cortical thickness, defects in the specification of cortical projection neuron subtypes and neonatal lethality [Knock et al., 2015]. The deletion of Mi-2/NuRD in neural stem cells (NSCs) of Drosophila triggered excessive Notch signalling which lead to supernumerary NSCs [Zacharioudaki et al., 2018]. The authors suggest that the NuRD complex is necessary for decommissioning stem-cell enhancers in their progeny, thereby enabling the switch towards more differentiated fates [Zacharioudaki et al., 2018]. It is therefore possible that MBD2/3-NuRD in planarians is involved directly in the repression of inappropriate transcription in stem cell progenies and differentiated cells cells to facilitate appropriate cell lineage choice and differentiation programmes or indirectly with its association with Dachshund [Knock et al., 2015].

MBD2/3 is required to repress inappropriate transcription in neoblasts

Several transcripts of mitofusins (including MFN1 and MFN2) and mitochondrial assembly regulatory factors were found to be upregulated in all datasets post knockdown of *mbd2/3*. Some mitofusin transcripts were found to be > 6-fold upregulated in X1 cells at both time points. An upregulation of these mitofusins was also observed after in the *Smed-chd4*(RNAi) transcriptome (data from Tu et al. [2015]). Further, I

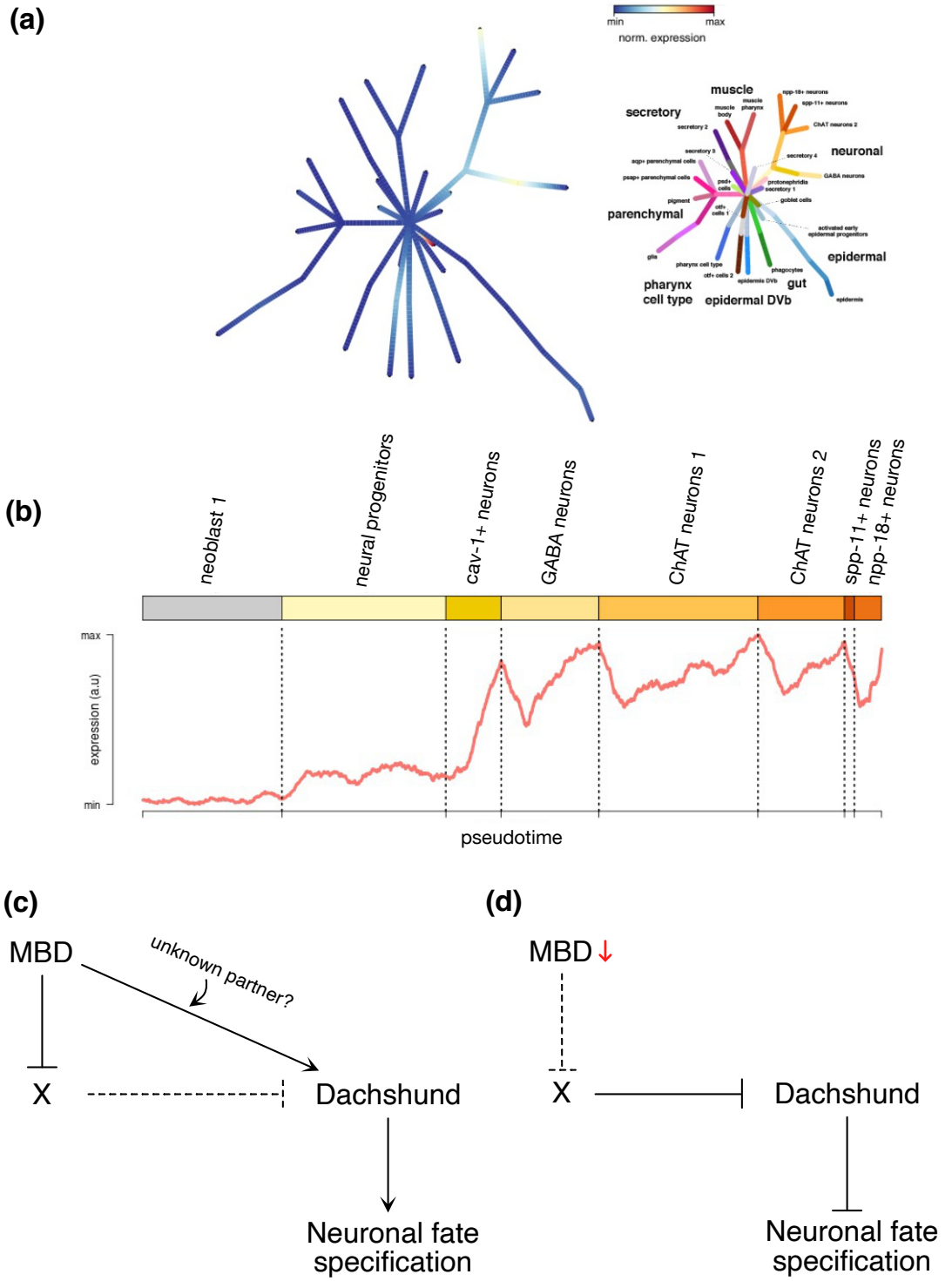

Figure 3.7: Dachshund-MBD2/3 may play a role in neuronal lineage specification in planarians. (a) Single cell lineage tree and (b) plot representing gene expression changes in pseudo-temporally ordered cells indicates that the Dachshund gene is expressed in many neuronal lineages (`https://shiny.mdc-berlin.de/psca/`) (c) Hypothesis of the role of MBD/Dachshund in the specification of neuronal lineages.

verified the expression of these five transcripts in the single cell datasets. All five genes have very low expression in stem cells and progenitors, and have high expression in differentiated cell types. Gene expression changes in pseudo-temporally ordered cells of epidermal, pharyngeal, protonephridial, neural, muscle, gut, and secretory lineages of all five mitochondrial genes show high expression in terminally differentiated cell types (Figs. 3.8 and 3.9). This suggests that these genes are markers of differentiated cells and are repressed by the MBD2/3-NuRD in X1 and X2 cells. Knockdown of *mbd2/3* has led differentiated cell-specific genes to undergo transcriptional activation in stem cells. Further work is required to gain an overall picture of the effect on the differentiation potential of these stem cells.

Mitofusins have been shown to be essential in stem cell fate determination to mediate EMT-associated stemness [Wu et al., 2019]. In HSCs, mitofusin 2 (MFN2) is involved in promoting mitochondrial fusion and mitochondria-ER tethering [Luchsinger et al., 2016]. In general, it is noted that ESCs have non-fused mitochondria, while lineage-primed or multipotent cells such as neural, mesenchymal, hematopoietic stem cells as well as hematopoietic progenitor cells and neurons have fused, elongated mitochondria [Seo et al., 2018]. During reprogramming and re-differentiation of iPSCs, mitochondrial morphology changes dynamically, with mitochondria elongating and adopting a circular shape after re-differentiating into a neural lineage [Suda et al., 2011, Zhou et al., 2012, Simsek et al., 2014, Khacho et al., 2016, Beckervordersandforth et al., 2017, Seo et al., 2018]. It has also been shown that *Mfn1* and *Mfn2* halt reprogramming by inhibiting cell proliferation through direct binding to *Ras* and *Raf* [Son et al., 2015]. It has also been suggested that *Mfn1* controls cell fate as a translational target [Son et al., 2015]. Depleting *Mfn1/2* results in reciprocal inhibition of the *p53-p21* pathway, promoting the transformation of somatic cells to a pluripotent state and maintaining pluripotency [Son et al., 2015]. Upon reprogramming, the fragmented mitochondria were markedly increased, whereas the fused form gradually disappeared [Son et al., 2015]. Activation of *Drp1*, a mitochondrial fission protein, promotes mitochondrial fragmentation contributing to the maintenance of pluripotency [Son et al., 2013]. Son et al. [2015] propose a model for the control of cellular stability and plasticity via the reciprocal interaction of

the *p53/p21* and *Mfn1/2* pathways. *Mfn1/2* depletion leads to mitochondrial fission and increased cell proliferation, thereby promoting maintenance of pluripotency while increased levels of *Mfn1/2* lead to mitochondrial fusion and cell cycle arrest aiding the exit from pluripotency (Fig. 3.10a). An upregulation of mitofusins in both X1 and X2 cells could hence be due to a mis-regulation of stem cell genes caused by the failure in cells to repress inappropriate transcription in stem and progenitor cells (Fig. 3.10b).

Functional characterisation of target genes

40 transcripts were selected from the four differentially expressed data sets based on their expression levels. These transcripts were cloned into pPR-T4P vector. Ten of these genes were knocked down in planarians using RNAi. 10 animals were used for the knockdown of each gene as well as control. After 6 injections, animals were amputated into three to observe the functions of these genes in regeneration. Animals were observed and visualised regularly after amputation. None of the knockdown animals showed any strong phenotypes during regeneration (Figs. 3.11 and 3.12). To ascertain any molecular changes caused by the knockdown of these genes, future experiments involving assessment of different stem cell and progeny markers using in situ hybridisation will be required. A broader and in-depth screen of target genes and functional characterisation is necessary for the identification of new markers as well as the targets of MBD2/3-NuRD.

3.3 Conclusion

The current study shows that there exists a robust transcriptional response to the knockdown of MBD2/3 that regulates a diverse range of biological processes including the determination of stem cell fate. A majority of misregulated transcripts are uncharacterised genes and a few known ones are involved in the regulation of mitochondrial morphology and metabolism, cell differentiation or have DNA binding activities. As expected, a downregulation of *mbd2/3* as well as *agat-1* transcripts was seen. A notable transcript that was downregulated upon knockdown of *mbd2/3* was Dachshund. Dachshund is a transcription factor involved in the regulation of gene expression and cell fate determination during development. It is expressed in neuronal cell types in planarians

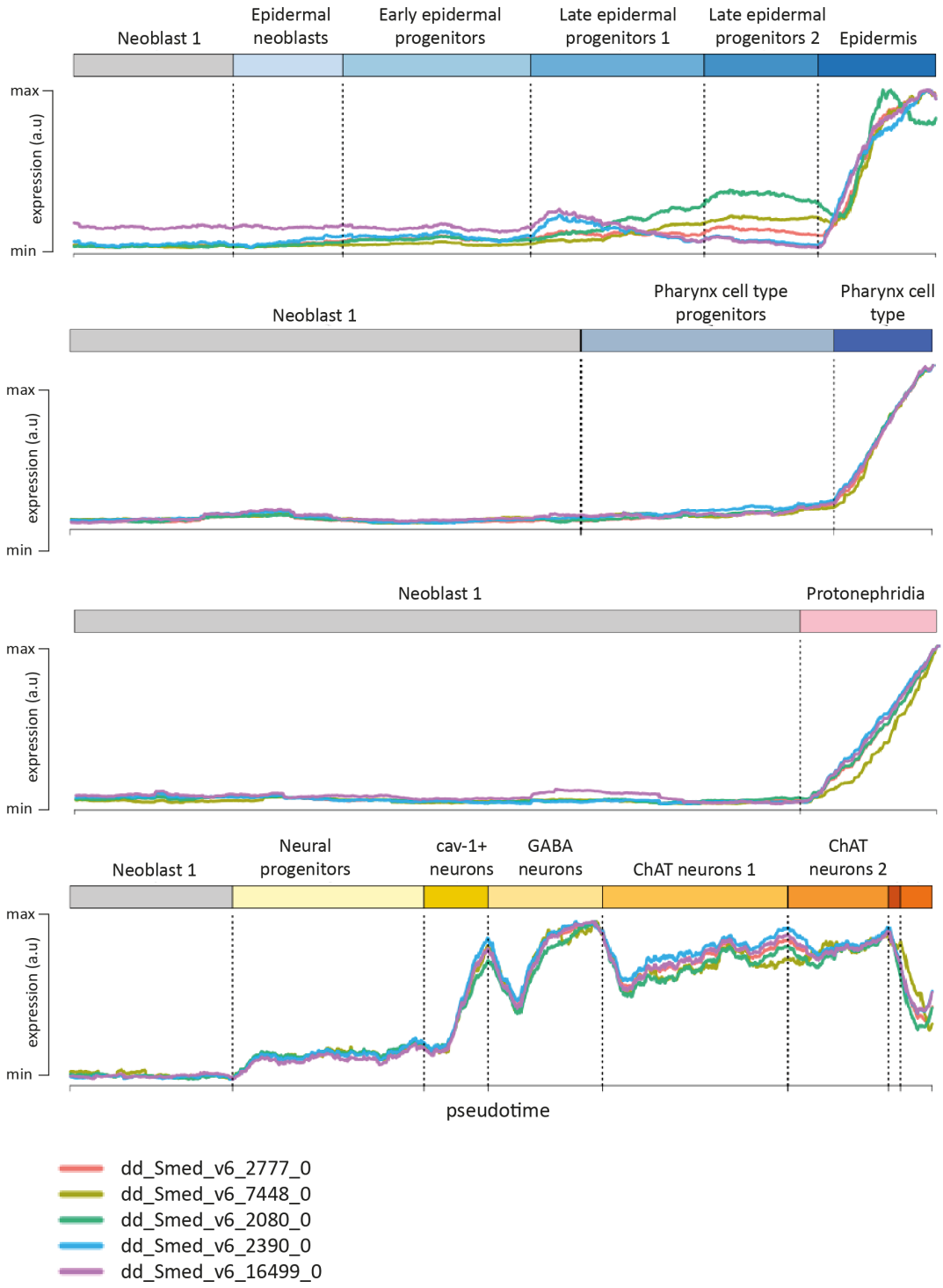

Figure 3.8: Gene expression changes in pseudo-temporally ordered cells of epidermal, pharyngeal, protonephridial and neural lineages of the five mitochondrial genes upregulated after *Smed-mbd2/3* knockdown. All 5 genes have low expression in neoblasts and committed progenitors but have high expression in differentiated cells (plotted using `https://shiny.mdc-berlin.de/psca/`).

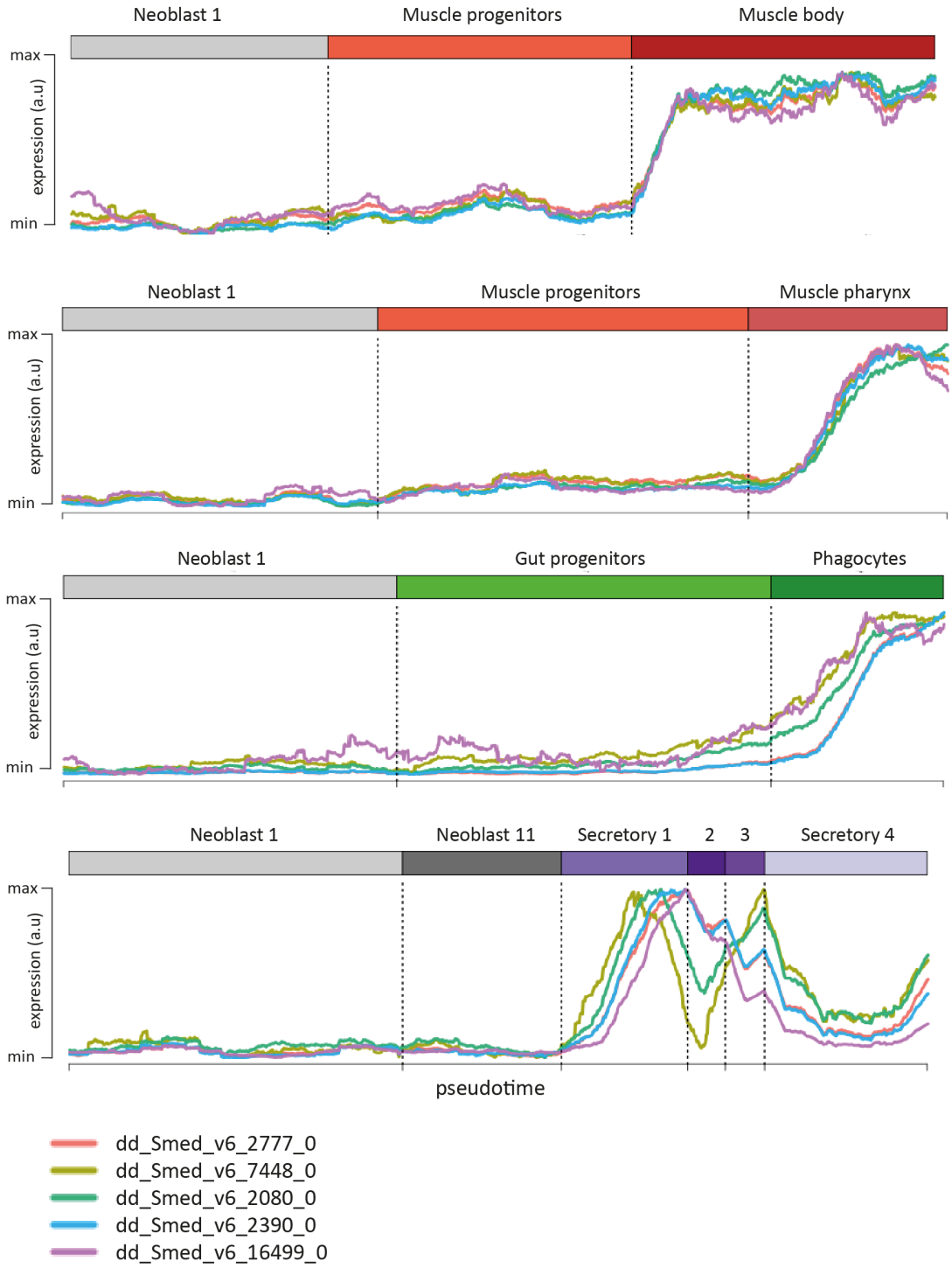

Figure 3.9: Gene expression changes in pseudo-temporally ordered cells of muscle, gut, and secretory lineages of the five mitochondrial genes upregulated after *Smed-mbd2/3* knockdown. All 5 genes have low expression in neoblasts and committed progenitors but have high expression in differentiated cells (plotted using `https://shiny.mdc-berlin.de/psca/`).

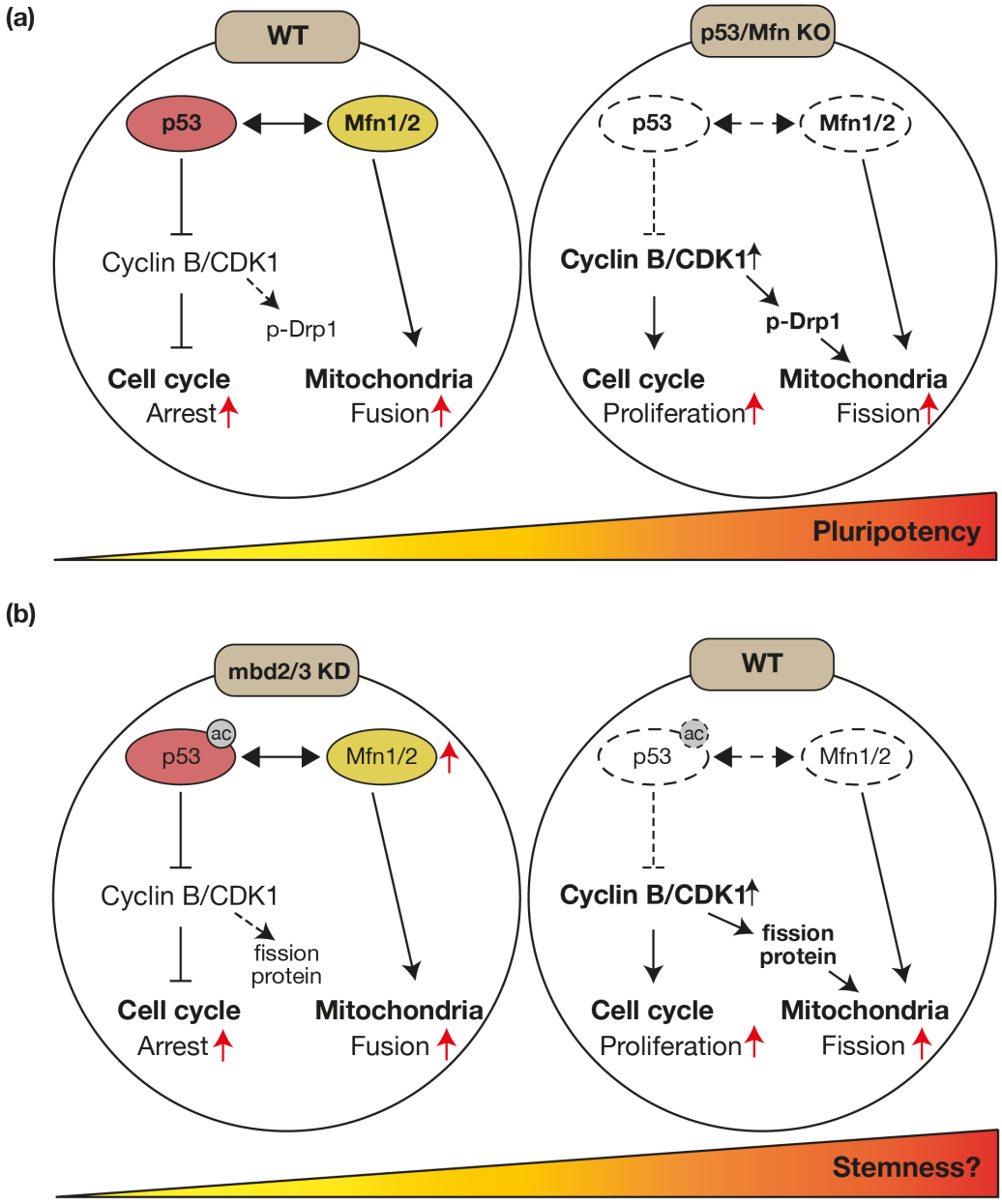

Figure 3.10: Silencing of *mbd2/3* leads to over-expression of mitochondrial fusion proteins in planarians. (a) p53- and p21-KO cells express low levels of mitochondrial fusion proteins [Son et al., 2015]. (b) In planarian stem cells, high ratio of fission to fusion proteins together with low levels of acetylated p53 lead to maintenance of pluripotency. After knockdown of *mbd2/3*, an increase in fusion proteins leads to mitochondrial fusion and potentially cell cycle arrest. Future experiments will help investigate the effect of this on the stemness or differential potential of stem cells in *mbd2/3* knockdown animals. Figure adapted from Son et al. [2015].

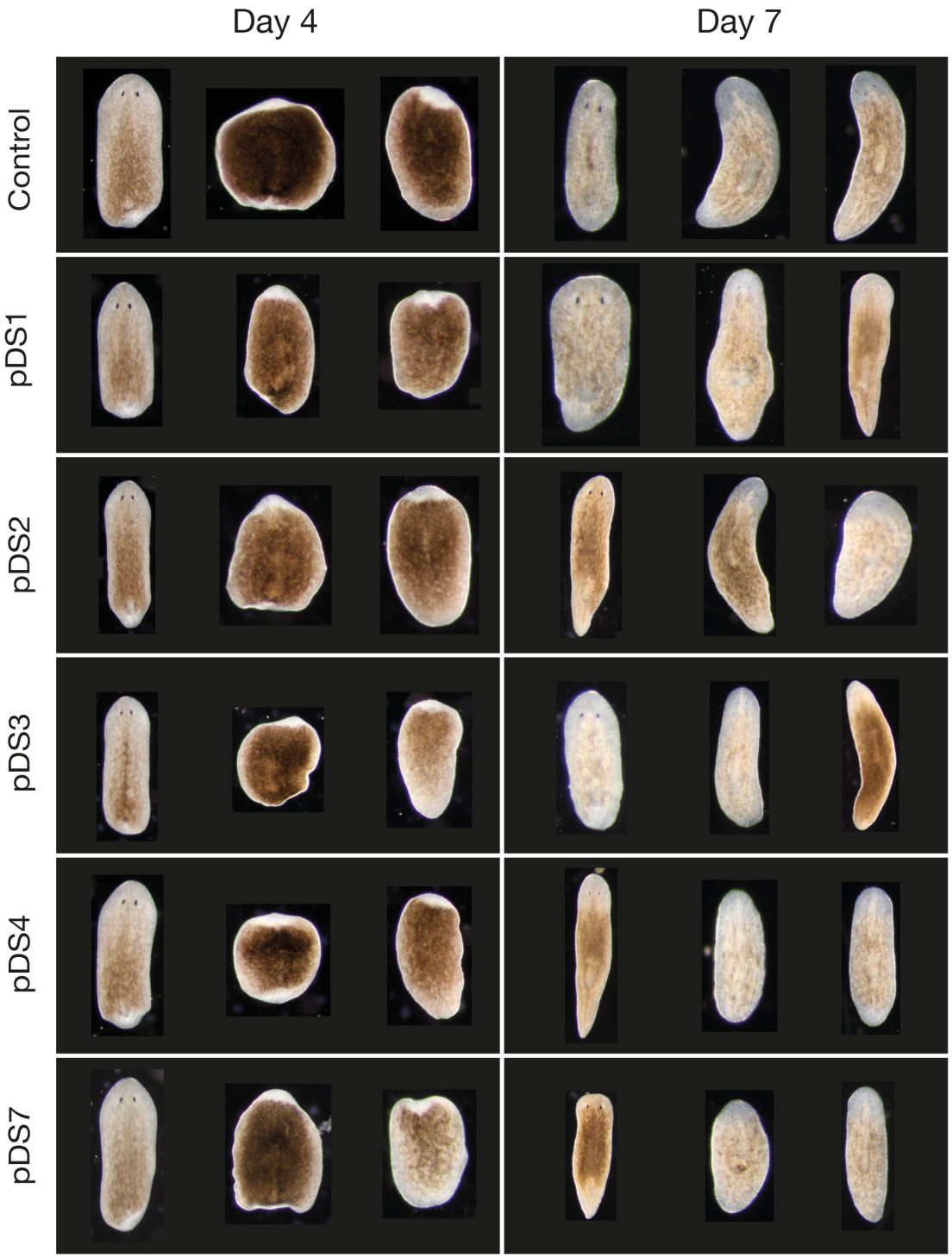

Figure 3.11: Silencing of select genes obtained from differential analysis does not affect regeneration. Head, middle and tail pieces are shown at day 4 and day 10 of regeneration after knockdown of target genes and control. Knockdown animals did not show any strong phenotype.

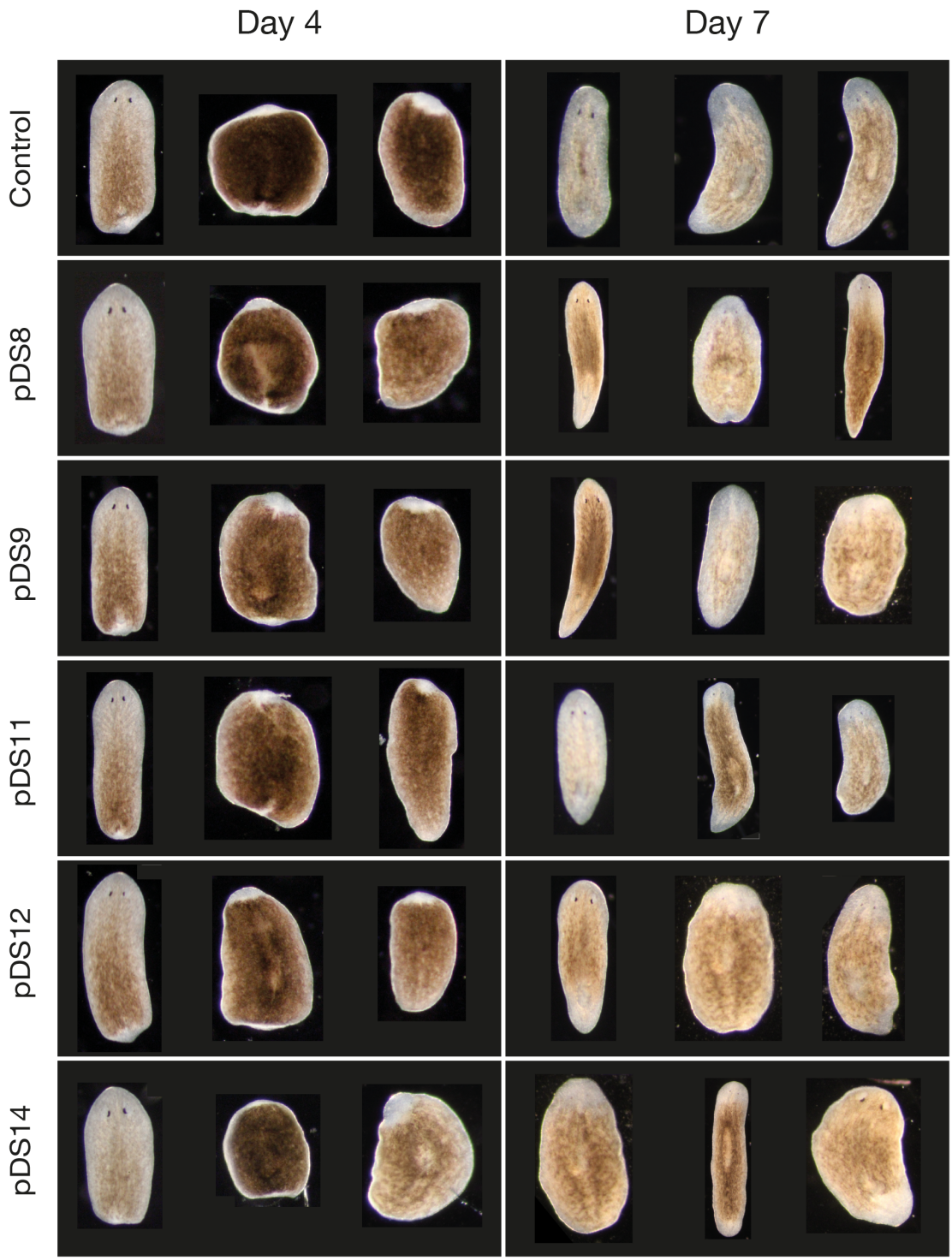

Figure 3.12: Knockdown of select genes affected by inhibition of *Smed-mbd2/3* does not affect regeneration. Head, middle and tail pieces are shown at day 4 and day 10 of regeneration after knockdown of target genes and control. Knockdown animals did not show any strong phenotype.

and plays a possible role in the development of nervous structures. MBD2/3-NuRD may be indirectly involved in the regulation of Dachshund and directly involved in the repression of inappropriate genes to facilitate appropriate lineage determination.

A significant upregulation of several mitofusins, which form part of the mitochondrial fusion machinery, was seen after the knockdown of *Smed-mbd2/3*. Modulation of mitochondrial fission and fusion has emerged as critical for regulation of both pluripotent and adult stem cells. The balance of mitochondrial fission-fusion is tilted more towards unopposed fission in embryonic stem cells or induced pluripotent stem cells [Fu et al., 2019, Spurlock et al., 2020]. However, transition from naïve to primed/committed pluripotent cells involves tilting the balance towards mitochondrial fusion [Bahat et al., 2018, Spurlock et al., 2020]. The mitochondrial morphology changes from non-fused and spherical to fused and elongated in terminally differentiated cells. The over-expression of these transcripts could indicate that more stem cells are exiting pluripotency or are losing their differentiation potential.

An upregulation of several transcripts encoding transcription factors including *zfp* and bicaudal/BTF3-like proteins was also seen. As the majority of over-expressed transcripts exhibit chromatin and gene expression modulating activity, it is evident that these targets of MBD2/3-NuRD should have been repressed to allow correct lineage commitment. These mis-regulated genes led to aberrant expression of proteins and lineage commitment, resulting in the failure of stem cells to differentiate into other lineages.

Future work involving functional analysis of the dataset of evolutionary conserved genes regulated by MBD2/3 can predict potential markers for various stem cell lineages and identify novel factors responsible for the determination of stem cell fate and differentiation. Further functional studies are required to understand the mechanistic role of these genes in stem cell differentiation. Like MBD2/3, the NuRD complex component CHD4 is essential for the maintenance of *prog-1* and *agat-1* cells [Scimone et al., 2010]. These abundant cell populations have rapid turnover kinetics and are routinely investigated to understand neoblast differentiation along the epidermal lineage. Nevertheless, the commonalities and differences between the lineage relationships of

these diverse cell types are little understood, and the underlying mechanisms controlling the progression of epidermal progenitors into mature cells remain completely unknown.

Nevertheless, although a majority of the genes and lineage markers still remain functionally unexplored, this work involving global gene-expression profiling holds great promise for unravelling functions of MBD2/3-NuRD complex and the mechanisms that control stem cell differentiation.

3.4 Methods

Methods for generation of dsRNA, riboprobe sythesis, RNA interference, in situ hybridisation, confocal imaging and cell counting are the same as those described in Chapter 2.

Statistical analysis

Prism 8.0 (`https://www.graphpad.com/`) was used to calculate statistical significance and plot curves. Results are expressed as mean $\pm$ standard deviation (SD). Statistical analyses were performed using Student's t-test using statistical significance at $p < 0.05$. Wherever cell numbers were compared between experimental condition and control, a two-tailed t-test assuming unequal variance was used. Each legend states the number of specimens per condition, where relevant. Bar graphs show the mean average and the error bars are always standard error of the mean.

For analysis of RNA-seq data, the Wald test (part of the Sleuth [Pimentel et al., 2017] software) was used for assessing differential expression.

Gamma irradiation

Gamma irradiation severely damages DNA, thereby killing proliferating cells. Proliferating planarian cells belong to two sub-populations (X1 and X2) of cells corresponding to distinct regions of a FACS plot that are markedly reduced following γ-ray-irradiation. The third sub-population containing differentiated cells are insensitive to radiation (Xins) [Hayashi et al., 2006, Higuchi et al., 2007, Romero et al., 2012]. After irradiation using a non-lethal dose, stem cells or X1 cells are drastically reduced by 3 days and stem cell progeny or X2 are reduced after 5-7 days. As irradiated worms are depleted in one or both

stem cell sub-populations, they are used as controls for gating in FACS analysis as they show that cells gated as stem cells and stem cell progeny are in fact proliferating cells.

Animals were starved for 7 days and exposed to 30 Gy (Gray) of 137 Cs gamma rays using a GSR D1 Gsm (Gamma service GmbH, serial number: A0108) gamma irradiator. Irradiation was performed using a 50 ml falcon tubes with worms settled at the bottom with minimal amount of planarian water ($\sim$2 ml). Worms were used as controls for cell sorting 3 and $5-7$ days post irradiation.

Cell dissociation and fluorescence activated cell sorting

Wild-type whole worms that were starved for 7 days were used for cell dissociation. Planarians were cut into small pieces using a scalpel and transferred into 1.5 ml micro-centrifuge tubes with CMFHE^{2+} solution (calcium-magnesium free buffer supplemented with 0.1 % BSA, 0.5 % glucose, 15 mM HEPES and 3 mM EDTA). The scalpel was regularly wiped while cutting planarians to remove mucus. The tissue pieces were digested using Papain (15 U/ml) for 1 h at 25 °C. CMFHE^{2+} and the digestion solution was always made fresh. The digested pieces were mechanically dissociated using a pipette and filtered through a 100 µm and 35 µm cell strainer into a pre-cooled tube. This step preferentially enriches the stem cell population by reducing cellular debris. The filtered cells were collected and the final volume of the cell suspension made up to 1 ml, adjusting the concentration depending on visual inspection of cell density. The cells were stained with the nuclear marker Hoechst 34580 (1 mg/ml stock concentration) at a final concentration of 20 µg/ml. Cytoplasmic staining was performed using Calcein AM (1 mg/ml stock concentration) at a final concentration of 0.5 µl/ml. Staining was performed for at least 1 h at room temperature and the samples were protected from light. Prior to FACS analysis, 1 µl of propidium iodide (10 µg/ml) was added to the samples to stain the nucleic acid of cells with a broken cell membrane, i.e. dead cells. Cells dissociated from irradiated worms were used for gating and cells were sorted using a FACS ARIA III. BD FACS DIVA software was used for the sequential gating of cell populations as outlined in [Romero et al., 2012]. Cells were collected into

individual collection tubes containing ice-cold PBS. Dissociating 100 worms will give $> 600,000$ X1 cells and $> 1,000,000$ X2 cells.

RNA extraction from sorted cells

FACS sorted cells (collected in PBS) were centrifuged at 4000 g at 4 °C. The pellet was re-suspended in 250 µl nuclease free water and 750 µl TRIzol LS was added. The suspension was homogenised by pipetting up and down several times. Samples were snap frozen and stored in -80 °C if RNA was not extracted immediately. 200 µl of chloroform was added to each sample and vigorously mixed by hand for 15 seconds. After 2–3 minutes, the samples were centrifuged for 15 minutes at $12,000 \times g$ at 4 °C. The mixture separates into a lower red phenol-chloroform, and interphase, and a colourless upper aqueous phase. The aqueous phase containing the RNA was transferred to a new tube. The RNA was purified from the aqueous phase using an RNA Clean & Concentrator-5 (Zymo Research) kit following manufacturer's guidelines. The total RNA was quantified using a Qubit RNA BR Assay Kits (Invitrogen). Around 0.5-0.7 µg of total RNA was used for library preparation to obtain good quality libraries for RNA-seq.

mRNA library preparation and sequencing

Total RNA purified from sorted cells was used for library preparation using Illumina TruSeq stranded mRNA LT kit, as per manufacturer's guidelines. The adapters were diluted $1 : 1$ using re-suspension buffer to reduce the chance of the formation of adapter dimers. The libraries were quantified using Qubit dsDNA High Sensitivity kit (Invitrogen). The quality and size distribution of the libraries were checked using an Agilent 4200 TapeStation system. The libraries were individually quantified using the qPCR-based KAPA library quantification kit. Libraries with different indices were pooled together and sequenced on an Illumina NextSeq machine. The samples were paired-end sequenced on an Illumina NextSeq.

Three biological replicates per condition per sample and a total of 24 libraries were made. Libraries for two timepoints (day 5 and day 10) were chosen for *Smed-mbd2/3*

and control (RNAi) and for two cell populations (X1 and X2). Two pools with 12 libraries each were prepared and sequenced.

RNA-seq data analysis

Raw reads were trimmed with Trimmomatic 0.32 [Bolger et al., 2014] and pseudo-aligned *S. mediterranea* transcriptome with Kallisto 0.42.3 [Bray et al., 2016]. Differential gene expression analysis was subsequently performed with Sleuth 0.28.1 [Pimentel et al., 2017]. The differential expression of genes was analysed from three replicates of each condition using the Wald test. The analysis and generation of figures was performed in R using Sleuth. Common transcripts in different data sets were analysed using custom python scripts.

4

Elucidating chromatin accessibility in planarian cells to identify enhancers

Contents

4.1 Introduction

Stem cells have the ability to self-renew and differentiate along all cell lineages. During stem cell differentiation, self-renewal genes are downregulated and cell lineage-specific regulatory genes are activated. These changes in gene expression are associated with marked changes in chromatin structure [Zhao et al., 2019]. Transcriptionally active and repressed genes contain "open" and "closed" chromatin, respectively. The "open" regions, referred to as euchromatin, contain actively transcribed genes and enhancers. In contrast, "closed" regions, referred to as heterochromatin, containing centromeres and telomeres, are relatively compact regions containing mostly inactive genes and reflect changes in the cell cycle [Bannister and Kouzarides, 2011]. Active chromatin is typically marked by the presence of H3K4me1, H3K4me3 and H3K27ac [Calo and Wysocka, 2013, Weber and Henikoff, 2014], while transcriptionally repressed genes are often marked by H3K27me3 and H3K9me3 [Long et al., 2016, Catarino and Stark, 2018]. Therefore, distinct cell types with varied expression of genes display unique chromatin structures. Studies investigating dynamic changes in chromatin structure during differentiation and between different cell types consequently provide fresh insights into the molecular mechanisms of cell-fate specification and differentiation [Zhao et al., 2019].

Promoters and enhancers are *cis*-regulatory DNA sequences and are key contributors of tissue-specific gene expression. Enhancers can regulate genes often over a considerable distance from the transcription start site (TSS). They control specificity as well as level of transcription and interact with sequence-specific transcription factors [Zhao et al., 2019]. These regions display specific signatures of post-translational modifications, lower nucleosomal density, as well as increased nuclease sensitivity. Enhancers can be located large distances away from the promoters of genes they control, making them challenging to identify [Spicuglia and Vanhille, 2012].

Owing to the recent technological advances in mapping histone modifications genome-wide, as well as the development of the assay for transposase-accessible chromatin using sequencing (ATAC-seq), several studies have elucidated the chromatin signatures of active enhancers [Visel et al., 2009, Daugherty et al., 2017]. ATAC-seq was used in *C. elegans* to identify putative enhancers that drive cell-type- and temporal-specific gene expression,

and their stage-specific expression patterns were validated using reporter constructs [Daugherty et al., 2017]. Studies using ATAC-seq have also led to novel insights into the dynamic changes in chromatin states during lineage commitment and differentiation [Bao et al., 2015, Heuston et al., 2018, Ruzycki et al., 2018, Zhao et al., 2019].

In a regenerative context, active enhancers have been identified at the whole-body level in the acoel *Hofstenia miamia* [Gehrke et al., 2019] and the cnidarian *Hydra vulgaris* [Murad et al., 2019] and at a tissue level in Drosophila imaginal discs [Harris et al., 2016] and Zebrafish (heart and fin) [Kang et al., 2016, Goldman et al., 2017, Yang and Kang, 2019]. A recent study in planarians identified an enhancer in the first intron of *wnt1* with FoxG binding motifs using ATAC-seq and ChIPmentation [Pascual-Carreras et al., 2020]. The changes in accessibility of anterior and posterior specific active enhancers were investigated after knockdown of *notum* and *wnt1* using whole tissue samples [Pascual-Carreras et al., 2020].

Planarians as a model to study epigenetic mechanisms controlling stem cells in vivo

Planarians are masters of regeneration and are best known for their ability to regenerate their whole bodies. Planarians owe this remarkable ability to neoblasts or pluripotent stem cells, the only dividing cells in their body. Neoblasts replace cells that are lost due to physiological turnover as well as injury [Tu et al., 2015]. Numerous studies have established the conservation of fundamental features of stem cell biology with other animals [Önal et al., 2012]. Planarians also allow for the study of stem cell heterogeneity and lineage progression from undifferentiated stem cells due to the availability of molecular markers for stem cells and their progeny. An advantage of using *Schmidtea mediterranea* as a model system for studying epigenetic regulation is the availability of an excellent array of genomic resources and tools to facilitate these studies. These include an excellent genome assembly [Grohme et al., 2018], annotations [Cantarel et al., 2008], a genome database [Robb et al., 2015], transcriptome repository [Brandl et al., 2016] and single cell RNA-seq data sets [Fincher et al., 2018, Plass et al., 2018]. Nonetheless, while planarians are a promising model system for in vivo stem cell biology,

we are only beginning to understand the molecular principles that govern associated regulatory mechanisms. Further research into planarian stem cells and other model organisms will help us understand fundamental stem cell properties and allow us to disentangle pluripotency and self-renewal [Dattani et al., 2018b]. Planarians, therefore, are a promising model organism to investigate the epigenetic regulation of pluripotency, stem cell function, cell-fate specification and differentiation during regeneration.

The understanding of the epigenetic regulation of regeneration in planarians is in its infancy. A number of convincing lines of evidence suggest that DNA methylation-based regulation of the planarian genome is not a feature of epigenetic regulation. *S. mediterranea* contains a single MBD2/3 (methyl-CpG binding domain) protein that lacks the conserved residues known to contact methylated DNA and a conserved DNMT2 with no apparent role in regeneration [Jaber-Hijazi et al., 2013]. The absence of 5-methyl cytosine in the *S. mediterranea* genome was also confirmed in numerous ways, including the lack of antibody staining against 5-methyl cytosine, and undetectable levels of 5-methyl cytosine in high performance liquid chromatography mass spectrometry [Jaber-Hijazi et al., 2013]. These different lines of evidence suggest that the function of MBD2/3 is likely independent of DNA methylation, and that DNA methylation is not involved in the epigenetic control of planarian stem cells. Along with Smed-MBD2/3, the functions of four other nucleosome remodelling and deacetylase (NuRD) complex components have also been investigated by RNA interference in planarians: Smed-HDAC1 [Zhu and Pearson, 2013, Robb and Alvarado, 2014, Ross et al., 2015], Smed-CHD4 [Scimone et al., 2010], *RbAp48* [Bonuccelli et al., 2010, Hubert et al., 2015], and GATAD2 (or *p66*) [Vásquez-Doorman and Petersen, 2016]. While all of these genes affect stem cell differentiation, HDAC1 and CHD4 also affect stem cell renewal or survival in vivo.

The phenotypic effects of the loss of these epigenetic regulators can be effectively assessed during planarian regeneration, and stem cell survival as well as differentiation defects can be assayed by in situ hybridization using a growing list of markers. With the advent of chromatin immunoprecipitation followed by sequencing (ChIP-seq) on planarian cells, one can now associate the phenotypic effects with epigenetic changes at specific loci. ChIP-seq was first used on dissociated planarians to show that histone

methyl-transferase enzymes Set1 and MLL1/2 target different genomic loci [Duncan et al., 2015]. Set1 targets are associated with the maintenance of the stem cells, while MLL1/2 targets were genes that are involved in ciliogenesis [Duncan et al., 2015]. Mihaylova et al. [2018] investigated the role of the MLL3/4 tumour suppressors in planarians by performing RNA-seq and ChIP-seq on the stem cell population after RNAi-mediated knockdown. Loss of MLL3/4 in planarians led to the formation of tumour-like outgrowths, confirming that this methyl-transferase is also a tumour suppressor. LPT-Trr/MLL3/4 proteins are involved in transcriptional regulation via mono- and/or tri-methylation of H3K4 at promoters and enhancers [Mihaylova et al., 2018]. RNA-seq on the same cells revealed a significant upregulation of genes involved in cell proliferation and differentiation, including potential oncogenes. The transcriptional changes of several genes following the inhibition of Smed-LPT correlate with differences in H3K4me1 peaks at their promoter region. However, a correlation between RNA-seq and ChIP-seq data at promoters is not apparent for many other genes [Mihaylova et al., 2018]. This could be because they are indirect regulatory effects of RNAi or because epigenetic changes at enhancers modulate these changes in transcription.

A recent study by Dattani et al. [2018a] developed an improved ChIP-seq protocol for stem cells in *S. mediterranea* to generate genome-wide profiles for the active marks (H3K4me3 and H3K36me3) and repressive marks (H3K4me1 and H3K27me3) to assess the epigenetic regulation of gene expression in neoblasts. As predicted from previous work in vertebrates and other protostomes, these marks showed conserved patterns of association with active and suppressed gene expression in planarian neoblasts. Significantly, loci with little or no transcriptional activity in the neoblast compartment that are known to switch on in neoblast progeny during differentiation were found to be bivalent, having both H3K4me3 and H3K27me3 marks at promoter regions. ChIP-seq analysis also revealed high levels of paused RNA Polymerase II at the promoter-proximal region providing further evidence that these gene are bivalent in neoblasts and are actively transcribed upon differentiation. This suggests that the epigenetic regulation of potency through bivalency is conserved across bilaterians, rather than being a special feature of vertebrates [Dattani et al., 2018a]. Together, these studies have established

a robust ChIP-seq protocol and analysis methodology, encouraging further studies in planarians to investigate histone modification-mediated regulation of stem cell fate.

Thus far, the identity of enhancers in planarians has remained elusive. The lack of transgenic constructs in planarians poses a challenge in the validation of both function and targets of putative enhancers and other regulatory elements. In this chapter, potential enhancers were identified in the planarian genome using ATAC-seq and complementary methodologies, mainly ChIP-seq and RNA-seq.

4.2 Results and discussion

Optimising a protocol to assay chromatin accessibility in planarians

Chromatin accessibility is a feature that can be used for the identification of active enhancers and other regulatory regions, including promoters, in eukaryotes [Zhu et al., 2015, Daugherty et al., 2017, Thomas et al., 2011, West et al., 2014, Klemm et al., 2019]. We employ the assay for transposase-accessible chromatin using sequencing (ATAC-seq) for the first time on the three planarian cell populations, X1, X2 and Xins, to identify enhancers or promoter-proximal cis-regulatory elements. This assay utilises a hyperactive Tn5 transposase enzyme that introduces two cuts 9 bases apart, along with the simultaneous insertion of sequencing adaptors into accessible regions of the chromatin [Buenrostro et al., 2015]. There is a greater probability of Tn5 insertion in open or more accessible chromatin when compared to less accessible chromatin such as condensed heterochromatin. The products of two adjacent Tn5 reactions are excised out and these transposed DNA fragments serve as the input for PCR amplification and purification followed by sequencing. Sequencing reads are then used to identify regions of increased accessibility as well as regions of transcription factor binding genome-wide [Buenrostro et al., 2015]. Figure 4.1 shows an overview of the ATAC-seq protocol.

To ensure adequate transposition, the protocol was optimised by trialling different durations and cell numbers for the transposition reaction. This was performed to factor the genome size and heterogeneity of the different cell populations, as chromatin configurations change during the cell-cycle (in particular, in the S phase during DNA

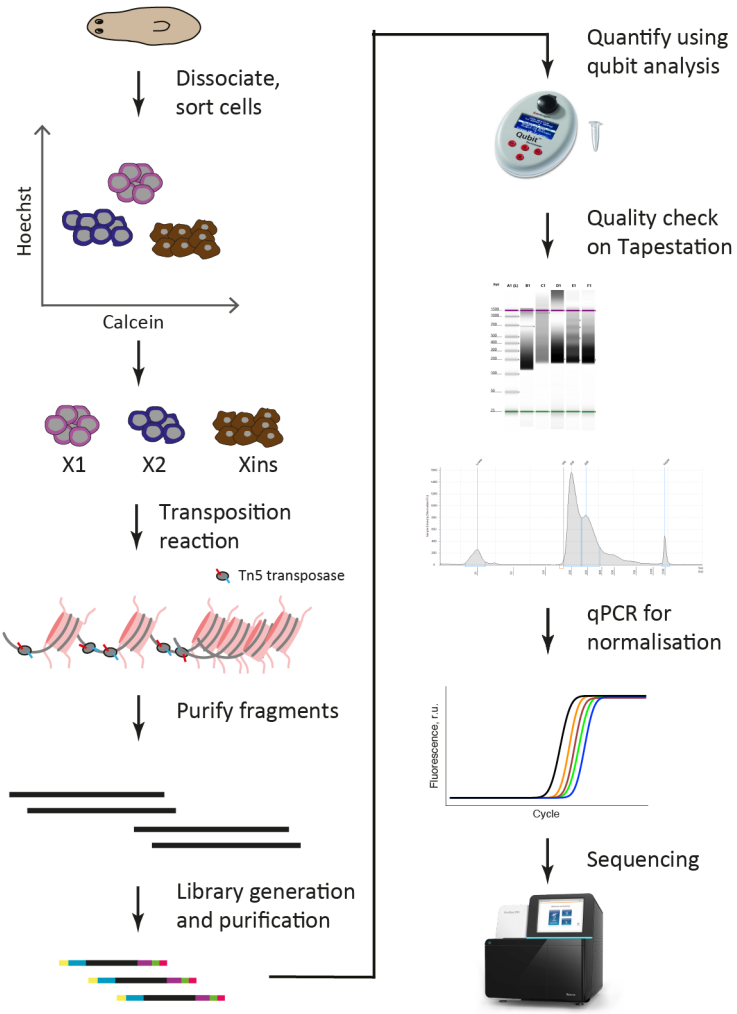

Figure 4.1: Overview of ATAC-seq methodology. Tn5 transposase transposes more often in open chromatin regions than in inaccessible chromatin. The Tn5 transposase then cuts the open chromatin with the simultaneous insertion the adaptors to generate DNA fragments. The transposed DNA is isolated, amplified by PCR and subjected to sequencing.

replication, and the M phase as chromatin condenses). The X1 compartment comprises a heterogeneous population of stem cells in the different stages of the cell-cycle (S/G2/M) while the X2 compartment contains post-mitotic stem cell progeny and G1 stem cells, and the Xins compartment contains differentiated cells. To ensure that the DNA is not over- or under-transposed, the transposition reaction was optimised for each cell population.

Using this approach, we were able to generate libraries with two replicates for X1, X2 and Xins samples. Transposed DNA libraries were checked on TapeStation to check

for the quality of prepared libraries. I found that TapeStation profiles of transposed DNA libraries prepared using 130,000 sorted X1 cells show weaker nucleosome phasing (Fig. 4.2a-b). TapeStation profiles of transposed DNA libraries for approximately 219,000 X2 cells show clear nucleosome phasing (Fig. 4.2c-d), as do profiles of libraries prepared with 248,000 Xins cells (Fig. 4.2e-f). The weaker phasing seen in X1 cells could reflect their inherent heterogeneity and may form an average of different chromatin configurations at the cell cycle phases.

A typical size distribution plot for a good ATAC-seq experiment shows periodical peaks in decreasing size corresponding to the nucleosome-free regions (NFR) (< 100 bp) and mono-, di-, and tri-nucleosomes [Yan et al., 2020]. After pre-processing sequencing reads, the fragment distributions of all six libraries were analysed and the majority of fragments were found to be less than 100 bp, and consequently found to correspond to nucleosome-free regions (Fig. 4.3a-c). As the size of DNA wrapped around a nucleosome is ~ 146 bp, periodicity in fragment size can be observed in X2 and Xins samples corresponding to the presumptive mono- and di-nucleosome length that the Tn5 enzyme cuts around.

Mapped fragments were filtered to only include those with a length of less than 100 bp, as the aim was to visualise the distribution of nucleosome-free regions across the genome. I used MACS2 to call peaks from these filtered NFR reads for each of sub-population of planarian cells, which allows for the determination of genomic sites with the greatest amount of chromatin accessibility [Zhang et al., 2008]. Further qualitative analysis was performed using DiffBind [Stark and Brown, 2016].

Plots based on principal components analysis (PCA) provide insight into how similar samples are. A PCA plot including normalised read counts for all differentially binding sites, was created using DiffBind [Stark and Brown, 2016]. The resulting plot (Fig. 4.3d) shows that the X1 and X2 samples are neither separable in the first (horizontal) nor in the second (vertical) component, while all Xins samples (blue) slow clustering on one side of the second (vertical) component. Spearman's correlation coefficient of the ATAC-seq coverage profiles for all three cell population samples was plotted (Fig. 4.3e). The replicates of each cell population cluster together. The correlation between X1

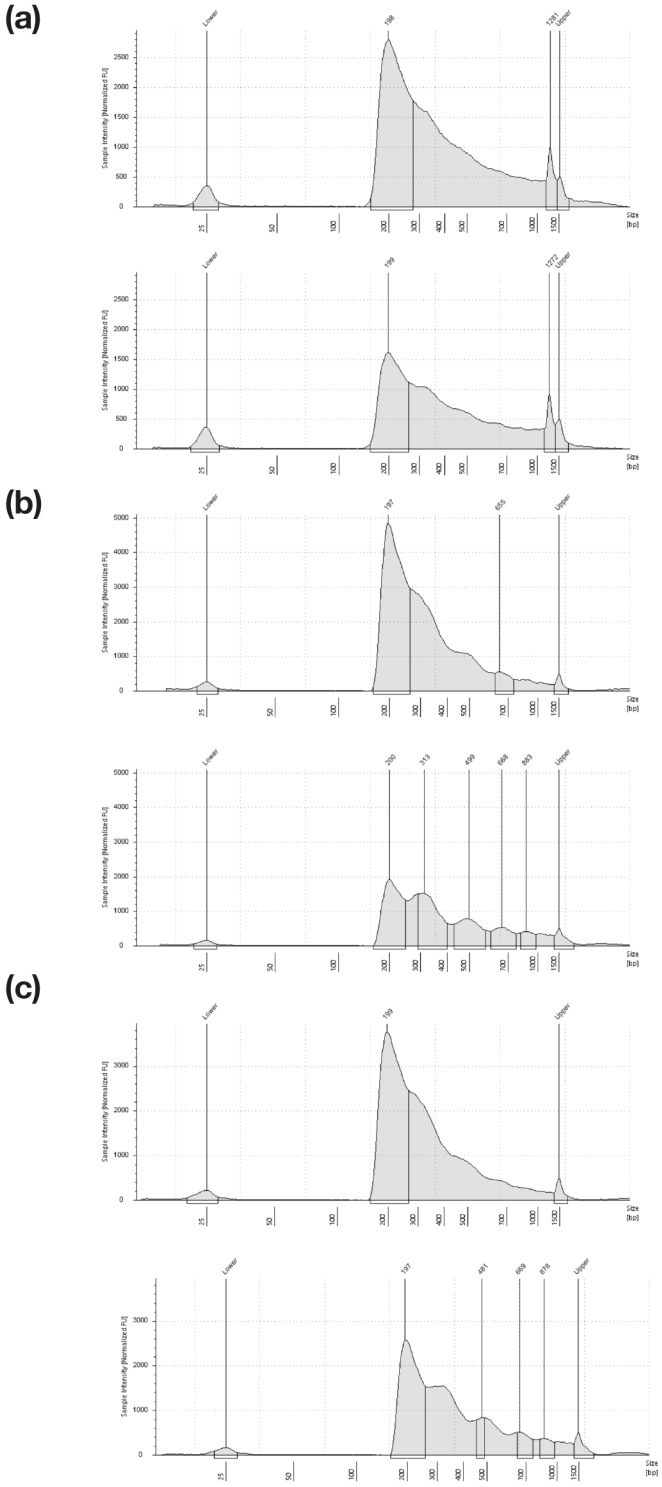

Figure 4.2: TapeStation profiles of transposed DNA libraries for sorted planarian cells. (a) and (b) X1 replicates, (c) and (d) X2 replicates and (e-f) Xins replicates.

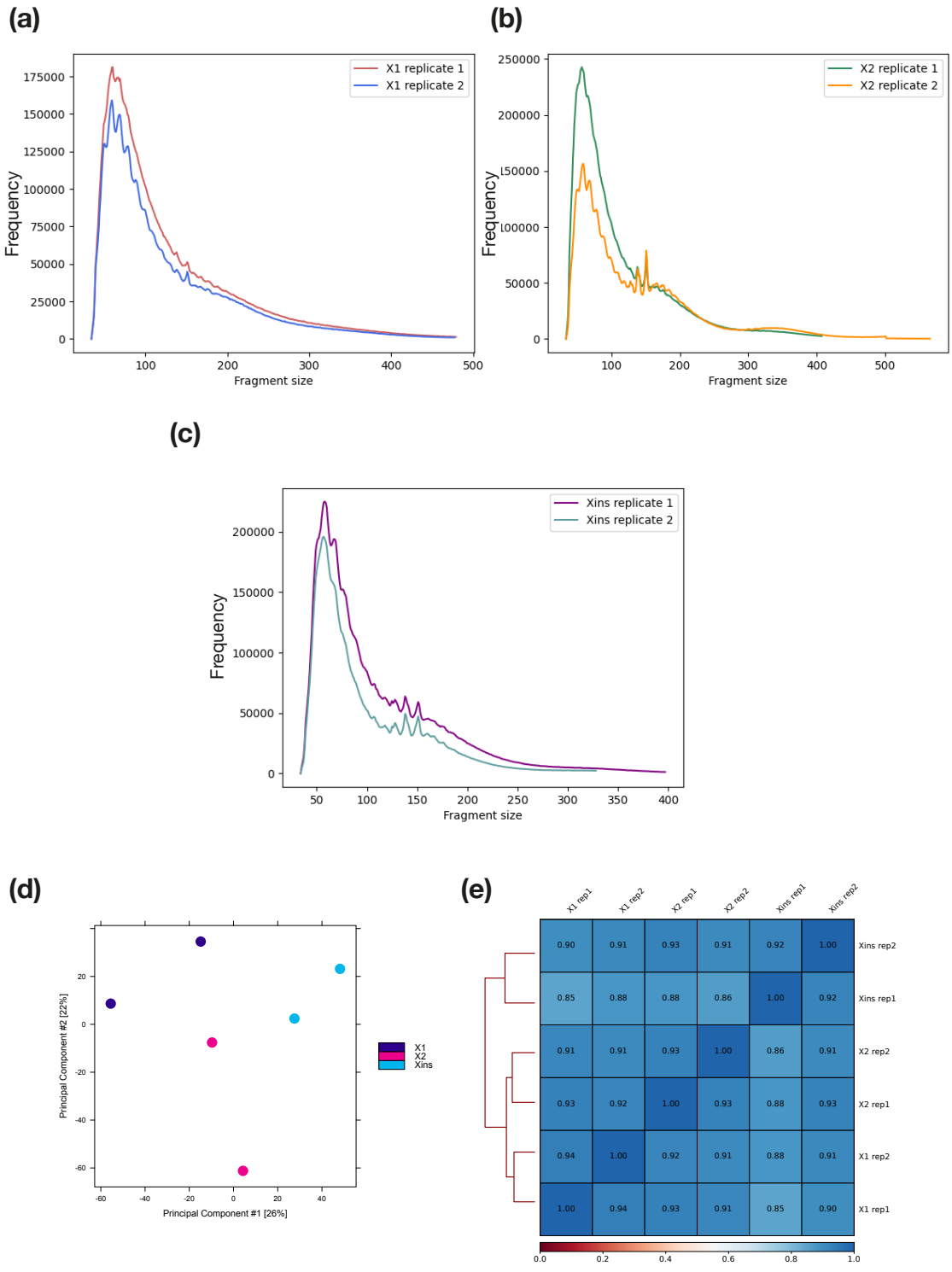

Figure 4.3: Fragment size distribution plots for (a) X1 (b) X2 and (c) Xins ATAC-seq libraries. (d) PCA plot of different ATAC-seq libraries. (e) Correlation of the ATAC-seq coverage profiles for the three cell populations samples.

and Xins was lower than that for either X1 and X2 or X2 and Xins. As expected, a clustering of X1 and X2 samples was observed.

Identification of regions of open chromatin in planarian cell populations

The intersection of BED intervals for MACS2 consensus peaks of X1, X2 and Xins libraries was determined and used to find unique and common regions using Intervene [Khan and Mathelier, 2017]. I found a total of 11,811 high confidence peaks in X1 cells, 22,702 in X2 cells, and 23,926 in Xins cells (Fig. 4.4a). Fewer accessible peaks are common to X1 and Xins cell populations than for either X1 and X2 or X2 and Xins cell populations. Xins cells also show the highest number of unique peaks. This pattern in peak similarities agrees with the pattern of coverage similarity previously seen (Fig. 4.3e). As the stem cells proceed towards lineage commitment and differentiation, the cells appear to have more open, active regions of chromatin. Lineage commitment and cell differentiation progresses by the repression of pluripotency and lineage-inappropriate genes. This might explain the presence of higher numbers of open accessible regions of chromatin in differentiated cells, compared to stem cells and progenitors.

Using the latest annotation of the planarian genome, the accessible peaks were assigned to genomic features for each cell population. It can be seen that the majority of called peaks for each of the three cell populations are in the intergenic regions of the genome, with comparatively fewer peaks being called at promoter-TSS annotated sites (Fig. 4.4b). X2 and Xins cells show a slightly higher number of peaks in introns and exons than in X1 cells (Fig. 4.4b).

Correlation heatmaps were plotted using DiffBind to visualise the differentially accessible regions in the three cell populations. Figure 4.5 shows the binding affinity heatmaps in X1, X2 and Xins cells depicting affinities for differentially acccessible sites. The figure illustrates the clustering of the differentially acessible sites as well as the sample clustering ([Stark and Brown, 2016]. Replicate samples of each cell population cluster together, as expected.

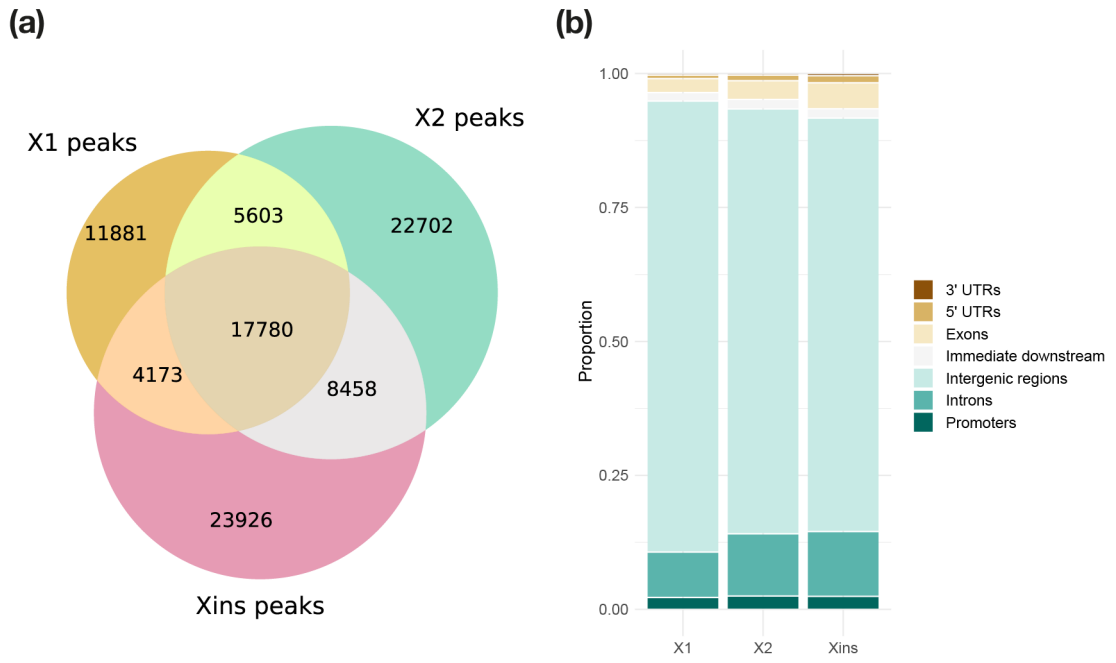

Figure 4.4: (a) Venn diagram of the common peak sets of X1, X2 and Xins cell populations found by Intervene. (b) The genomic location of ATAC-seq peaks in the X1, X2 and Xins populations.

Figure 4.6 shows the MA plots for chromatin accessibility in the three cell populations. MA plots effectively visualise the relationship between Tn5 insertions at each site and the magnitude of change in insertions between cell populations [Stark and Brown, 2016]. An MA plot is obtained for the X1 vs X2 contrast, X1 vs Xins contrast and X2 vs Xins contrast (Fig. 4.6a-c). Each point represents a binding site, with the points in magenta representing sites with differential Tn5 insertion [Stark and Brown, 2016]. The plot shows how the Tn5 insertion sites appear to exhibit a minimum absolute log-fold difference between two and three. It also shows that more Tn5 insertion sites are lost in X1 cells than gained, as evidenced by more red dots below the centre line than above. This agrees with the previous findings that the X2 and Xins cells have more unique accessible peaks than the X1 cells (Fig. 4.4a). I compared each two populations to determine the number of open chromatin regions that were present in one population but not the other. The data showed that the X1 and Xins populations showed the greatest number of non-overlapping regions (Fig. 4.6b). This indicates that a contrasting set of regions are open or active in stem cells and differentiated cells. The transcriptional statuses of

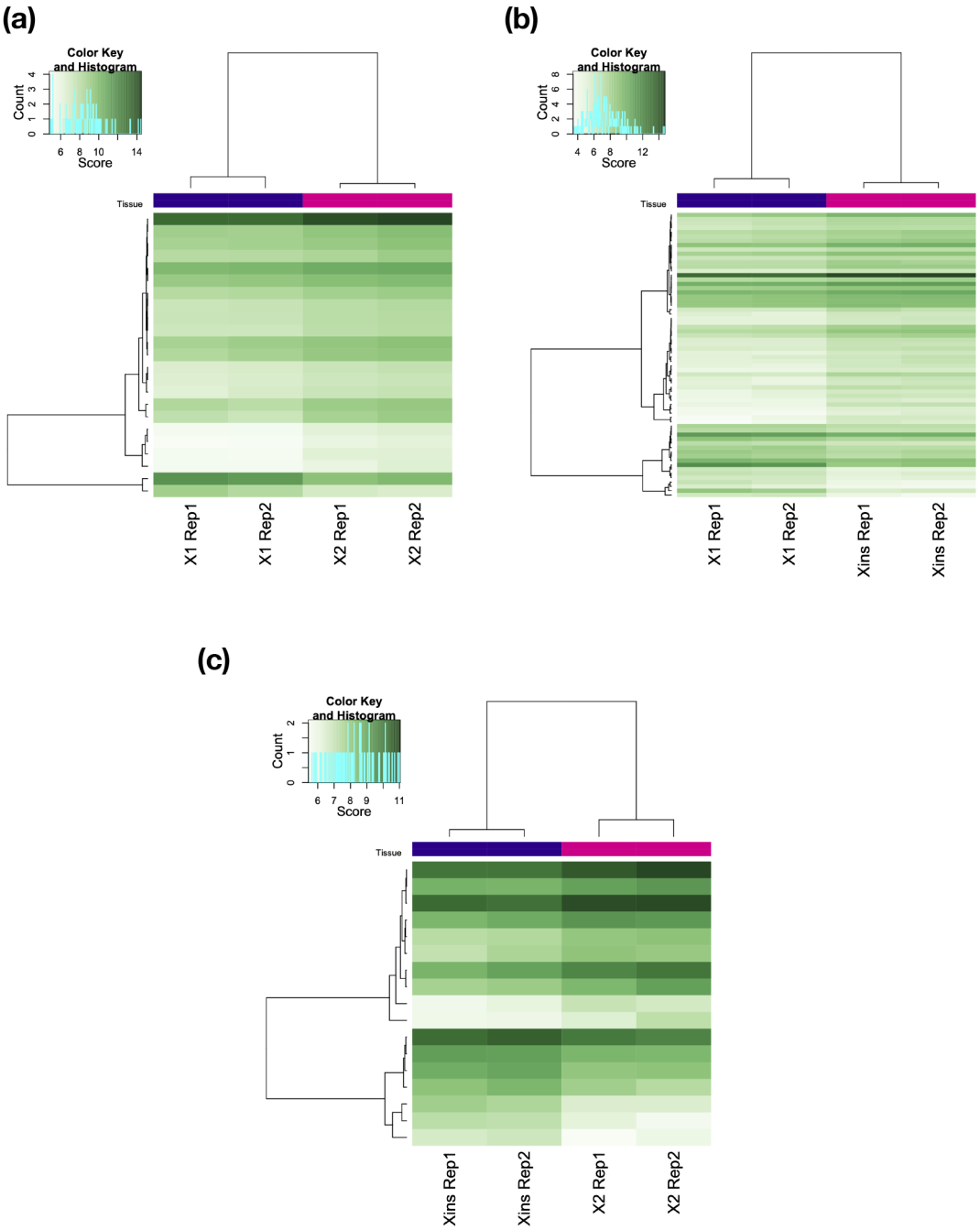

Figure 4.5: Correlation heatmaps visualising the differentially accessible regions in the three cell populations. Binding affinity heatmap showing affinities for differentially acessible sites between (a) X1 and X2, (b) X1 and Xins and (c) X2 and Xins populations. Replicates of each cell population cluster together.

stem cells and differentiated cells are perhaps the most different – stem cells contain active pluripotency genes and repressed lineage commitment genes, while differentiated cells have repressed pluripotency genes and active lineage-appropriate genes.

A heatmap of all the accessible peaks in three cell populations was plotted and it was found that the peaks were less than 100 bp wide (Fig. 4.7).

Chromatin signatures of active enhancers in planarians

Functional enhancers are cis-acting DNA elements that establish long-range interactions with the promoters of genes they regulate [Spicuglia and Vanhille, 2012]. They are key contributors of time- and tissue-specific gene regulation [Natoli, 2010]. Technological advances in ChIP methodologies have led to the characterisation of chromatin signatures of active enhancers. Together with technologies available to help identify accessible chromatin, ChIP has been used in numerous mammalian studies to obtain a global view of cis-regulatory elements, functions of transcription factors, epigenetic regulation of gene expression and, more importantly, on the identity, organisation, and functions of enhancers [Bulger and Groudine, 2011, Daugherty et al., 2017, Jin et al., 2018].

Active enhancers have a histone modification signature containing H3K4me1, H3K4me2, H3K27ac and the additional presence of H3K4me3 and H3K9ac in a subset of active enhancers [Ernst et al., 2011, Spicuglia and Vanhille, 2012, Calo and Wysocka, 2013]. Making use of the ChIP-seq protocol standardised in our laboratory, I used histone marks as a starting point for the identification of enhancer elements. An overview of the protocol can be seen in Fig. 4.8.

I generated ChIP libraries for planarian stem cells (X1 cell compartment) for H3K9ac, H3K27ac and H3K9me3 histone marks with two replicates and sequenced. The specificity of ChIP-grade antibodies was tested using Western blot on planarian total protein lysate and specific bands were seen for all antibodies used (Fig. 4.9a). ChIP libraries generated using the antibodies as well as input control were analysed on TapeStation to verify the size of PCR-enriched fragments and check the size distribution (Fig. 4.9b-c). All samples were of good quality and showed a narrow distribution with a peak size of approximately 275 bp. I analysed previously generated ChIP-seq data with respect to the

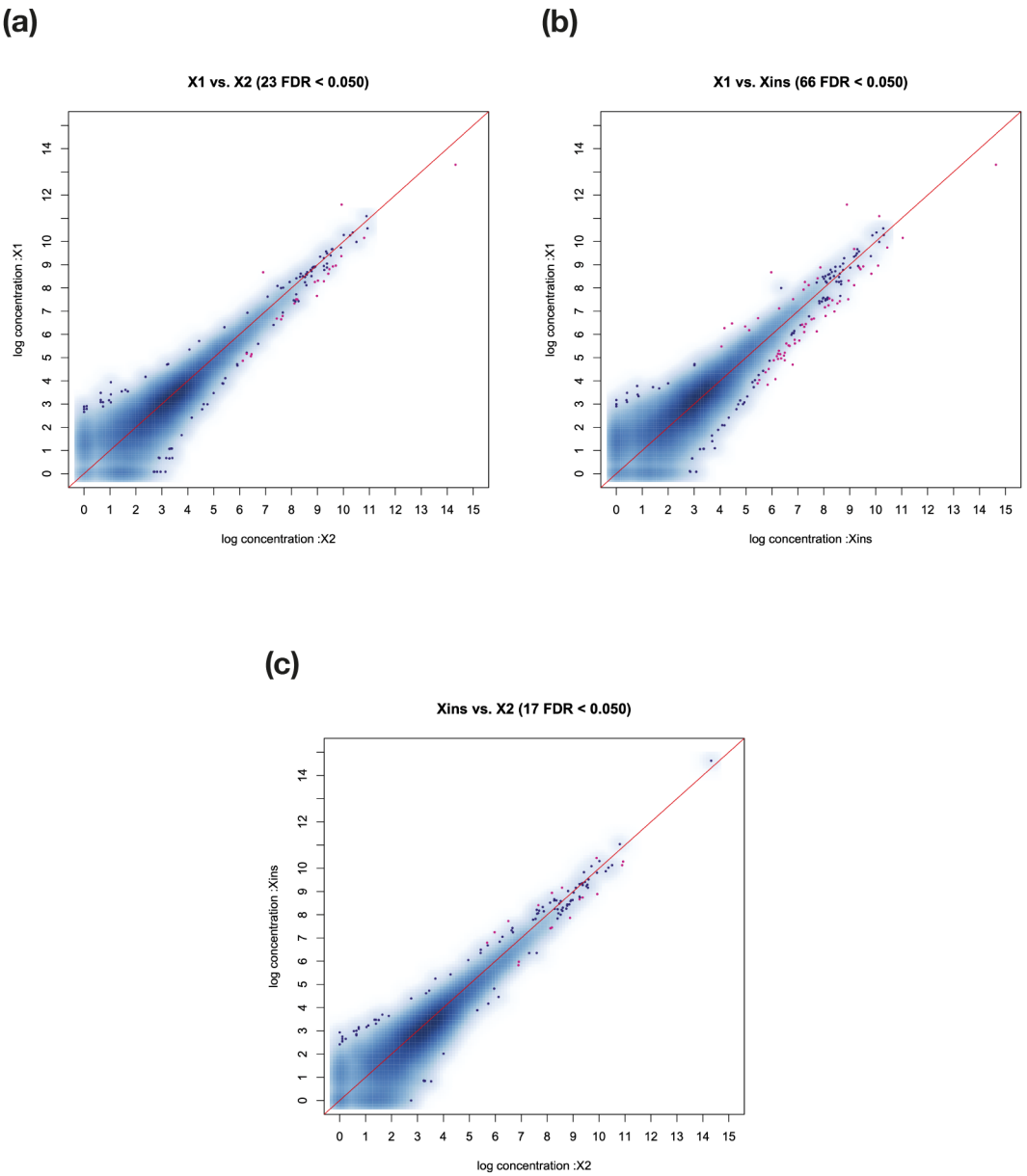

Figure 4.6: MA plots visualising the relationship between the Tn5 insertions at each site for (a) X1 vs X2 contrast, (b) X1 vs Xins contrast and (c) X2 vs Xins contrast. Each point represents an insertion site. The points in magenta represent sites identified as differentially accessible.

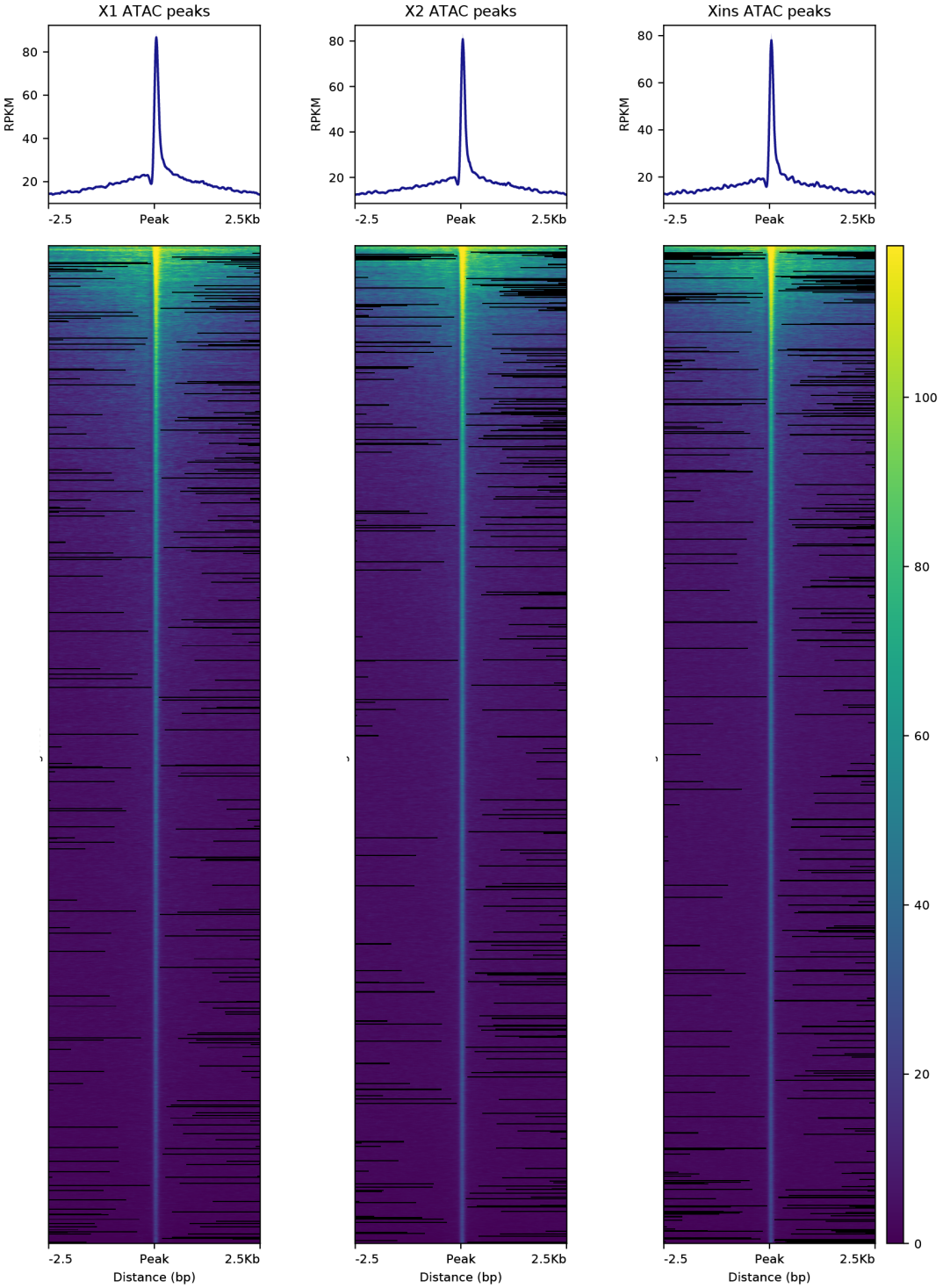

Figure 4.7: The profile of ATAC-seq peaks in the X1, X2, and Xins cell populations. The profiles ATAC peaks called by MACS2 using nucleosome-free reads (< 100 bp) indicate that peaks are generally narrow.

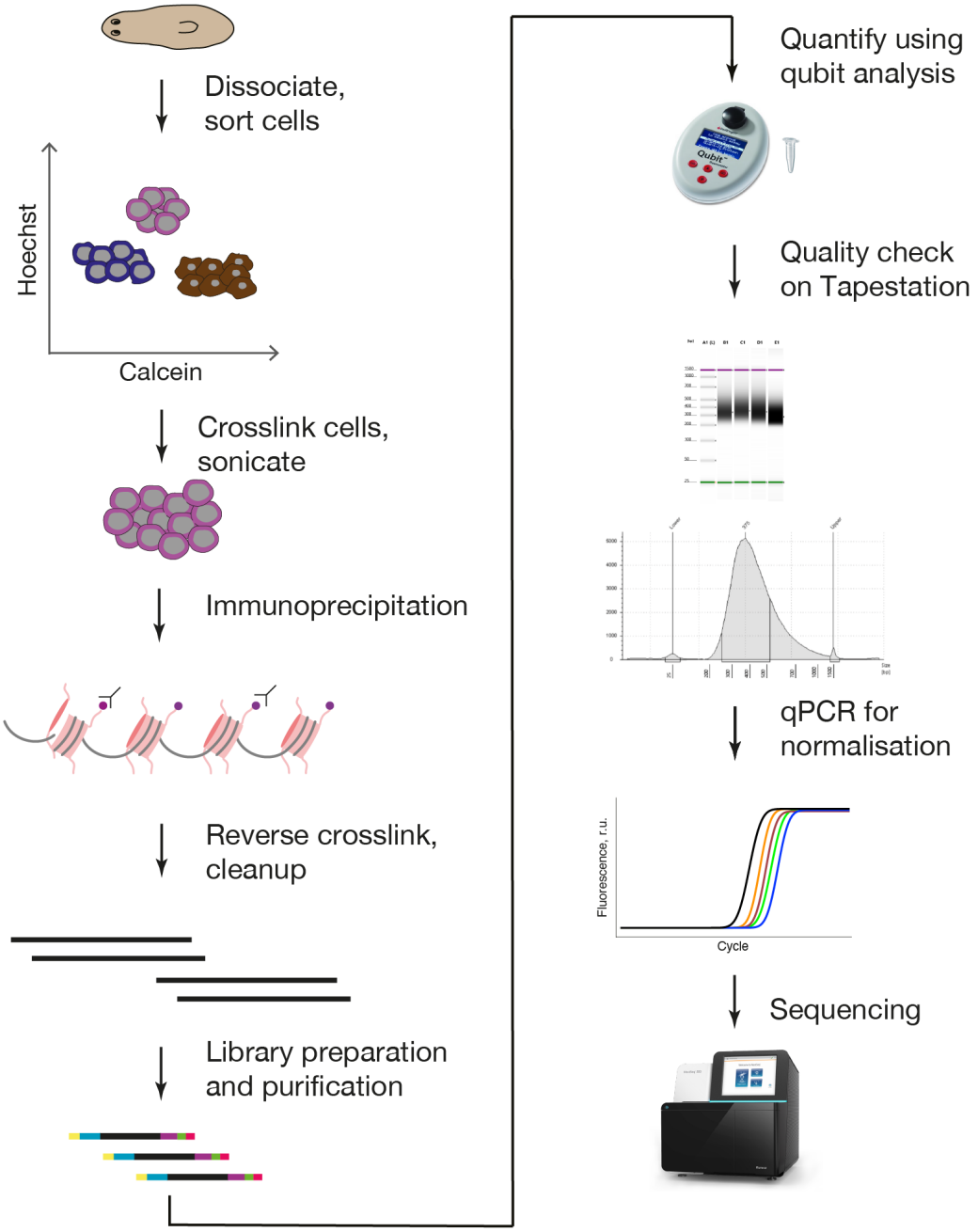

Figure 4.8: An overview of ChIP-seq methodology. Wild-type planarians are dissociated and stem cells are sorted using a flow cytometer. Sonicated chromatin from these cells is used for immunoprecipitation with different antibodies against histone marks. Immunoprecipitated samples are used for library preparation and sequencing. Prior to immunoprecipitation, *Drosophila* S2 chromatin spike-in is used to normalise any technical differences across replicate libraries of an IP.

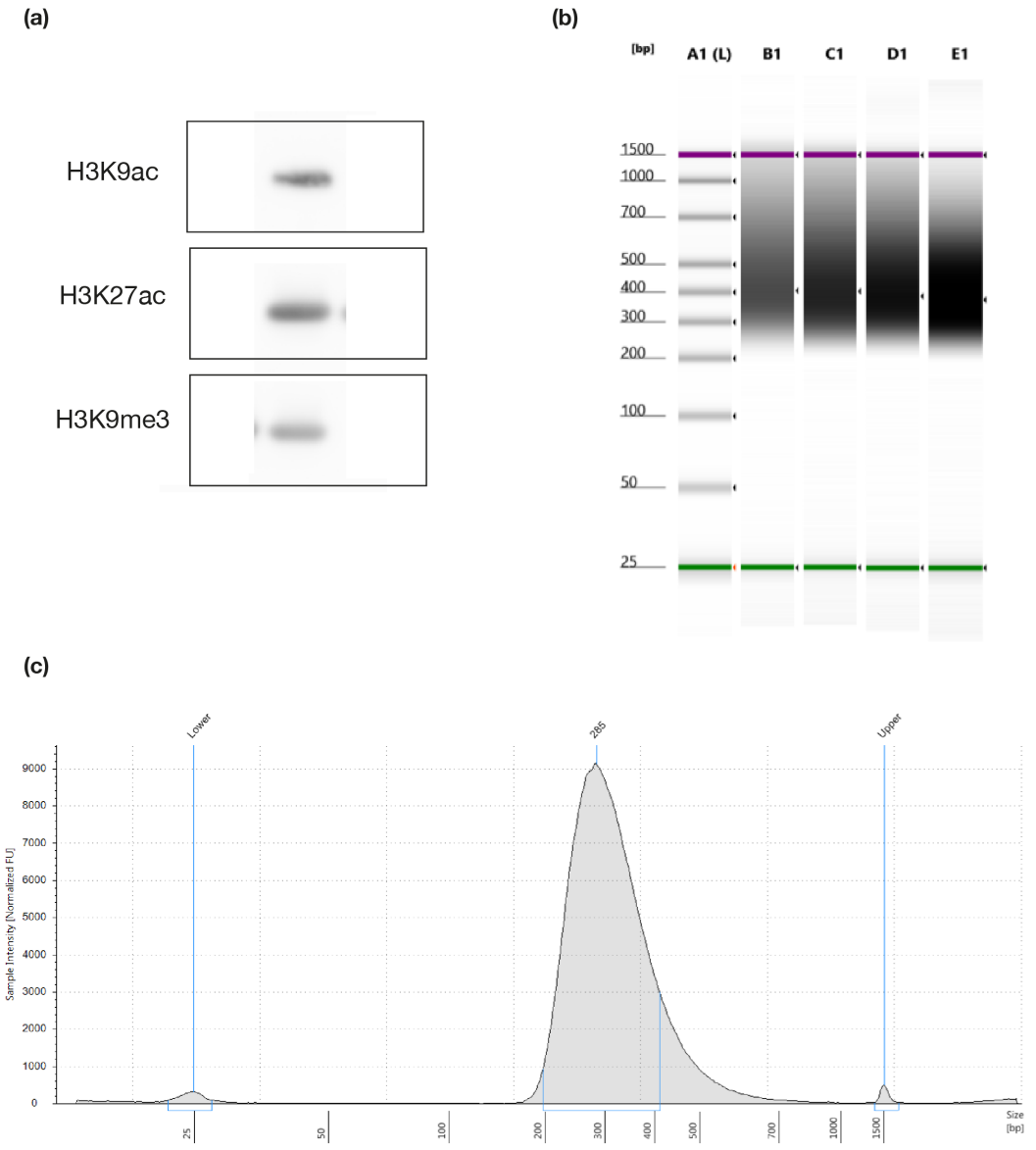

Figure 4.9: Quality control in ChIP-seq library generation. (a) Western blot image showing bands obtained using ChIP-grade antibodies against planarian protein lysate. (b) TapeStation profile of the first replicate of ChIP libraries. A1-Ladder, B1-H3K9ac, C1-H3K27ac, D1-H3K9me3, F1-Input control. (c) Representative image of a typical TapeStation profile showing the size distribution of ChIP libraries.

epigenetic enhancer-associated mark H3K4me1 and H3K4me3 in X1 cells [Mihaylova et al., 2018] and the H3K27ac data that was generated to identify genomic regions indicative of an active enhancer state.

ChIP peaks for H3K27ac and H3K4me1 histone marks were called using MACS2 and found to be generally less than 200-300 bp wide (Fig. 4.10a-b). H3K27ac and

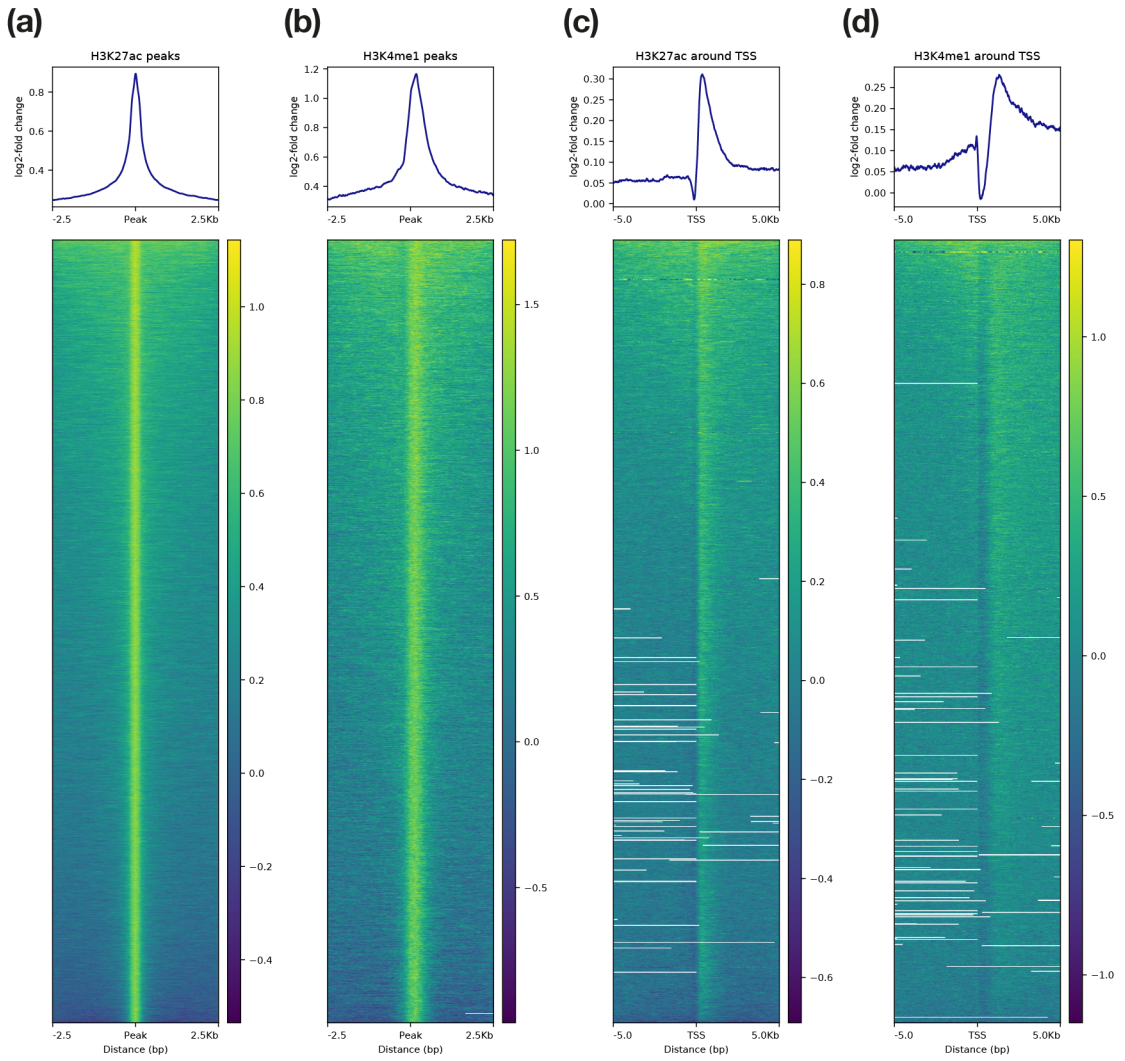

Figure 4.10: Heatmaps visualising H3K27ac and H3K4me1 signals in planarian stem cells. The profile of (a) H3K27ac and (b) H3K4me1 signal. The profile of (c) H3K27ac and (b) H3K4me1 signal around transcription start site (TSS) regions.

H3K4me1 signals were also plotted around TSS regions (Fig. 4.10c-d).

The intersection of BED intervals for MACS2 consensus peaks of H3K27ac and H3K4me1 libraries was determined. A total of 46,829 high confidence H3K27ac peaks and 17,384 high confidence H3K4me1 peaks were found in X1 cells, of which 3516 overlapped (Fig. 4.11a). Using the latest annotation of the planarian genome, the peaks were assigned to genomic features for each histone modification. Overall, a majority of called peaks for each histone modification lay in the intronic regions of the genome (Fig. 4.11b). H3K27ac peaks were slightly more over-represented in exons

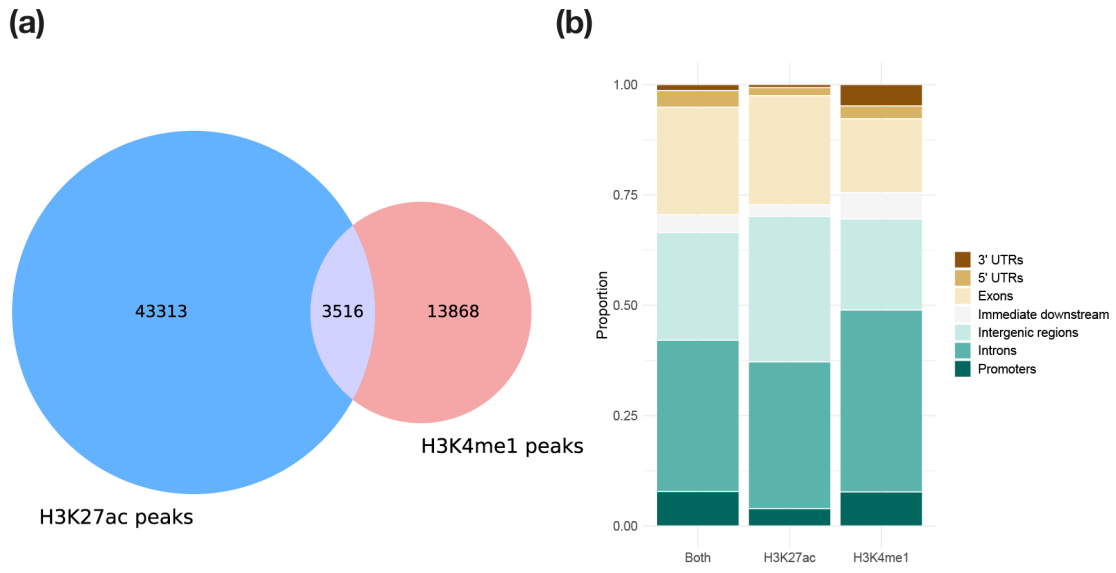

Figure 4.11: (a) Venn diagram showing the unique and common MACS2 called peaks in the two sample sets. (b) The genomic location of individual ChIP peaks as well as peaks containing both H3K27ac and H3K4me1 signal.

and intergenic regions than H3K4me1 peaks.

H3K4me1 signal exhibited two narrow peaks at the peak centre of most H3K27ac peaks (Fig. 4.12a). A peak-valley-peak or bimodal peak of H3K4me1, flanking p300/Pol II/H3K27ac peaks, has been implicated in active mammalian enhancers, and may be evolutionarily conserved within metazoans [Gorkin et al., 2012, Pundhir et al., 2016]. H3K27ac signal around H3K4me1 peaks can be seen to be strongest at the peak centre for almost all peaks (Fig. 4.12b). In summary, this ChIP-seq analysis establishes that enhancers in planarians show signatures also found in other organisms.

Leveraging chromatin accessibility and histone modifications to identify enhancers in planarian stem cells

The analysis of ChIP-seq data of planarian stem cells has identified the presence of conserved chromatin signatures of active enhancers. On the basis of this data, I identify regions with the two histone modifications consistent with enhancer identity and activity, and use ATAC-seq data to pinpoint accessible chromatin overlapping with

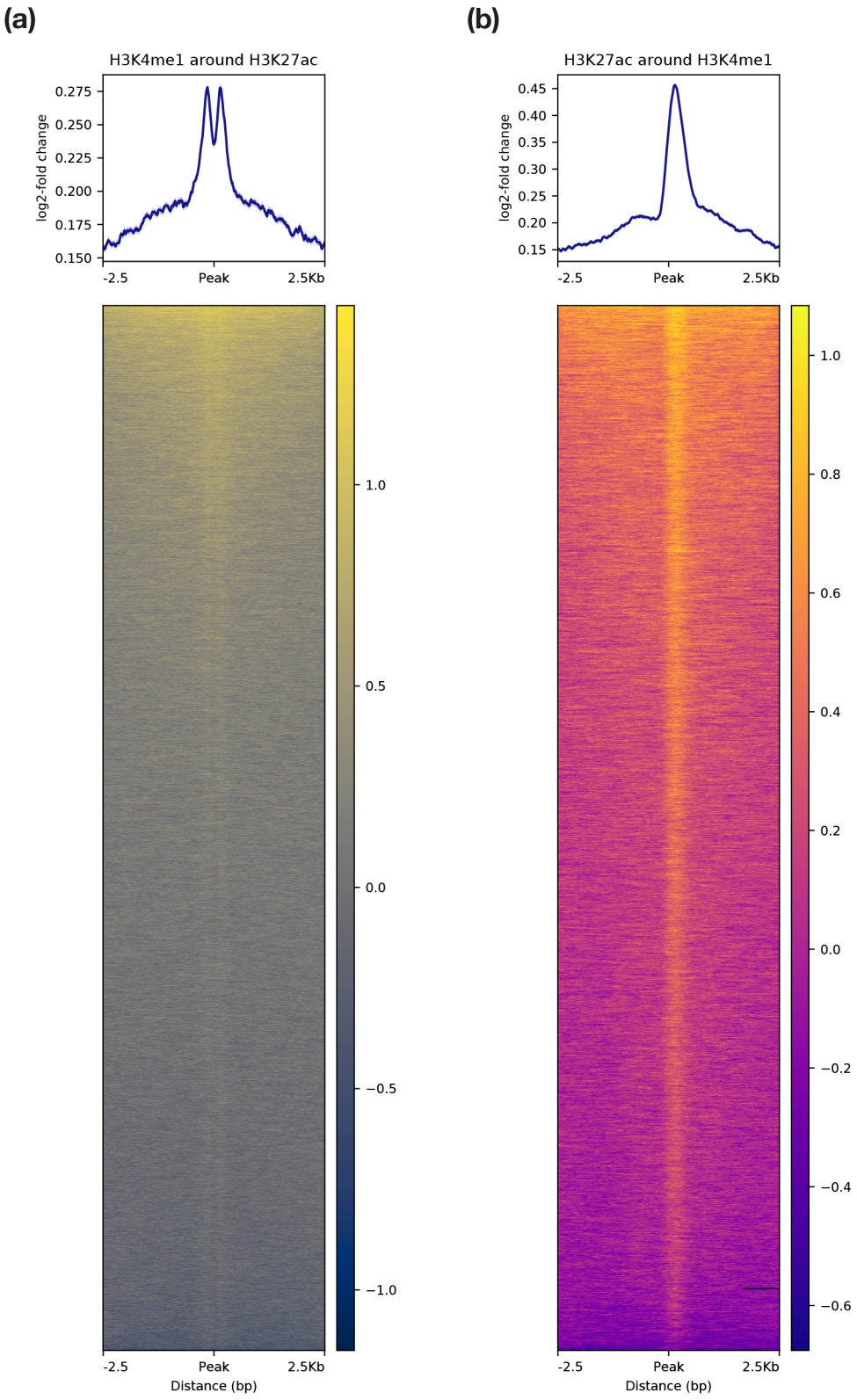

Figure 4.12: The profile of (a) H3K27ac signal (log2-fold change) at H3K4me1 peaks and (b) H3K4me1 signal (log2-fold change) at H3K27ac peaks.

regions that exhibit these enhancer hallmarks, allowing me to locate potential enhancers genome-wide. These analyses were used in the construction of a catalogue of putative enhancer elements in planarian stem cells. ChIP peaks of histone marks were used as the starting point for the gradual selection of putative enhancers. This selection was performed by integrating information from regions of open/accessible chromatin, histone modifications, and RNA-seq coverage in stem cells.

The enhancer identification pipeline involved four criteria – H3K27ac and H3K4me1 signals, H3K4me3 signal, accessibility of chromatin, and RNA evidence. Firstly, I selected regions with overlapping H3K27ac and H3K4me1 peaks. Secondly, I determined the presence of H3K4me3 signal at the peaks. Thirdly, I selected peaks with an open chromatin configuration based on ATAC-seq data. ATAC signals from stem cells peaked at the centre of most intergenic and intronic enhancer-like regions that were filtered based on evidence from histone modifications (Fig. 4.13). Lastly, I identified regions with RNA expression in stem cell transcriptomics data. The intergenic regions with RNA evidence were identified as transcriptionally active putative enhancers (TAPEs). The gradual selection of ChIP-seq peaks resulted in the construction of a catalogue of regions with different combinations of enhancer-like properties (Fig. 4.14a). This selection produced a list of 3294 TAPEs with overlapping H3K27ac-H3K4me1 peaks. Putative enhancer catalogues were also created for those genomic regions that contained only H3K27ac or H3K4me1 peaks (Fig. 4.14b-c).

Verification of enhancer-target gene pairs is not straightforward in planarians due to the absence of transgenic reporters. Therefore, I relied on indirect evidence for the existence of these putative enhancers by correlating chromatin accessibility with target gene expression in stem cells [Duren et al., 2017]. I assigned putative enhancers of the three catalogues to target genes using linear distance as well as the expression of the target gene. The majority of enhancers were located less than 10 000 bp from the nearest gene (Fig. 4.15).

However, assigning an enhancer to the closest gene may not be the most fruitful strategy, as there are many examples of enhancers skipping over nearby genes in favour of more distal targets [Lettice et al., 2003, Moore et al., 2020]. Future experiments

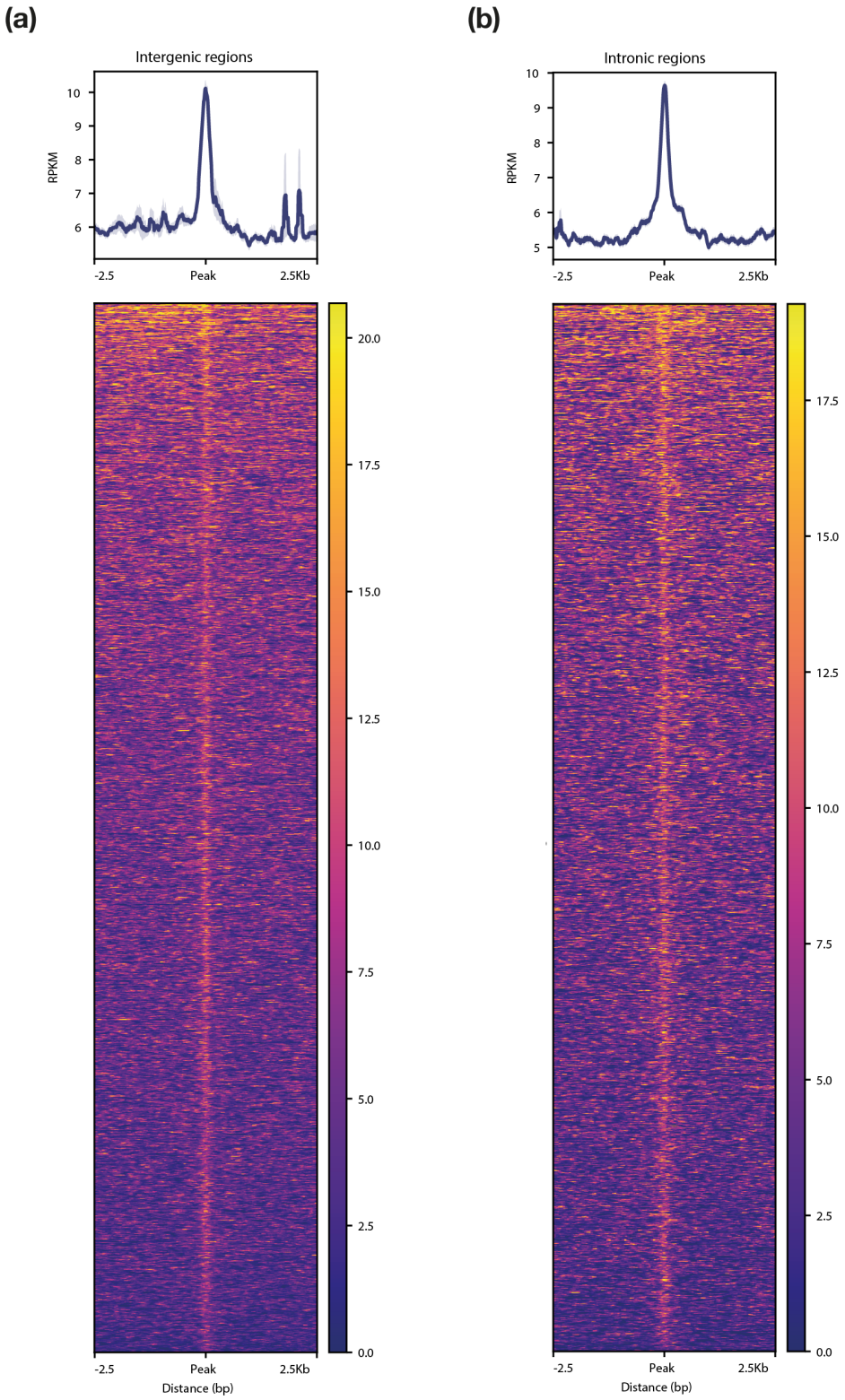

Figure 4.13: ATAC signals of stem cells peaked at the centre of most enhancer-like regions that were filtered based on evidence from histone modifications. Heatmaps of (a) intergenic regions and (b) intronic regions.

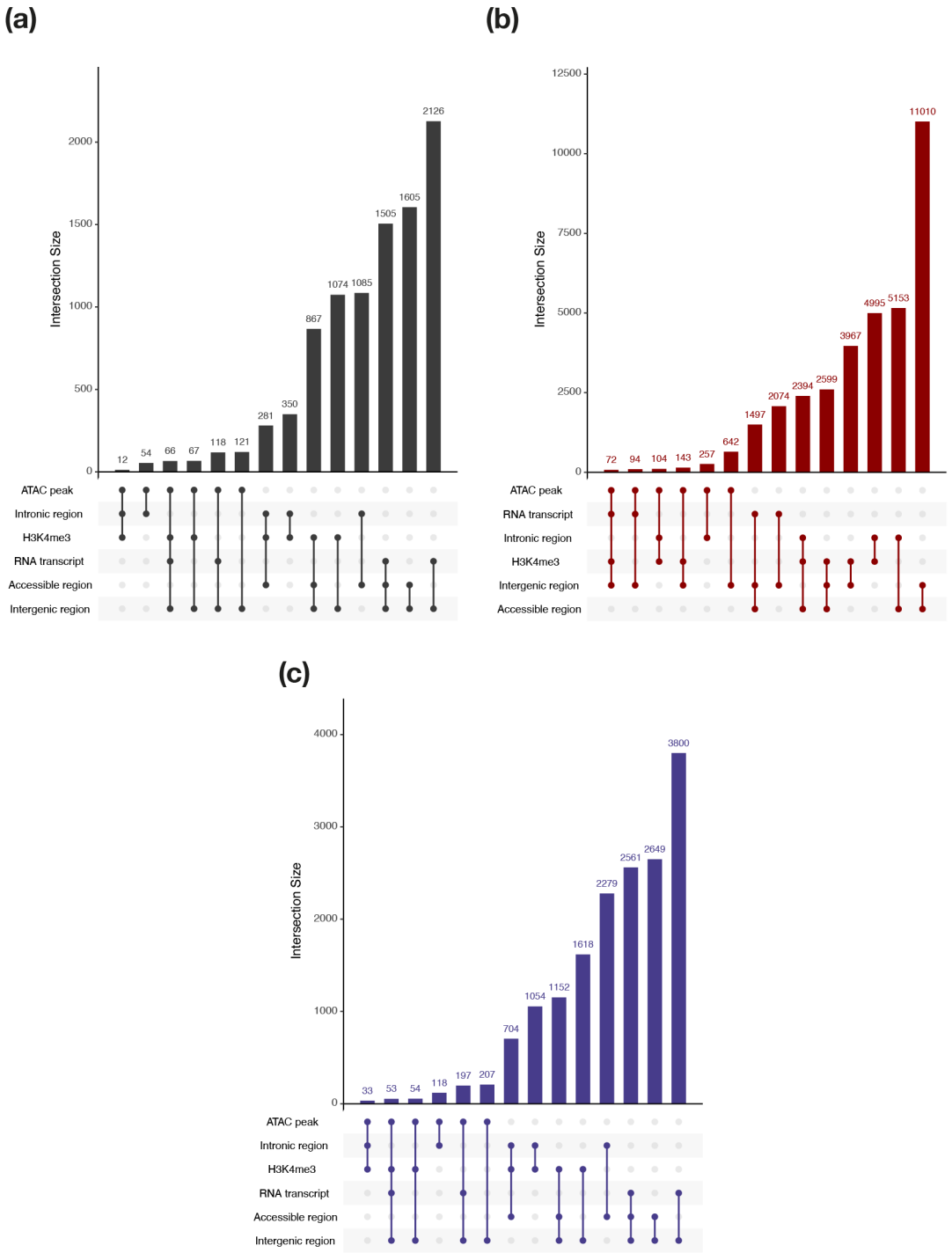

Figure 4.14: Upset plots summarising the different enhancer properties in the constructed enhancer catalogues of planarian stem cells. Putative intergenic and intronic enhancers with different enhancer properties were identified in the planarian genome with (a) both H3K4me1 and H3K27ac peaks, (b) only H3K27ac peaks and (c) only H3K4me1 peaks.

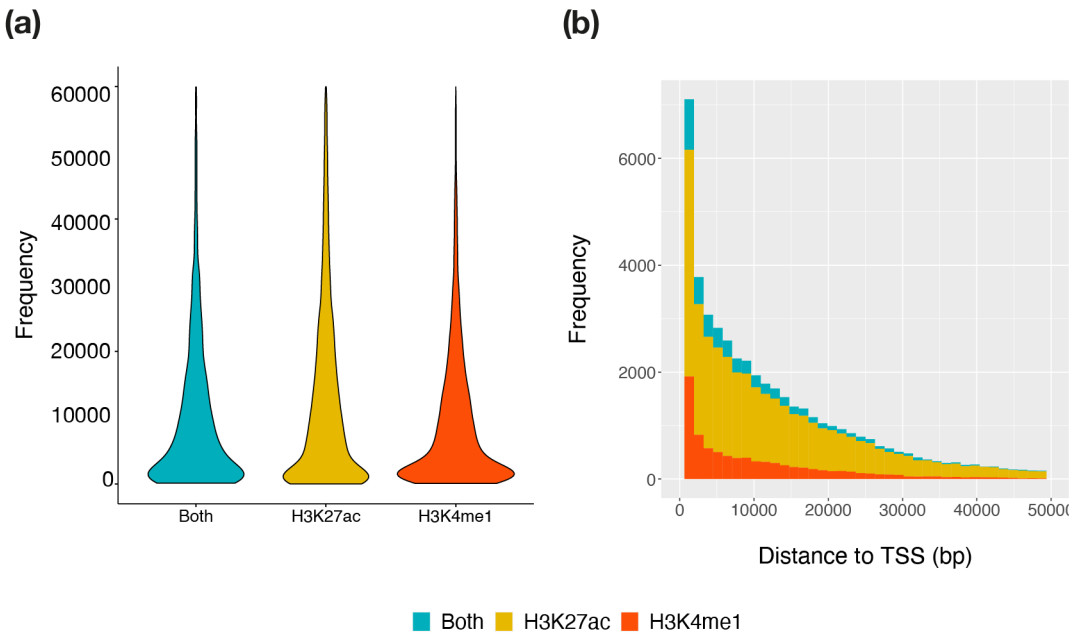

Figure 4.15: Target genes of putative enhancers in planarian stem cells. (a) A violin plot of the enhancers showing that most of the enhancers are located within 10 000 bp from the nearest gene. (b) Histogram showing the frequency of the three enhancer catalogues at different distances from the TSS.

using complementary techniques such as Hi-C (genome-wide chromosome conformation capture) and ChIA-PET (chromatin interaction analysis with paired-end tag sequencing) along with scATAC-seq will aid in the identification of regulatory connections [Li et al., 2012, Lieberman-Aiden et al., 2009, Rao et al., 2014, Tang et al., 2015, Mifsud et al., 2015, Pliner et al., 2018].

4.3 Conclusion

Enhancers that regulate stem cell maintenance and differentiation have remained elusive in planarians. This chapter constitutes a first attempt at uncovering their identity genome-wide. While the identification of enhancers is challenging, the advent of improved sequencing methodologies has led to numerous genomics studies that have revealed signatures of active enhancers [Akhtar-Zaidi et al., 2012, Lara-Astiaso et al., 2014, Daugherty et al., 2017, Carullo et al., 2020]. Together with previous data of

histone modifications in planarians, libraries generated in this chapter contribute to a better understanding of the planarian epigenome.

ATAC-seq and differential accessibility analysis have become the standard methods for mapping regions of open chromatin in various systems [Buenrostro et al., 2013]. Chromatin accessibility is a key feature of regulatory regions including enhancers across animals. Furthermore, enhancers possess distinct histone modification signatures that can be assayed using ChIP-seq [Lara-Astiaso et al., 2014, Shlyueva et al., 2014, Pradeepa, 2017]. H3K27ac and H3K4me1 are known to mark active enhancers in both mammals and Drosophila. Broad H3K4me3 domains mark super-enhancers in cancer tissues [Cao et al., 2017, Li et al., 2019], whereas H3K4me1 alone marks predetermined or poised enhancers [Heintzman et al., 2007, Creyghton et al., 2010, Rada-Iglesias et al., 2011, Ernst et al., 2011, Bonn et al., 2012, Arnold et al., 2013, Calo and Wysocka, 2013, Dorighi et al., 2017]. Nevertheless, not all poised enhancers are activated and some are later actively repressed, while most enhancers are activated without a prior poised state, indicating that the poised enhancer state is not necessarily indicative of a pre-activated state [Rada-Iglesias et al., 2011, Bonn et al., 2012, Koenecke et al., 2017]. Nonetheless, the epigenetic signature of H3K27ac and H3K4me1 appears to be a conserved feature in active enhancers across metazoans and has been successfully used in the identification of enhancers in less traditional model organisms like *Amphimedon* [Gaiti et al., 2017].

The results presented in this chapter for the identification of differentially accessible regions of chromatin in the three FACS-isolated cell populations of planarians form part of the first study to assay chromatin accessibility in planarian stem cells. We achieve the expected result in terms of the size distribution of ATAC seq fragments, and a clear periodicity was seen in X2 and Xins libraries. However, X1 libraries did not show distinct nucleosome peaks. This could be due to over-transposition of DNA in X1 samples. Given that neoblasts are the only dividing cells in planarians and these cells are in various stages of the cell cycle in the X1 population, standardising the optimal tagmentation time might be less straightforward. Regions of open chromatin were identified from nucleosome-free fragments using MACS2 peak caller. A significantly higher number of high confidence peaks was identified in X2 (54,543) and Xins (54,337) populations

compared to X1 (39,437). 45% of these X1 peaks were also found in the X2 and Xins populations. Together, these difference shows that cells undergoing lineage commitment and differentiation have more accessible regulatory regions compared to stem cells.

Using ChIP-seq of two histone modifications – H3K27ac and H3K4me1 – I show that putative enhancers in planarians contain conserved patterns of active enhancers. The newly generated data, together with previous data of RNA coverage, were used to identify putative intergenic and intronic enhancers. This led to the generation of various catalogues of putative enhancers that contain different combinations of enhancer features. Transcriptionally active putative enhancers (TAPEs) with overlapping H3K27ac-H3K4me1 peaks as well as unique ChIP peaks were also identified.

Further analyses of the data will help identify the potential targets of these enhancers as well as conserved transcription factor binding sites. Analysis for transcription factor motif enrichment in the identified catalogue of putative enhancers can help identify motifs that are enriched in these regions. Integration of this information with transcription factor and gene expression data will aid construct gene regulatory networks in planarians. Studying the function of transcription factors together with their regulatory networks in stem cells will lead to a better understanding of how pluripotency and lineage commitment is controlled in planarians.

Analysis of non-coding RNAs will help identify putative enhancer RNAs (eRNAs) in planarians. eRNA transcripts are short, unstable, often bidirectional, have low copy numbers and are often not spliced or polyadenylated [Azofeifa et al., 2018, Carullo et al., 2020]. eRNAs can be identified using nascent RNA sequencing technique such as global nuclear run-on sequencing (GRO-seq) and precision run-on sequencing (PRO-seq) [Wang et al., 2018]. Although potential enhancer-target gene pairs can be identified in planarians, validating them is challenging as no transgenic reporter lines are available. The scientific community can only rely on indirect evidence until enhancer-reporter constructs become accessible in planarians.

Future studies generating the epigenome of X2 and Xins cell populations are essential to identifying cell type specific enhancers and chromatin changes that govern cellular differentiation and lineage commitment. This combined data will help identify enhancers

that are active in individual FACS isolated populations, and therefore indicative of either a stem cell, progenitors or differentiated cell-type specific function. A similar comprehensive study after the knockdown of NuRD components like MBD2/3, p66 and PRC2 complex components like MLL2/3 might be able to disentangle the role of the NuRD complex in planarians. These data will also test if real enhancers were identified using this pipeline. ChIP-seq for acetylation marks, as well as identification of regions of open chromatin after knockdown of *Smed-mbd2/3*, will help determine the changes at promoters and/or enhancers, thereby both elucidating genetic targets of *Smed-mbd2/3* as well as the involvement of the NuRD complex. Investigating the role of MBD2/3 in a DNA methylation-free organism has important evolutionary significance, owing to their association as a methyl-binding protein, and may clarify an important DNA-methylation independent role. ChIP-seq will help resolve whether genes are aberrantly marked by H3K27 acetylation and methylation in both stem cells and stem cell progeny following *Smed-mbd2/3* knockdown. This could also resolve the long standing dispute that the ancestral role of MBD2/3 did not involve binding to methylated DNA.

4.4 Methods

Methods used for cell dissociation and FACS are described in Chapter 3.

Western blotting

To test the reactivity of planarian histone modification epitopes against commercial antibodies, a western blot was carried out. 10 asexual planarians (5-7 mm) were thoroughly rinsed and all the water was removed. 90 µl PBS, 4 µl phenylmethylsulfonyl fluoride and 4 µl $50\times$ cOmplete protease inhibitor (Roche) was added and the animals were homogenised with a Kontes pellet pestle motor. 100 µl $2\times$ Laemmli sample buffer and 10 µl dithiothreitol (DTT) was added, the sample boiled at 100 °C for 5 minutes, and then centrifuged at 13 000 g for 5 minutes. The protein lysate was run on a 4-12 $\%$ Bis/Tris gel and transferred to a PVDF membrane. The PVDF membrane was blocked with 5 $\%$ dry skimmed milk in PBS with 0.05 $\%$ Tween-20 for 1 h, and incubated with

HRP conjugated secondary antibodies (1:4000 dilution). Tubulin was used as a loading control for all western blots and bands were detected using a SuperSignal West Pico kit.

Chromatin immunoprecipitation followed by sequencing

For each experimental replicate (ChIP of three histone marks and one input control), $600,000 - 700,000$ planarian X1 cells were isolated. Immunoprecipitation was performed using ChIP-grade H3K9ac, H3K37ac and H3K9me3 antibodies and a protocol adapted from Dattani et al. [2018a]. Libraries were prepared using NEB Ultra II kit following manufacturers guidelines.

The libraries were quantified using Qubit dsDNA High Sensitivity kit. The quality and size distributions of the libraries were checked on an Agilent 4200 TapeStation system using a TapeStation High Sensitivity DNA ScreenTape. The libraries were individually quantified using the qPCR-based KAPA library quantification kit. Libraries were pooled together to 4nM concentration and then paired-end sequenced at a length of 75 nt on an Illumina NextSeq machine.

ChIP-seq data analysis

ChIP libraries with > 15 million uniquely mapped paired-end fragments were generated with two replicates for each histone mark. Trimmomatic [Bolger et al., 2014] and BWA-MEM (version 0.7.12) [Li and Durbin, 2009] were used to trim and align sequenced reads to the annotated *S. mediterranea* and *D. melanogaster* reference genome [Hoskins et al., 2015]. Duplicate reads were eliminated with Picard tools (version 1.115, `https://broadinstitute.github.io/picard/`). Custom Python scripts were used to filter and separate out read pairs belong to either genome. A normalisation factor to control for technical variations among IP replicates was obtained by counting the number of reads that align to the *Drosophila* genome. A BED file was created by parsing and adding the coordinates representing 100 bp at the centre of the sequence for each read. Coverage tracks were created in the bigwig format using the genomecov method of Bedtools 2.27.0 [Quinlan and Hall, 2010] and the bedgraphtoBigWig tool [Kent et al., 2010]. For each sample and corresponding input, the coverage around 2.5 kb

on either side of the annotated TSS for each annotated locus in 50 bp windows was computed using deepTools' computeMatrix method [Ramírez et al., 2016]. A scaling factor for the input ChIP-seq libraries was determined using deepTools (using the SES method) [Diaz et al., 2012]. In order to generate the final coverage track, used for visualisation and analyses, the mean normalised input coverage was subtracted from the normalised sample coverage. Jupyter notebooks containing ChIP-seq analysis are available at `https://github.com/divya-sridhar/DPhil-thesis/`.

Assay for transposase-accessible chromatin using sequencing

For each experimental replicate, approximately 120,000-250,000 planarian X1, X2 and Xins cells were isolated by sorting dissociated planarian cells. A protocol adapted from Buenrostro et al. [2013] was used. Sorted cells were washed in $1\times$ PBS, and the supernatant was discarded. For cells undergoing lysis, 50 µl of cold lysis buffer (10 mM Tri-Cl (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40) was added and re-suspended by pipetting up and down. Lysed cells were centrifuged at 500 rpm for 10 minutes at 4 °C. The cytoplasmic content within the supernatant was discarded and the nuclei pellet was kept. For cells undergoing transposition without lysis, cells were centrifuged in PBS, supernatant removed and cell pellet kept. 25 µl $2\times$ TD Buffer, 2.5 µl Tn5 Transposase and 22.5 µl of nuclease-free H_2O was added to each nuclei pellet. The pellets were re-suspended by pipetting up and down and then incubated at 37 °C for 60 minutes. The transposed DNA was isolated using a Zymogen Clean and Concentrator Kit, and eluted in 10 µl of EB buffer. The sample was either stored at -20 °C or used directly for PCR amplification and library purification. For each sample, 10 µl of purified transposed DNA, 10 µl of nuclease-free water was combined with 15 µl Nextera PCR Master Mix, 5 µl of PCR primer cocktail, 5 µl Index Primer 1, and 5 µl Index Primer 2. The samples were amplified using 12-14 PCR cycles (72 °C for 3 minutes, 98 °C for 30 seconds, [98 °C for 10 seconds, 63 °C for 30 seconds, 72 °C for 1 minute]). The samples were purified using Agencourt AMPure XP beads. Briefly, $1.8\times$ volume of Agencourt AMPure XP beads was added to each samples and pipetted up and down $10\times$ to mix thoroughly. The PCR-bead mixture was placed on a magnetic rack for 5 minutes,

the supernatant discarded, and washed once with 200 µl 80% EtOH. After drying on rack, to ensure all EtOH removal, the beads were re-suspended in 20 µl H_2O. The fragmentation patterns of the samples were checked on a TapeStation (Agilent). The transposition reaction was optimised by varying time and number of cells used. Fig. 4.2d shows an example of optimum fragmentation patterns while Fig. 4.2a shows sub-optimal transposition of the X1 FACS isolated cells. Samples with the best transposition profiles obtained so far for each cell population were then sequenced.

ATAC-seq data analysis

Libraries were sequenced with a depth of > 50 million reads and we were able to generate > 12 million uniquely mapping paired-end fragments per library, with two replicates in each of the X1, X2 and Xins samples.

Trimmomatic 0.32 [Bolger et al., 2014] and BMA-MEM [Li and Durbin, 2009] were used to trim and map paired-end reads to the sexual genome assembly. In a next step, the resulting data was filtered to retain only uniquely mapped reads with a quality score of at least 10. Duplicates were identified and eliminated with Picard tools (version 1.115, `https://broadinstitute.github.io/picard/`), and the resulting BAM file was filtered by size. Peaks were called using MACS2 in order to determine enriched Tn5 'cutting sites' and extend 5' ends of sequenced reads uniformly in both directions by 100 bp in order to smooth pile-up signals. Subsequently, shared bed intervals between replicates peak files were computed with bedtool's intersectBed method. Furthermore, the Intervene tool [Khan and Mathelier, 2017] was applied to identify unique and common intervals in the three cell populations. In order to compute coverage, the bamCoverage method of deepTools [Ramírez et al., 2016] was applied with a genome binsize of 10 bp (normalised to RPKM) and a smoothing length of 50 bp. Differential binding sites were identified using DiffBind [Pimentel et al., 2017]. ATAC peaks were assigned to genomic regions using CHIPpeakAnno in the following order of precedence ([Zhu et al., 2010]): promoters, immediately downstream, 5' UTRs, 3' UTRs, exons, and introns. The Matplotlib-venn package was used to illustrate the intersections of ATAC-seq peak overlaps computed by Intervene. DeepTools's compute

Matrix and plotHeatmap methods were applied to generate heatmaps of ATAC-seq profiles 2500 bp up- and downstream of the peak centre, and around the TSS, while the Spearman's correlation between ATAC-seq samples was computed using deepTools' multiBigwigSummary [Ramírez et al., 2016]. The enhancer catalogues were visualised using UpSetR (`https://github.com/hms-dbmi/UpSetR`). Putative enhancers were assigned to target genes using linear distance as well as the expression of the target genes in X1 data. Jupyter notebooks containing ATAC-seq analysis are available at `https://github.com/divya-sridhar/DPhil-thesis/`.

5

General discussion and future directions

Contents

Planarians are masters of regeneration and owe this remarkable ability to neoblasts or pluripotent stem cells. Neoblasts, the only dividing cells in planarians, replace cells lost due to normal physiological turnover as well as injury. Planarians are a promising model system to study the epigenetic regulation of pluripotency, stem cell function, cell fate specification, and differentiation. Transcriptomic analyses have shown a molecular conservation between mammalian ESCs and neoblasts [Reddien et al., 2005b, Labbé et al., 2009, Rouhana et al., 2010, Resch et al., 2012, Solana et al., 2012, Önal et al., 2012], suggesting the molecular conservation of key features in stem cell biology across animal pluripotent cells. Planarian homologues of several ESC genes that play a role in stem cell self-renewal and differentiation have been found. These include epigenetic regulators and RNA binding proteins [Scimone et al., 2010, Jaber-Hijazi et al., 2013, Vásquez-Doorman and Petersen, 2016, Solana et al., 2016, Duncan et al., 2015, Mihaylova et al., 2018].

Planarians also permit the study of stem cell heterogeneity and lineage progression of undifferentiated stem cells due to the availability of molecular markers for stem cells and their progeny [van Wolfswinkel and Ketting, 2010, Fincher et al., 2018, Plass et al., 2018, Swapna et al., 2018, Zeng et al., 2018]. Another advantage of using *S. mediterranea* as a model organism for studying epigenetics is the availability of an excellent array of bioinformatics tools to facilitate these studies. These tools include genome assembly, annotations, transcriptome repository and a new contiguous genome [Cantarel et al., 2008, Robb et al., 2015, Brandl et al., 2016, Grohme et al., 2018]. Nevertheless, while planarians are a promising model system for in vivo stem cell biology, we are only beginning to understand the molecular principles governing the associated regulatory mechanisms. Further research into stem cells in planarians and other model organisms could help understand fundamental stem cell properties, including the disentanglement of pluripotency and self-renewal, as well as uncover universal regulatory features.

The last 25 years have been crucial in establishing planarians as a model system to study regeneration and stem cell biology. The cellular and molecular basis of regeneration as well as neoblast pluripotency and differentiation has been well-understood. The past 5 years have shed more light on neoblast heterogeneity and the different cell types in planarians [Fincher et al., 2018, Plass et al., 2018, Swapna et al., 2018, Zeng et al., 2018]. The absence of cell lines and genetic manipulation of neoblasts poses significant challenges for the in-depth understanding of the molecular mechanisms governing cellular processes. Moreover, the epigenetic control of neoblast pluripotency, cell differentiation and fate remain understudied. In light of these hurdles, the primary aim of this thesis has been to understand the epigenetic mechanisms regulating stem cell fate in planarians using an array of next-generation sequencing methods.

5.1 Understanding the role of NuRD complex in planarians

The nucleosome remodelling and deacetylation (NuRD) complex is a transcriptional repressor of gene regulation that modulates the chromatin accessibility of target genes to transcription factors and RNA polymerase II [Xue et al., 1998, Wade et al., 1999,

Shao et al., 2020]. Methyl-CpG-binding domain protein 3 (Mbd3), a scaffolding protein component of the NuRD complex, is a regulator of pluripotency that is essential for mouse embryogenesis [Hendrich et al., 2001]. Deletion of MBD3 in mESCs has been shown to affect their ability to differentiate. MBD3-null cells initiate the formation of embryoid bodies but are unable to commit to later stages of differentiation, as they fail to express lineage-specific markers [Kaji et al., 2006].

In Chapter 2, I focused on the NuRD complex in the context of planarian pluripotent adult stem cells. The knockdown of a NuRD exclusive component, the methyl-binding domain containing gene *Smed-mbd2/3*, leads to perturbed neoblast differentiation but does not affect neoblast self-renewal. This phenotype is exciting as it mirrors observations in mammals, suggesting a fundamental conservation and the possibility of disentangling the key stem cell properties of self-renewal and pluripotency. Post amputation, *Smed-mbd2/3* knockdown animals are able to start blastema formation but are unable to completely regenerate. Regenerating pieces are unable to form any visible eyes or pharynx. During homeostasis, there is overall tissue turn-over and *Smed-mbd2/3* knockdown animals show phenotypic defects like head regression which leads to loss of eyes and eventually death. To understand these defects at a molecular level, I performed in situ hybridisation of known markers. *Smed-mbd2/3* (RNAi) animals show no change in *Smedwi-1* transcript levels or distribution. In order to determine whether the stem cell progenies are affected by knockdown, animals were double-stained using *Smedwi-1* and sigma, zeta or gamma pool probes. After quantifying double-positive cells, I found that the numbers and distributions of sigma, zeta and gamma cells were comparable to those in control animals. The animals showed slightly increased numbers of early epithelial progeny *prog-1* and lower numbers of late epithelial progeny *agat-1*. This change became more significant with time. Together, these results suggest that the loss of MBD2/3 in planarians does not affect stem cell self-renewal or the lineage-committed stem cells. The animals have stem cells that start to differentiate in the epidermal lineage but cannot differentiate into later stages. This defect in differentiation is fatal during regeneration as well as homeostasis.

A similar set of experiments was performed after the knockdown of *Smed-hdac1* and *Smed-p66*. Histone deacetylase (or HDAC1) and p66/GATA2D are two other components of the NuRD complex. Both HDAC1 and p66 are essential for planarian regeneration and homeostasis. Although *Smed-p66* presents phenotypic defects similar to that of MBD2/3, the defects are more pronounced and regenerating animals survive for a shorter time compared to *Smed-mbd2/3*(RNAi) animals. After knockdown of *p66*, animals show no changes in neoblast self-renewal. Knockdown of *p66* in planarians also causes a differentiation defect; moreover, animals are unable to form *agat-1* cells and show an accumulation of *prog-1* cells. Additionally, silencing of *p66* does not appear to affect neoblast heterogeneity. Knockdown of *Smed-p66* led to an increase in photoreceptor neurons (PRNs), but did not affect eye pigment cup cell production, indicating that *p66* acts to suppress PRN production in wild-type worms [Vásquez-Doorman and Petersen, 2016]. Knockdown of S*med-hdac1* results in a much more severe phenotype, as loss of HDAC1 affects neoblast self-renewal. Animals show a "no neoblast" phenotype that is typical after irradiation or ablation of stem cells. Knockdown animals had reduced numbers of *Smedwi-1* cells as well as sigma, zeta and gamma neoblasts. As a result, the number of *prog-1* and *agat-1* cells was also significantly reduced. This difference could be explained by the fact that MBD2/3 and p66 are unique to NuRD while HDAC1 is associated with other complexes. Paralogs of other NuRD complex components – MTA1-like-1, MTA1-like-2, CHD3, CHD4 – were cloned and knocked down in planarians. Using RNAi-mediated knockdowns, it was shown that MTA1-1 and CHD3 were not essential to planarian regeneration, while MTA1-2 and CHD4 were found to be essential. Scimone et al. [2010] showed that *Smed-chd4*(RNAi) animals have lower numbers of *prog-1* cells at 11 days post-RNAi and lower numbers of *agat-1* cells at all timepoints, starting as early as 6 days post knockdown [Scimone et al., 2010].

The functions of *Smed-mbd2/3* are analogous to those of mammalian MBD3 [Jaber-Hijazi et al., 2013]. Mbd3 is required for the stable formation of the NuRD complex in pre-implantation embryos and plays an essential role in the differentiation of mESCs [Kaji et al., 2006]. *Mbd3*$^{-/-}$ ESCs are capable of self-renewal but are unable to differentiate properly, with ESCs defaulting to a trophectodermal fate [Kaji

et al., 2006, 2007, Zhu et al., 2009]. *Mbd3*$^{-/-}$ ESCs inappropriately express genes normally active in preimplantation embryos and generally repressed in ESCs. Furthermore, *Mbd3*$^{-/-}$ ESCs were unable to downregulate expression of undifferentiated cell markers [Kaji et al., 2006]. These cells were unable to progress through differentiation, even though they were able to initiate one early stage, and instead commit towards the trophectoderm lineage [Kaji et al., 2006].

Taken together, this indicates that all NuRD genes have distinct functions in the lineage differentiation process – *Smed-chd4* and *Smed-hdac-1* are required at an earlier point, while *Smed-mbd2/3* is involved at a later point in differentiation. Stem cell proliferation is reduced after the knockdown of *Smed-chd4*, *Smed-hdac-1* and *Smed-RbAp48*, while animals have normal levels of stem cell proliferation and are able to form the blastema after knockdown of *Smed-mbd2/3* and *Smed-p66* [Bonuccelli et al., 2010, Zhu and Pearson, 2013, Jaber-Hijazi et al., 2013, Robb and Alvarado, 2014, Vásquez-Doorman and Petersen, 2016]. The diverging phenotypic responses of different components of NuRD may be attributed to the association of most of the components with other complexes. Mammalian RbAp48 and HDAC-1, for example, have been shown to be a part of the Sin3 deacetylase complex which together with Nanog is involved in the activation of pluripotency factors as well as the suppression of differentiation genes [Cunliffe, 2008, Dattani et al., 2018b]. Additionally, RbAp48 and CHD4 are involved in the chromatin assembly factor (CAF-1) complex which initiates nucleosome assembly by adding histones H3 and H4 onto newly formed DNA and dMec [Marheineke and Krude, 1998, Kunert et al., 2009]. Therefore their role in planarian stem cell maintenance may be due to their association with these complexes [Aboobaker, 2011]. HDAC1 is associated with CoREST/REST, NCoR/SMRT and SHIP1 [Choi et al., 2008, Hayakawa and Nakayama, 2011]. The role of these proteins in stem cell differentiation, however, appears to be related to their participation in the NuRD complex.

The role of the NuRD complex in planarians can, therefore, be studied using the *Smed-mbd2/3* phenotype, as the protein is specific to this complex. Of significance is the fact that DNA methylation is absent in *S. mediterranea* and that the function of NuRD can be assessed without the confounding factor of whether MBD2 (methyl

binding) or MBD3 (hemi-methyl-binding, or non-methyl-binding) is present in NuRD. Additionally, research into planarian MBD2/3 will help resolve the long-standing question of whether the ancestral MBD2/3 guides the NuRD chromatin machinery to target loci independent of DNA methylation. It will be interesting to investigate whether this mechanism has been conserved in mammals [Yildirim et al., 2011, Baubec et al., 2013, Shimbo et al., 2013, Hainer et al., 2016]. The results in this chapter show that *Smed-mbd2/3* is not essential for stem cell maintenance or proliferation but is essential for differentiation. This phenotype is exciting as it mirrors observations in mammals, suggesting a fundamental conservation between neoblasts and mammalian stem cells and the possibility of disentangling the two key stem cell properties of self-renewal and pluripotency.

5.2 Identifying the targets of MBD2/3-NuRD in planarians

Studies from various animals have shown that the NuRD complex is essential for embryonic development. In mammalian ESCs, NuRD-mediated gene silencing contributes to the capacity of ESCs to exhibit pluripotency and self-renewal while preserving their ability to differentiate [Kaji et al., 2006, Denslow and Wade, 2007, Mcdonel et al., 2009]. NuRD directs the deacetylation of target genes, which are then available for trimethylation by PRC2, consequently leading to transcriptional repression [Reynolds et al., 2012b]. Deacetylation fails to occur at these target genes if NuRD function is inhibited. This leads to an increase in acetylation, a subsequent reduction in H3K27me3 due to the inability of the mark to attract PRC2, and an increase in transcription [Reynolds et al., 2012b]. While some specific targets of NuRD are known, understanding the genome-wide regulatory networks regulated in mammalian cells has been challenging for a number of reasons, including a cryptic and complex interplay with DNA methylation.

Studying the role of the NuRD complex in the context of planarian pluripotent adult stem cells – where knockdown of a NuRD exclusive component, *mbd2/3*, leads to perturbed differentiation but does not affect stem cell self-renewal – carries evolutionary significance. The phenotype is exciting as it mirrors observations made in mammals

that suggest a molecular conservation of the role of epigenetic silencing in the cell-fate commitment of pluripotent cells and hint at the possibility of uncoupling the key stem cell properties of self-renewal and pluripotency. This interplay could be more readily studied in *S. mediterranea*, owing to the absence of a major factor, DNA methylation. With this in mind, I assessed the genome-wide regulatory roles of NuRD in planarians by exploiting the *Smed-mbd2/3*(RNAi) phenotype using transcriptomics.

Planarian cells can be sorted into three compartments to yield stem cells, stem cell progeny and differentiated cells. The transcriptomes of X1 and X2 cells in control and *Smed-mbd2/3* knockdown animals were generated 5 days and 10 days post RNAi. The results provide an overview of the transcriptional response of planarians to the loss of MBD2/3. A small number of genes were downregulated after the knockdown of *Smed-mbd2/3* with the only notable transcript being that of *Smed-mbd2/3* itself. No changes were seen in the *Smedwi-1* levels in all datasets which corroborates in situ hybridisation results indicating that MBD2/3 does not play a role in regulating pluripotency in planarian stem cells. An expected downregulation of *Smed-agat-1* transcripts was seen in X2 cells. Transcription factors such as dachshund, bicaudal-like and zfp1-like were upregulated after the knockdown of *mbd2/3*. Several mitofusins and mitochondrial transcripts were upregulated in stem cells at both time points after the inhibition of *mbd2/3*. A majority of genes upregulated after the knockdown of MBD2/3 are uncharacterised with a few containing chromatin modifying activity or DNA binding domains. Whether the uncharacterised targets of MBD2/3 are genes involved in lineage specification and determination of cell fate is yet to be determined. An RNAi-based screen was performed with the aim of identifying new markers for lineage progression and differentiation in planarians. Further experiments using a more robust screen has the potential to identify lineage markers in planarians as well as targets of NuRD complex. Although a majority of the genes and lineage markers still remain functionally unexplored, this work involving global gene-expression profiling holds great promise for unravelling functions of MBD2/3-NuRD complex and the mechanisms that control lineage commitment and differentiation.

5.3 Constructing an enhancer catalogue of planarian stem cells

Enhancers are distal-acting regulatory elements that, together with partner transcription factors, can influence the expression of their target genes. Enhancers have received a lot of attention following recent advancements in sequencing technologies and have been studied to understand time- and tissue-specific regulation during embryonic development and regeneration [Kang et al., 2016, Daugherty et al., 2017, Gehrke et al., 2019, Reddington et al., 2020, Thompson et al., 2020, Pascual-Carreras et al., 2020]. Chromatin accessibility has been used to identify tissue regeneration enhancer elements (TREEs) in zebrafish (heart and fin) [Kang et al., 2016, Goldman et al., 2017, Thompson et al., 2020], Drosophila imaginal wing discs [Harris et al., 2016, Vizcaya-Molina et al., 2018], acoels [Gehrke et al., 2019], planarians [Pascual-Carreras et al., 2020] and Hydra [Murad et al., 2019]. A recent study in zebrafish assessed chromatin accessibility and gene expression to identify dynamic changes during regeneration, revealing thousands of putative enhancer regions [Thompson et al., 2020]. Experiments using transgenic constructs validated the identity of several previously unknown non-coding regulatory sequences near essential genes during fin regeneration [Thompson et al., 2020].

The identities of enhancers and other regulatory elements remain unknown in planarians. Moreover, the role of chromatin modifiers and regulatory elements in the spatio-temporal control of gene expression in different planarian cell-types also remains elusive thus far. In Chapter 4, I described first efforts to identify accessible regions of chromatin in the three planarian cell populations using ATAC-seq. ChIP-seq of planarian neoblasts works robustly and was used to generate histone modification profiles for genes. The histone modification signals for H3K27ac, H3K4me3 and H3K4me1 were correlated with gene expression and open regions of chromatin in neoblasts in order to identify putative enhancers. This can also be extended to both X2 and Xins cells by performing ChIP-seq for the histone marks on these cell populations. Considering the fact that these cell populations contain the various lineage committed progenies and differentiated cells, it may prove challenging to uncover useful information, not least since genes are differentially regulated in different lineages. Single-cell sequencing

methods for accessible regions of chromatin and gene expression profiles could, however, provide better insights into these diverse cell types.

Analysis of the ChIP-seq data showed that the profiles of the histone modifications studied follow well-established patterns of enhancers. A bimodal peak of H3K4me1 flanking H3K27ac peaks was observed. H3K27ac signals around H3K4me1 peaks were seen to be strongest at the peak centre at almost all peaks. A correlation of ChIP profiles with regions of open chromatin in neoblasts was used for the identification of potential enhancers. A series of filters were applied using all three types of datasets to create a catalogue of enhancers with different combinations of enhancer properties. ATAC signals in neoblasts peaked at the centre of most enhancer-like regions that were filtered based on the basis of histone modifications. From the enhancer catalogue created, intergenic enhancers with RNA expression in stem cells were identified as transcriptionally active putative enhancers (TAPEs). Putative enhancers of the three catalogues were assigned to target genes using distance as well as the expression of the target gene. The majority of enhancers were located less than 10 000 bp from the nearest gene. Confirming the identify of these enhancers is difficult in planarians due to the lack of synthetic constructs and genetic manipulations such as transgenics. Future experiments using ChIP-seq against transcription factors, Pol II or p300 can be used to independently verify the identity of enhancers. This would be useful in constructing gene regulatory networks by deducing which transcription factor binds to an enhancer, and the target genes whose expression they control.

The transcriptomes of several flatworm species, along with the latest highly contiguous genome of *S. mediterranea*, are available to the planarian community via PlanMine (`http://planmine.mpi-cbg.de`) [Brandl et al., 2016, Grohme et al., 2018, Rozanski et al., 2019]. A comprehensive study generating ATAC- and ChIP-seq datasets for different cell types is essential and could then be used to create a similar accessible resource for the epigenome of planarians. Together with the current genome and improved transcriptomes, this resource will help future research in the epigenetic regulation of planarians which, in turn, will identify mechanisms that underpin regeneration, lineage-commitment, growth and de-growth, disease, and the unlimited

capacity for self-renewal. The identification of regulatory elements, transcription factors and target genes will help construct gene regulatory networks that underlie lineage commitment in stem cells as well as regulatory differences between different lineage-committed cells. It will also help unravel the regulatory network underlying the transcriptional heterogeneity within stem cells.

5.4 Limitations of the research

Transcriptomics data provides targets regulated by MBD2/3 but, as the interactions of NuRD components in planarians is still unknown, we cannot be certain that the targets of MBD2/3 are indeed the targets of the NuRD complex. Moreover, many of the effects observed are due to molecular changes associated with the manifestation of the phenotype. An in-depth analysis of the expression profile of lineage-specific markers of cells from knockdown animals should be performed to verify which cells can undergo differentiation and which ones fail to differentiate; and, moreover, whether these are the same as known NuRD targets. ChIP-seq following *Smed-mbd2/3* knockdown will help resolve whether genes are aberrantly marked by H3K27 acetylation in both stem cells and stem cell progeny. Further ChIP-seq and ATAC-seq in control and *Smed-mbd2/3*(RNAi) stem cells will help us understand the genome-wide regulatory changes that underpin this phenotype (Fig. 5.1). This will complement the RNA-seq approach that has already been performed. By integrating these three data sets, we can understand how changes in the marks mediated by NuRD (H3K27ac and H3K9ac) lead to changes in enhancer activity (revealed by ATAC-seq) and in gene expression (revealed by RNA-seq). This will allow us to elucidate the key changes associated with loss of pluripotency but maintenance of self-renewal in these stem cells. Alternatively, the development of a ChIP-grade antibody would help to determine the targets of Smed-MBD2/3 protein as well as the interactions with NuRD components.

Although I was able to construct a catalogue of enhancers in planarian stem cells, I was unable to identify enhancers specific to the cell type. The analysis was limited by the absence of histone modification data for X2 and Xins cell compartments. An updated ATAC-seq data set, together with RNA and histone-modification profiles of the

three planarian cell populations, will be instrumental in the identification of cell-type specific enhancers in planarians. Alternatively, the gene regulatory landscape of single planarian cells can be ascertained using single-cell ChIP and ATAC-seq methodologies. This will provide a better understanding of the heterogeneity of planarian stem cells and the identification of gene regulatory networks specific to progression of different lineages.

5.5 Future perspectives

NuRD, Polycomb and Trithorax proteins are essential for development; their mis-expression, mutation or deletion leads to a plethora of developmental conditions which are often lethal [Mills, 2010, Richly et al., 2011, Morishita and Di Luccio, 2011, Steffen and Ringrose, 2014]. Studying the underlying molecular mechanisms of these proteins and their interaction with chromatin is crucial to understanding their biological function, as the crosstalk between these histone modification pathways regulates the recruitment and accumulation of many chromatin modifiers at target sites. Chromatin modifiers read pre-existing histone modifications and maintain transcriptional states even when the initial stimulus is gone. Moreover, the interplay between histone marks and chromatin modifiers of opposing functions is essential for the establishment of transcriptional states amenable to dynamic changes in gene expression during development.

The generation of antibodies to known planarian proteins could be useful in the purification of the NuRD complex. The biochemical characterisation of the complex can, in turn, lead to the identification of all components and accessory proteins. An antibody against Smed-MBD2/3 would constitute an excellent starting point, as MBD2/3 bridges the association between MTA1-3 and the deacetylase core with GATAD2A/B and CHD3/4. Further biochemical and structural characterisation of the interactions within NuRD will be useful in establishing the presence of different NuRD complexes. Distinct NuRD complexes may be specifically targeted to active gene bodies through the interaction of additional components with target DNA. Large-scale genome-wide mapping targets of different epigenetic regulators will help untangle the interactions between various proteins. However, the identification of NuRD complex targets might be difficult because many components have independent functions or associate with

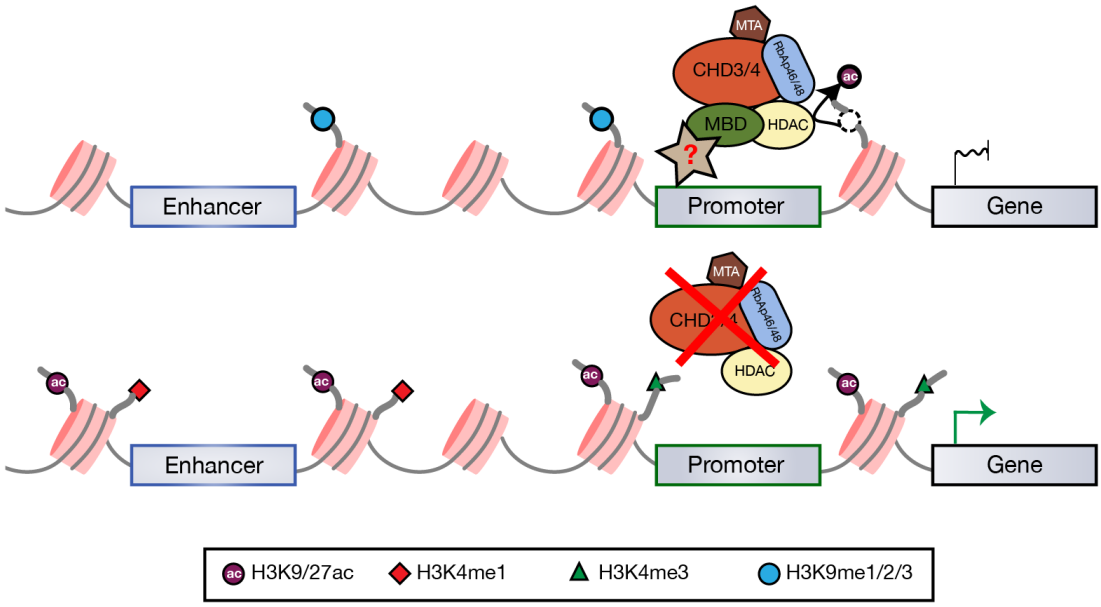

Figure 5.1: An ideal future experiment. ChIP-seq and ATAC-seq in control and *Smed-mbd2/3*(RNAi) stem cells will help us identify the genome-wide targets of MBD-NuRD by determining changes in histone marks and chromatin accessibility. This will complement the RNA-seq approach that has already been done in this thesis.

other complexes. An ideal experiment will obtain transcriptomics and ChIP-seq dataset for all NuRD components and use the overlapping targets to identify NuRD specific targets. A smaller experiment might involve using an antibody against MBD2/3, as it is unique to NuRD. The major drawback of this would be that low or subtle differences in their localisation may be difficult to detect.

The interplay between NuRD and Polycomb complex is yet another interesting area of study, as NuRD together with PRC2 (Polycomb repressive complex 2) co-regulate the expression of a subset of developmentally important genes through the mediation of an H3K27 acetylation to methylation switch that promotes gene repression [Reynolds et al., 2012b]. How NuRD is recruited to target sites, and whether the recognition of methylated DNA by MBD2/3 is crucial for this, remains an important question in the field. Recruitment of NuRD to various targets can be attributed in part to the diverse complexes it is able to form. Planarians offer themselves as an ideal model to unravel this puzzle due to the lack of DNA methylation. The recognition of targets through MBD2/3 in this case is independent of methyl-CpG and could help uncover

if the ancestral role of MBD2/3 is independent of DNA methylation.

All core components of mammalian NuRD have two to three paralogs which may confer different biological functions to the complex and target the complex to different sites. Furthermore, the association of NuRD with an array of accessory proteins like transcriptional factors and chromatin binding proteins is also important for its recruitment to genomic targets. RNA-, ChIP- and ATAC-seq datasets from all wild-type animals or controls will also be useful in the identification and characterisation of enhancers of different cell types. Single cell transcriptomics in planarians have led to the identification of more cell types and a better understanding of the heterogeneity within neoblasts [Fincher et al., 2018, Plass et al., 2018, Zeng et al., 2018]. Single-cell ChIP- and ATAC-seq can be used to identify interactions between different chromatin modifiers as well as gene regulatory networks. Chromatin conformation capture (3C) based sequencing can also be used for the identification of enhancers, enhancer targets as well as enhancer-promoter interactions by the biochemical mapping of DNA looping interactions [Dekker et al., 2002, He et al., 2014]. These genomics approaches can therefore be used to create an epigenetic landscape of individual planarian cells. Furthermore, the identification of regulatory differences in cells can lead to a better understanding of lineage progression and cell fate in planarians.

5.6 Conclusion

In this work, I present the first efforts to identify accessible regions of chromatin in the three planarian cell compartments using ATAC-seq. Chromatin accessibility, together with histone modification signatures in stem cells, was used to identify enhancers in planarians. Using the data from different sequencing methods, I constructed a catalogue of putative enhancers with different combinations of enhancer properties, shedding light on the function of MBD2/3, and identified the genes regulated by MBD2/3. Elucidating the mechanisms behind the recruitment of MBD2/3-NuRD and its role at gene regulatory regions will contribute to understanding the function of NuRD and MBD2/3.

This work presents a number of opportunities for the continued study of the epigenetic regulation of planarian stem cells. Of particular interest are the means by which planarian

neoblasts exit pluripotency and how their lineage fate is determined. Elucidating this mechanism will not only improve our understanding of the specification of the different planarian cell type but also permit comparisons between mammalian systems and provide insights into long-standing questions in the evolution of chromatin remodelling pathways. In addition to understanding the determination of stem cell fate, continued investigation of epigenetic mechanisms in a highly regenerative model organism would aid in the broader understanding of the regulation of stem cell pluripotency and unlock the secrets of planarians' extraordinary regenerative abilities.

Bibliography

A. A. Aboobaker. Planarian stem cells: A simple paradigm for regeneration. *Trends in Cell Biology*, 21(5):304–311, 2011.

C. E. Adler and A. Sánchez Alvarado. Types or States? Cellular Dynamics and Regenerative Potential. *Trends in Cell Biology*, 25(11):687–696, 2015.

N. Agarwal, T. Hardt, A. Brero, D. Nowak, U. Rothbauer, A. Becker, H. Leonhardt, and M. C. Cardoso. MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Research*, 35(16):5402–5408, 2007.

B. Akhtar-Zaidi, R. Cowper-Sal · lari, O. Corradin, A. Saiakhova, C. F. Bartels, D. Balasubramanian, L. Myeroff, J. Lutterbaugh, A. Jarrar, M. F. Kalady, J. Willis3, J. H. Moore, P. J. Tesar, T. Laframboise, S. Markowitz, M. Lupien, and P. C. Scacheri. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*, 336(6082): 736–739, 2012.

R. Albalat. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus Branchiostoma floridae. *Development Genes and Evolution*, 218(11-12):691–701, 2008.

R. Albalat, J. Mart, C. Can, J. Martí-Solans, and C. Cañestro. DNA methylation in amphioxus : from ancestral functions to new roles in vertebrates. *Briefings in functional genomics*, 11(2):142–155, 2012.

A. Alié, T. Hayashi, I. Sugimura, M. Manuel, W. Sugano, A. Mano, N. Satoh, K. Agata, and N. Funayama. The ancestral gene repertoire of animal stem cells. *Proceedings of the National Academy of Sciences*, page 201514789, 2015.

A. S. Alvarado and H. Kang. Multicellularity, stem cells, and the neoblasts of the planarian Schmidtea mediterranea. *Experimental Cell Research*, 306:299–308, 2005.

J. Anderson, C. L. Salzer, and J. P. Kumar. Regulation of the retinal determination gene dachshund in the embryonic head and developing eye of Drosophila. *Developmental Biology*, 297(2):536–549, 2006.

C. D. Arnold, D. Gerlach, C. Stelzer, Ł. M. Boryń, M. Rath, and A. Stark. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339(6123): 1074–1077, 2013.

J. G. Azofeifa, M. A. Allen, J. R. Hendrix, T. Read, J. D. Rubin, and R. D. Dowell. Enhancer RNA profiling predicts transcription factor activity. *Genome Research*, 28 (3):334–344, 2018.

J. Baguñà, E. Saló, and C. Auladell. Regeneration and pattern formation in planarians III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development*, 107:77–86, 1989.

A. Bahat, A. Goldman, Y. Zaltsman, D. H. Khan, C. Halperin, E. Amzallag, V. Krupalnik, M. Mullokandov, A. Silberman, A. Erez, A. D. Schimmer, J. H. Hanna, and A. Gross. MTCH2-mediated mitochondrial fusion drives exit from naïve pluripotency in embryonic stem cells. *Nature Communications*, 9(1):1–11, 2018.

A. J. Bannister and T. Kouzarides. Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395, 2011.

X. Bao, A. J. Rubin, K. Qu, J. Zhang, P. G. Giresi, H. Y. Chang, and P. A. Khavari. A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biology*, 16(1):1–17, 2015.

O. Bar-Nur, H. A. Russ, S. Efrat, and N. Benvenisty. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell*, 9(1):17–23, 2011.

J. Basta and M. Rauchman. The Nucleosome Remodeling and Deacetylase (NuRD) complex in development and disease. *Translational Research*, 165(1):36–47, 2015.

T. Baubec, R. Ivánek, F. Lienert, and D. Schübeler. Methylation-dependent and -independent genomic targeting principles of the mbd protein family. *Cell*, 153(2): 480–492, 2013.

H. I. Baymaz, A. Fournier, S. Laget, Z. Ji, P. W. T. C. Jansen, A. H. Smits, L. Ferry, A. Mensinga, I. Poser, A. Sharrocks, P.-a. Defossez, and M. Vermeulen. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics*, pages 2179–2189, 2014.

P. B. Becker and J. L. Workman. Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.*, 2013.

R. Beckervordersandforth, B. Ebert, I. Schäffner, J. Moss, C. Stockburger, K. Friedland, K. Steib, J. V. Wittgenstein, C. Redecker, S. M. Hölter, W. Xiang, W. Wurst, A. F. Schinder, G.-l. Ming, N. Toni, and S. Jessberger. Role of Mitochondrial Metabolism in the Control of Early Lineage Progression and Aging Phenotypes in Adult Hippocampal Neurogenesis. *Neuron*, 93(3):560–573, 2017.

B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Medisableissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2):315–326, 2006.

A. Bird. DNA methylation patterns and epigenetic memory DNA methylation patterns and epigenetic memory. *Genes & Development*, pages 6–21, 2002.

M. J. Blythe, D. Kao, S. Malla, J. Rowsell, R. Wilson, D. Evans, J. Jowett, A. Hall, V. Lemay, S. Lam, and A. Aziz Aboobaker. A dual platform approach to transcript discovery for the planarian schmidtea mediterranea to establish RNAseq for stem cell and regeneration biology. *PLoS ONE*, 5(12), 2010.

J. Boeke, O. Ammerpohl, S. Kegel, U. Moehren, and R. Renkawitz. The minimal repression domain of MBD2b overlaps with the Methyl-CpG binding domain and binds directly to Sin3A. *Journal of Biological Chemistry*, 2000.

A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014.

S. Bonn, R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczyåski, A. Riddell, and E. E. Fdisableurlong. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics*, 44(2):148–156, 2012.

L. Bonuccelli, L. Rossi, A. Lena, V. Scarcelli, G. Rainaldi, M. Evangelista, P. Iacopetti, V. Gremigni, and A. Salvetti. An RbAp48-like gene regulates adult stem cells in planarians. *Jounal of Cell Science*, 2:690–698, 2010.

L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, 122(5):947–956, 2005.

M. Brackertz, J. Boeke, R. Zhang, and R. Renkawitz. Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. *Journal of Biological Chemistry*, 277(43):40958–40966, 2002.

H. Brandl, H. K. Moon, M. Vila-Farré, S. Y. Liu, I. Henry, and J. C. Rink. PlanMine - A mineable resource of planarian biology and biodiversity. *Nucleic Acids Research*, 44 (D1):D764–D773, 2016.

N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527, 2016.

J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin , DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12), 2013.

J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 2015(January):21.29.1–21.29.9, 2015.

M. Bulger and M. Groudine. Functional and mechanistic diversity of distal transcription enhancers. *Cell*, 144(3):327–339, 2011.

Y. Cai, E.-J. Geutjes, K. de Lint, P. Roepman, L. Bruurs, L.-R. Yu, W. Wang, J. van Blijswijk, H. Mohammad, I. de Rink, R. Bernards, and S. Baylin. The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. *Oncogene*, 33(17):2157–2168, 2014.

E. Calo and J. Wysocka. Modification of enhancer chromatin: what, how and why? *Molecular Cell*, 49(5), 2013.

B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, and M. Yandell. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18(1):188–196, 2008.

F. Cao, Y. Fang, H. K. Tan, Y. Goh, J. Y. H. Choy, B. T. H. Koh, J. Hao Tan, N. Bertin, A. Ramadass, E. Hunter, J. Green, M. Salter, A. Akoulitchev, W. Wang, W. J. Chng, D. G. Tenen, and M. J. Fullwood. Super-enhancers and broad h3k4me3 domains form complex gene regulatory circuits involving chromatin interactions. *Scientific Reports*, 7(1):1–14, 2017.

N. V. Carullo, R. A. Phillips, R. C. Simon, S. A. Roman Soto, J. E. Hinds, A. J. Salisbury, J. S. Revanna, K. D. Bunner, L. Ianov, F. A. Sultan, K. E. Savell, C. A. Gersbach, and J. J. Day. Enhancer RNAs predict enhancer–gene regulatory links and are critical for enhancer function in neuronal systems. *Nucleic Acids Research*, 48(17):9550–9570, 2020.

R. R. Catarino and A. Stark. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes and Development*, 32(3-4):202–223, 2018.

E. Choi, C. Han, I. Park, B. Lee, S. Jin, H. Choi, H. K. Do, Y. P. Zee, E. M. Eddy, and C. Cho. A novel germ cell-specific protein, SHIP1, forms a complex with chromatin remodeling activity during spermatogenesis. *Journal of Biological Chemistry*, 283(50): 35283–35294, 2008.

J. M. Cramer, D. Pohlmann, F. Gomez, L. Mark, B. Kornegay, C. Hall, E. Siraliev-Perez, N. M. Walavalkar, M. J. Sperlazza, S. Bilinovich, J. W. Prokop, A. L. Hill, and D. C. Williams Jr. Methylation specific targeting of a chromatin remodeling complex from sponges to humans. *Scientific Reports*, pages 1–15, 2017.

M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50): 21931–21936, 2010.

V. T. Cunliffe. Eloquent silence: developmental functions of Class I histone deacetylases. *Current Opinion in Genetics and Development*, 18(5):404–410, 2008.

A. Dattani, D. Kao, Y. Mihaylova, P. Abnave, S. Hughes, A. Lai, S. Sahu, and A. A. Aboobaker. Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells. *Genome research*, 28(10): 1543–1554, 2018a.

A. Dattani, D. Sridhar, and A. A. Aboobaker. Planarian flatworms as a new model system for understanding the epigenetic regulation of stem cell pluripotency and differentiation. *Seminars in Cell & Developmental Biology*, 2018b.

A. C. Daugherty, R. W. Yeo, J. D. Buenrostro, W. J. Greenleaf, A. Kundaje, and A. Brunet. Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. *Genome Research*, 27(12):2096–2107, 2017.

J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, 2002.

S. A. Denslow and P. A. Wade. The human Mi-2/NuRD complex and gene regulation. *Oncogene*, pages 5433–5438, 2007.

A. Dhasarathy and P. A. Wade. The MBD protein family-Reading an epigenetic mark? *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 647 (1-2):39–43, 2008.

A. Diaz, K. Park, D. A. Lim, and J. S. Song. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol.*, 23(1):1–7, 2012.

K. M. Dorighi, T. Swigut, T. Henriques, N. V. Bhanu, B. S. Scruggs, N. Nady, C. D. Still, B. A. Garcia, K. Adelman, and J. Wysocka. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell*, 66(4):568–576.e4, 2017.

Q. Du, P.-L. Luu, C. Stirzaker, and S. J. Clark. Methyl-CpG-binding domain proteins : readers of the epigenome. *Epigenomics*, 7:1051–1073, 2015.

E. M. Duncan, A. D. Chitsazan, C. W. Seidel, and A. S. Alvarado. Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *Cell Reports*, 13: 2741–2755, 2015.

Z. Duren, X. Chen, R. Jiang, Y. Wang, and W. H. Wong. Modeling gene regulation from paired expression and chromatin accessibility data. *Proceedings of the National Academy of Sciences of the United States of America*, 114(25):E4914–E4923, 2017.

G. T. Eisenhoffer, H. Kang, and A. S. Alvarado. Molecular analysis of stem cells and their descendents during cell turnover and regeneration in the planarian Schmidtea mediterranea. *Cell Stem Cell*, 3(3):1841–1850, 2008.

J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, and B. E. Bernstein. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

C. Falandry, G. Fourel, V. Galy, T. Ristriani, B. Horard, E. Bensimon, G. Salles, E. Gilson, and F. Magdinier. CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. *Journal of Biological Chemistry*, 285(26):20234–20241, 2010.

M. R. Fantappie, E. Rodrigues, P. Gimba, and F. D. Rumjanek. Lack of DNA methylation in Schistosoma mansoni. *Experimental Parasitology*, 166:162–166, 2001.

Q. Feng, R. Cao, L. Xia, H. Erdjument-Bromage, P. Tempst, and Y. Zhang. Identification and Functional Characterization of the p66/p68 Components of the MeCP1 Complex. *Molecular and Cellular Biology*, 22(2):536–546, 2002.

S. Feng, S. J. Cokus, X. Zhang, P.-y. Chen, M. Bostick, and M. G. Goll. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 2010.

C. T. Fincher, O. Wurtzel, T. de Hoog, K. M. Kravarik, and P. W. Reddien. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science*, 1736(April): eaaq1736, 2018.

D. J. Forsthoefel, N. P. James, D. J. Escobar, J. M. Stary, A. P. Vieira, F. A. Waters, and P. A. Newmark. An RNAi Screen Reveals Intestinal Regulators of Branching Morphogenesis, Differentiation, and Stem Cell Proliferation in Planarians. *Developmental Cell*, 23(4):691–704, 2012.

W. Fu, Y. Liu, and H. Yin. Mitochondrial Dynamics : Biogenesis , Fission , Fusion , and Mitophagy in the Regulation of Stem Cell Behaviors. *Stem Cells International*, 2019, 2019.

N. Fujita, S. Watanabe, T. Ichimura, S. Tsuruzoe, Y. Shinkai, M. Tachibana, T. Chiba, and M. Nakao. Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *Journal of Biological Chemistry*, 278(26):24132–24138, 2003.

F. Fuks, P. J. Hurd, D. Wolf, X. Nan, A. P. Bird, and T. Kouzarides. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *Journal of Biological Chemistry*, 278(6):4035–4040, 2003.

F. Gaiti, K. Jindrich, S. L. Fernandez-Valverde, K. E. Roper, B. M. Degnan, and M. Tanurdžić. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *eLife*, 6:1–33, 2017.

A. R. Gehrke, E. Neverett, Y. J. Luo, A. Brandt, L. Ricci, R. E. Hulett, A. Gompers, J. Graham Ruby, D. S. Rokhsar, P. W. Reddien, and M. Srivastava. Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science*, 363(6432), 2019.

K. K. Geyer, C. M. R. López, I. W. Chalmers, S. E. Munshi, M. Truscott, J. Heald, M. J. Wilkinson, and K. F. Hoffmann. Cytosine methylation regulates oviposition in the pathogenic blood fluke Schistosoma mansoni. *Nature Communications*, 2011.

K. K. Geyer, I. W. Chalmers, N. Mackintosh, J. E. Hirst, R. Geoghegan, M. Badets, P. M. Brophy, K. Brehm, and K. F. Hoffmann. Cytosine methylation is a conserved epigenetic feature found throughout the phylum Platyhelminthes. *BMC Genomics*, pages 1–13, 2013.

J. A. Goldman, G. Kuzu, N. Lee, J. Karasik, M. Gemberling, M. J. Foglia, R. Karra, A. L. Dickson, F. Sun, M. Y. Tolstorukov, and K. D. Poss. Resolving Heart Regeneration by Replacement Histone Profiling. *Physiology & behavior*, 176(10):139–148, 2017.

D. U. Gorkin, D. Lee, X. Reed, C. Fletez-Brant, S. L. Bessling, S. K. Loftus, M. A. Beer, W. J. Pavan, and A. S. McCallion. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Research*, 22(11):2290–2301, 2012.

G. Grafi, A. Zemach, and L. Pitto. Methyl-CpG-binding domain (MBD) proteins in plants. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1769(5-6): 287–294, 2007.

M. A. Grohme, S. Schlodisableissnig, A. Rozanski, M. Pippel, G. R. Young, S. Winkler, H. Brandl, I. Henry, A. Dahl, S. Powell, M. Hiller, E. Myers, and J. C. Rink. The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms. *Nature Publishing Group*, 2018.

X. L. Guezennec, M. Vermeulen, A. B. Brinkman, W. A. M. Hoeijmakers, A. Cohen, E. Lasonder, and H. G. Stunnenberg. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and cellular biology*, 26(3):843–851, 2006.

K. Günther, M. Rust, J. Leers, T. Boettger, M. Scharfe, M. Jarek, M. Bartkuhn, and R. Renkawitz. Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Research*, 41(5): 3010–3021, 2013.

T. Guo, A. H. Peters, and P. A. Newmark. A bruno-like Gene Is Required for Stem Cell Maintenance in Planarians. *Developmental Cell*, 11(2):159–169, 2006.

A. Gutierrez and R. J. Sommer. Evolution of dnmt -2 and mbd -2-like genes in the free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis briggsae. *Nucleic Acids Research*, 32(21):6388–6396, 2004.

S. J. Hainer, K. N. McCannell, J. Yu, L. S. Ee, L. J. Zhu, O. J. Rando, and T. G. Fazzio. DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells. *eLife*, 5(November 2016):1–24, 2016.

R. E. Harris, L. Setiawan, J. Saul, and I. K. Hariharan. Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs. *eLife*, 5(FEBRUARY2016):1–28, 2016.

R. D. Hawkins, G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker, and B. Ren. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5):479–491, 2010.

T. Hayakawa and J. I. Nakayama. Physiological roles of class i HDAC complex and histone demethylase. *Journal of Biomedicine and Biotechnology*, 2011, 2011.

T. Hayashi, M. Asami, S. Higuchi, N. Shibata, and K. Agata. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development Growth and Differentiation*, 48(6):371–380, 2006.

B. He, C. Chen, L. Teng, and K. Tan. Global view of enhancer-promoter interactome in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 111(21), 2014.

N. D. Heintzman, R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3):311–318, 2007.

B. Hendrich and A. Bird. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology*, 18(11):6538–47, 1998.

B. Hendrich and S. Tweedie. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics*, 19(5):269–277, 2003.

B. Hendrich, U. Hardeland, H.-h. Ng, J. Jiricny, and A. Bird. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, 401 (September):301–304, 1999.

B. Hendrich, J. Guy, B. Ramsahoye, V. A. Wilson, and A. Bird. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes & Development*, pages 710–723, 2001.

E. F. Heuston, C. A. Keller, J. Lichtenberg, B. Giardine, S. M. Anderson, R. C. Hardison, and D. M. Bodine. Establishment of regulatory elements during erythro-megakaryopoiesis identifies hematopoietic lineage-commitment points. *Epigenetics and Chromatin*, 11(1):1–18, 2018.

S. Higuchi, T. Hayashi, I. Hori, N. Shibata, H. Sakamoto, and K. Agata. Characterization and categorization of fluorescence activated cell sorted planarian stem cells by ultrastructural analysis. *Development Growth and Differentiation*, 49(7):571–581, 2007.

R. A. Hoskins, J. W. Carlson, K. H. Wan, S. Park, I. Mendez, S. E. Galle, B. W. Booth, B. D. Pfeiffer, R. A. George, R. Svirskas, M. Krzywinski, J. Schein, M. C. Accardo, E. Damia, G. Messina, M. Méndez-Lago, B. De Pablos, O. V. Demakova, E. N. Andreyeva, L. V. Boldyreva, M. Marra, A. B. Carvalho, P. Dimitri, A. Villasante, I. F. Zhimulev, G. M. Rubin, G. H. Karpen, and S. E. Celniker. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research*, 25(3):445–458, 2015.

A. Hubert, J. M. Henderson, M. W. Cowles, K. G. Ross, M. Hagen, C. Anderson, C. J. Szeterlak, and R. M. Zayas. A functional genomics screen identifies an Importin-α homolog as a regulator of stem cell function and tissue patterning during planarian regeneration. *BMC Genomics*, 16(1):1–18, 2015.

T. Ichimura, S. Watanabe, Y. Sakamoto, T. Aoto, N. Pujita, and M. Nakao. Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. *Journal of Biological Chemistry*, 280(14):13928–13935, 2005.

H. Iwano, M. Nakamura, and S. Tajima. Xenopus MBD3 plays a crucial role in an early stage of development. *Developmental Biology*, 268:416–428, 2004.

F. Jaber-Hijazi, P. J. K. P. Lo, Y. Mihaylova, J. M. Foster, J. S. Benner, B. Tejada Romero, C. Chen, S. Malla, J. Solana, A. Ruzov, and A. Aziz Aboobaker. Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Developmental Biology*, 384(1):141–153, 2013.

R. Jaenisch and A. Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, 33 Suppl(march): 245–254, 2003.

L. Jeffery and S. Nakielny. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *Journal of Biological Chemistry*, 279(47):49479–49487, 2004.

A. Jeltsch. Phylogeny of methylomes. *Science*, 328(May):837–839, 2010.

Y. Jin, K. Chen, A. De Paepe, E. Hellqvist, A. D. Krstic, L. Metang, C. Gustafsson, R. E. Davis, Y. M. Levy, R. Surapaneni, A. Wallblom, H. Nahi, R. Mansson, and Y. C. Lin. Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood*, 131(19):2138–2150, 2018.

K. Kaji, I. M. Caballero, R. MacLeod, J. Nichols, V. a. Wilson, and B. Hendrich. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature cell biology*, 8(3):285–92, 2006.

K. Kaji, J. Nichols, and B. Hendrich. Mbd3, a component of the NuRD co-repressor complex, is reqiured for development of pluripotent cells. *Development (Cambridge, England)*, 134(6):1123–1132, 2007.

J. Kang, J. Hu, R. Karra, A. L. Dickson, V. A. Tornini, G. Nachtrab, M. Gemberling, J. A. Goldman, B. L. Black, and K. D. Poss. Modulation of tissue repair by regeneration enhancer elements. *Nature*, 532(7598):201–206, 2016.

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

M. Kashiwagi, B. A. Morgan, and K. Georgopoulos. The chromatin remodeler Mi-2β is required for establishment of the basal epidermis and normal differentiation of its progeny. *Development*, 134(8):1571–1582, 2007.

E. L. Keisman and B. S. Baker. The Drosophila sex determination hierarchy modulates wingless and decapentaplegic signaling to deploy dachshund sex-specifically in the genital imaginal disc. *Development*, 128(9):1643–1656, 2001.

W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics*, 26(17): 2204–2207, 2010.

K. D. Kernohan, D. Vernimmen, G. B. Gloor, and N. G. Bérubé. Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Research*, 42(13):8356–8368, 2014.

M. Khacho, A. Clark, D. S. Svoboda, J. Azzi, J. G. MacLaurin, C. Meghaizel, H. Sesaki, D. C. Lagace, M. Germain, M. E. Harper, D. S. Park, and R. S. Slack. Mitochondrial Dynamics Impacts Stem Cell Identity and Fate Decisions by Regulating a Nuclear Transcriptional Program. *Cell Stem Cell*, 19(2):232–247, 2016.

A. Khan and A. Mathelier. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*, 18(1):1–8, 2017.

R. S. King and P. A. Newmark. In situ hybridization protocol for enhanced detection of gene expression in the planarian Schmidtea mediterranea. *BMC Developmental Biology*, 13(1), 2013.

S. L. Klemm, Z. Shipony, and W. J. Greenleaf. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20(4):207–220, 2019.

E. Knock, J. Pereira, P. D. Lombard, A. Dimond, D. Leaford, F. J. Livesey, and B. Hendrich. The methyl binding domain 3/nucleosome remodelling and deacetylase complex regulates neural cell fate determination and terminal differentiation in the cerebral cortex. *Neural Development*, 10(1):1–20, 2015.

N. Koenecke, J. Johnston, Q. He, S. Meier, and J. Zeitlinger. Drosophila poised enhancers are generated during tissue patterning with the help of repression. *Genome Research*, 27(1):64–74, 2017.

T. Kouzarides. Chromatin modifications and their function. *Cell*, pages 693–705, 2007.

N. Kunert, E. Wagner, M. Murawska, H. Klinker, E. Kremmer, and A. Brehm. dMec : a novel Mi-2 chromatin remodelling complex involved in transcriptional repression. *The EMBO Journal*, 28(5):533–544, 2009.

M. Kurusu, T. Nagao, U. Walldorf, S. Flister, W. J. Gehring, and K. Furukubo-Tokunaga. Genetic control of development of the mushroom bodies, the associative learning centers in the Drosophila brain, by the eyeless, twin of eyeless, and dachshund genes. *Proceedings of the National Academy of Sciences of the United States of America*, 97(5):2140–2144, 2000.

R. M. Labbé, M. Irimiac, K. W. Currie, A. Lina, S. J. Zhua, D. D. Brown, E. J. Rosse, V. Voisinc, G. D. Baderb, B. J. Blencowec, and B. J. Pearson. A Comparative Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians and Mammals Roselyne. *Stem Cells*, 49(18):1841–1850, 2009.

S. Laget, M. Joulie, F. Le Masson, N. Sasai, E. Christians, S. Pradhan, R. J. Roberts, and P. A. Defossez. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS ONE*, 5(8), 2010.

G. Lagger, D. O'Carroll, M. Rembold, H. Khier, J. Tischler, G. Weitzer, B. Schuettengruber, C. Hauser, R. Brunmeir, T. Jenuwein, and C. Seiser. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression Gerda. *The EMBO Journal*, 21(11):2672–2681, 2002.

A. G. Lai, N. Kosaka, P. Abnave, S. Sahu, and A. A. Aboobaker. The abrogation of condensin function provides independent evidence for defining the self-renewing population of pluripotent stem cells. *Developmental Biology*, 433(2):218–226, 2017.

A. Y. Lai and P. A. Wade. Cancer biology and NuRD: A multifaceted chromatin remodelling complex. *Nature Reviews Cancer*, 11(8):588–596, 2011.

S. W. Lapan and P. W. Reddien. Dlx and sp6-9 control optic cup regeneration in a prototypic eye. *PLoS Genetics*, 7(8), 2011.

D. Lara-Astiaso, A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Keren-Shaul, A. Mildner, D. Winter, S. Jung, N. Friedman, and I. Amit. Chromatin state dynamics during blood formation. *Science*, 345(6199):943–950, 2014.

G. Lauter, I. Söll, and G. Hauptmann. Two-color fluorescent in situ hybridization in the embryonic zebrafish brain using differential detection systems. *BMC Developmental Biology*, 11, 2011.

L. A. Lettice, S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14):1725–1735, 2003.

E. Li and Y. Zhang. DNA methylation in mammals. *Cold Spring Harbor Perspectives in Biology*, 6(5), 2014.

H. Li and R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009.

M. Li, G.-h. Liu, J. Carlos, and I. Belmonte. Navigating the epigenetic landscape of pluripotent stem cells. *Nature Reviews Molecular Cell Biology*, 13(August), 2012.

Q. L. Li, D. Y. Wang, L. G. Ju, J. Yao, C. Gao, P. J. Lei, L. Y. Li, X. L. Zhao, and M. Wu. The hyper-activation of transcriptional enhancers in breast cancer. *Clinical Epigenetics*, 11(1):1–17, 2019.

E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.

A. Y. Lin and B. J. Pearson. Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. *Development*, 141(6):1197–1208, 2014.

H. K. Long, S. L. Prescott, and J. Wysocka. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, 167(5):1170–1187, 2016.

A. Loyola, H. Tagami, T. Bonaldi, D. Roche, J. P. Quivy, A. Imhof, Y. Nakatani, S. Y. Dent, and G. Almouzni. The HP1α-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Reports*, 10(7):769–775, 2009.

X. Lu, B. S. Zhao, and C. He. TET family proteins: Oxidation activity, interacting molecules, and functions in diseases, 2015.

L. L. Luchsinger, M. J. De Almeida, D. J. Corrigan, M. Mumau, and H. W. Snoeck. Mitofusin 2 maintains haematopoietic stem cells with extensive lymphoid potential. *Nature*, 529(7587):528–531, 2016.

N. M. Luis, L. Morey, L. Di Croce, and S. A. Benitah. Polycomb in stem cells: PRC1 branches out. *Cell Stem Cell*, 11(1):16–21, 2012.

G. Mardon, N. M. Solomon, and G. M. Rubin. dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila. *Development*, 120(12): 3473–3486, 1994.

K. Marheineke and T. Krude. Nucleosome assembly activity and intracellular localization of human CAF-1 changes during the cell division cycle. *Journal of Biological Chemistry*, 273(24):15279–15286, 1998.

J. Marhold, A. Brehm, and K. Kramer. the The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2 complex via p55 and MI-2. *BMC molecular biology*, 5(1):20, 2004.

J. M. Martín-Durán, F. Monjo, and R. Romero. Morphological and molecular development of the eyes during embryogenesis of the freshwater planarian Schmidtea polychroa. *Development Genes and Evolution*, 222(1):45–54, 2012.

S. R. Martini, G. Roman, S. Meuser, G. Mardon, and R. L. Davis. The retinal determination gene, dachshund, is required for mushroom body cell differentiation. *Development*, 127(12):2663–2672, 2000.

C. Mayer, K. M. Schmitz, J. Li, I. Grummt, and R. Santoro. Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. *Molecular Cell*, 22(3):351–361, 2006.

P. Mcdonel, I. Costello, and B. Hendrich. Keeping things quiet : Roles of NuRD and Sin3 co-repressor complexes during mammalian development. *The International Journal of Biochemistry & Cell Biology*, 41:108–116, 2009.

R. Meehan, J. D. Lewis, and A. P. Bird. Characterization of MECP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research*, 20(19): 5085–5092, 1992.

R. Menafra and H. G. Stunnenberg. MBD2 and MBD3: Elusive functions and mechanisms. *Frontiers in Genetics*, 5(DEC):1–7, 2014.

B. Mifsud, F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, P. A. Ewels, B. Herman, S. Happe, A. Higgs, E. Leproust, G. A. Follows, P. Fraser, N. M. Luscombe, and C. S. Osborne. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, 47(6):598–606, 2015.

Y. Mihaylova, P. Abnave, D. Kao, S. Hughes, A. Lai, F. Jaber-hijazi, N. Kosaka, A. A. Aboobaker, and A. Aziz. Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Nature Communications*, 9(1):3633, 2018.

A. A. Mills. Throwing the cancer switch: Reciprocal roles of polycomb and trithorax proteins. *Nature Reviews Cancer*, 10(10):669–682, 2010.

A. M. Molinaro and B. J. Pearson. In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome biology*, 17(1):87, 2016.

J. E. Moore, H. E. Pratt, M. J. Purcaro, and Z. Weng. A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biology*, 21(1):1–16, 2020.

L. Morey, A. Santanach, and L. Di Croce. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. *Molecular and Cellular Biology*, 35(16): 2716–2728, 2015.

M. Morishita and E. Di Luccio. Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. *Biochemical and Biophysical Research Communications*, 412(2):214–219, 2011.

R. Murad, A. Macias-Muñoz, A. Wong, X. Ma, and A. Mortazavi. Integrative analysis of Hydra head regeneration reveals activation of distal enhancer-like elements. *bioRxiv*, 2019.

Y. J. Nam, K. Song, X. Luo, E. Daniel, K. Lambeth, K. West, J. A. Hill, J. M. Di Maio, L. A. Baker, R. Bassel-Duby, and E. N. Olson. Reprogramming of human fibroblasts toward a cardiac fate. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14):5588–5593, 2013.

X. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683):386–389, 1998.

G. Natoli. Maintaining cell identity through global control of genomic organization. *Immunity*, 33(1):12–24, 2010.

P. A. Newmark and A. S. Alvarado. Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. *Developmental Biology*, 220(2):142–153, 2000.

H.-h. Ng, P. Jeppesen, and A. Bird. Active Repression of Methylated Genes by the Chromosomal Protein MBD1. *Molecular and cellular biology*, 20(4):1394–1406, 2000.

S. Y.-M. Ng, T. Yoshida, J. Zhang, and K. Georgopoulos. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in Hematopoietic Stem Cells. *Immunity*, 23(1):1–7, 2009.

J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, and A. Smith. Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell*, 95(3):379–391, 1998.

J. Nitarska, J. G. Smith, W. T. Sherlock, M. M. Hillege, A. Nott, W. D. Barshop, A. A. Vashisht, J. A. Wohlschlegel, R. Mitter, and A. Riccio. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Reports*, 17(6):1683–1698, 2016.

A. Noveen, A. Daniel, and V. Hartenstein. Early development of the Drosophila mushroom body: The roles of eyeless and dachshund. *Development*, 127(16):3475–3488, 2000.

J. Ogas, S. Kaufmann, J. Henderson, and C. Somerville. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24):13839–13844, 1999.

I. Ohki, N. Shimotake, N. Fujita, J.-g. Jee, T. Ikegami, M. Nakao, and M. Shirakawa. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, 105:487–497, 2001.

P. Önal, D. Grün, C. Adamidi, A. Rybak, J. Solana, and N. Rajewsky. Gene expression of pluripotency determinants is conserved between mammalian and planarian. *The EMBO journal*, 31(December 2011):2755–2769, 2012.

L. P. O'Neill, M. D. VerMilyea, and B. M. Turner. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nature Genetics*, 38(7):835–841, 2006.

S. H. Orkin and K. Hochedlinger. Chromatin connections to pluripotency and cellular reprogramming. *Cell*, 52(6):566–584, 2016.

E. Pascual-Carreras, M. Marín-Barba, S. Castillo-Lara, P. Coronel-Córdoba, M. Magri, G. Wheeler, F. Abril, J. Gomez-Skarmeta, E. Saló, and T. Adell. Genomic analyses reveal FoxG as an upstream regulator of wnt1 required for posterior identity specification in planarians. *bioRxiv*, pages 1–46, 2020.

R. Peña-Hernández, R. Aprigliano, S. Frommel, K. Pietrzak, S. Steiger, M. Roganowicz, J. Bizzarro, and R. Santoro. BAZ2A association with H3K14ac is required for the dedifferentiation of prostate cancer cells into a cancer stem-like state. *bioRxiv*, 2020.

C. F. Pereira, I. R. Lemischka, and K. Moore. Reprogramming cell fates: Insights from combinatorial approaches. *Annals of the New York Academy of Sciences*, 1266(1): 7–17, 2012.

H. Pimentel, N. L. Bray, S. Puente, P. Melsted, and L. Pachter. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, 14(7):687–690, 2017.

M. Plass, J. Solana, F. A. Wolf, S. Ayoub, A. Misios, P. Glažar, B. Obermayer, F. J. Theis, C. Kocks, and N. Rajewsky. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science*, 1723(April):1–17, 2018.

H. A. Pliner, J. S. Packer, J. L. McFaline-Figueroa, D. A. Cusanovich, R. M. Daza, D. Aghamirzaie, S. Srivatsan, X. Qiu, D. Jackson, A. Minkina, A. C. Adey, F. J. Steemers, J. Shendure, and C. Trapnell. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Molecular Cell*, 71(5): 858–871.e8, 2018.

M. M. Pradeepa. Causal role of histone acetylations in enhancer function. *Transcription*, 8(1):40–47, 2017.

S. Pundhir, F. O. Bagger, F. B. Lauridsen, N. Rapin, and B. T. Porse. Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic Acids Research*, 44(9):4037–4051, 2016.

Y. Qiao, R. Wang, X. Yang, K. Tang, and N. Jing. Dual roles of histone H3 lysine 9 acetylation in human embryonic stem cell pluripotency and neural differentiation. *Journal of Biological Chemistry*, 290(16):9949, 2015.

A. R. Quinlan and I. M. Hall. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

A. Rada-Iglesias, R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, and J. Wysocka. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333):279–285, 2011.

G. Raddatz, P. M. Guzzardo, N. Olova, M. Rosado, and M. Rampp. Dnmt2-dependent methylomes lack de fi ned DNA methylation patterns. *Proceedings of the National Academy of Sciences*, 110(21):2–6, 2013.

F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, and T. Manke. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*, 44(W1):W160–W165, 2016.

S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.

C. Rauskolb. The establishment of segmentation in the Drosophila leg. *Development*, 128(22):4511–4521, 2001.

P. W. Reddien and A. S. Alvarado. Fundamentals of Planarian Regeneration. *Annual Review of Cell and Developmental Biology*, 20(1):725–757, 2004.

P. W. Reddien, A. L. Bermange, K. J. Murfitt, J. R. Jennings, and A. Sánchez Alvarado. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental Cell*, 8(5):635–649, 2005a.

P. W. Reddien, N. J. Oviedo, J. R. Jennings, P. W. Reddien, J. C. Jenkin, and A. S. Alvarado. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science*, 310(November):1327–1331, 2005b.

J. P. Reddington, D. A. Garfield, O. M. Sigalova, A. Karabacak Calviello, R. Marco-Ferreres, C. Girardot, R. R. Viales, J. F. Degner, U. Ohler, and E. E. Fdisableurlong. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. *Developmental Cell*, 55(5):648–664.e9, 2020.

A. Regev, M. J. Lamb, and E. Jablonka. The role of DNA methylation in invertebrates: developmental regulation or genome defense ? *Mol. Biol. Evol.*, 15(7)(January): 880–891, 1998.

A. M. Resch, D. Palakodeti, Y. C. Lu, M. Horowitz, and B. R. Graveley. Transcriptome analysis reveals strain-specific and conserved stemness genes in schmidtea mediterranea. *PLoS ONE*, 7(4), 2012.

N. Reynolds, P. Latos, A. Hynes-Allen, R. Loos, D. Leaford, A. O'Shaughnessy, O. Mosaku, J. Signolet, P. Brennecke, T. Kalkan, I. Costello, P. Humphreys, W. Mansfield, K. Nakagawa, J. Strouboulis, A. Behrens, P. Bertone, and B. Hendrich. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell*, 10(5):583–594, 2012a.

N. Reynolds, M. Salmon-Divon, H. Dvinge, A. Hynes-Allen, G. Balasooriya, D. Leaford, A. Behrens, P. Bertone, and B. Hendrich. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO Journal*, 31(3):593–605, 2012b.

H. Richly, L. Aloia, and L. Di Croce. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death and Disease*, 2(9):e204–7, 2011.

J. C. Rink. Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, 223:67–84, 2013.

J. C. Rink, H. T.-K. Vu, and A. S. Alvarado. The maintenance and regeneration of the planarian excretory system are regulated by EGFR signaling. *Development*, 138(17): 3769–3780, 2011.

S. M. Robb and A. S. Alvarado. Histone Modifications and Regeneration in the Planarian Schmidtea mediterranea. *Current Topics in Developmental Biology*, 108:71–93, 2014.

S. M. Robb, K. Gotting, E. Ross, and A. S. Alvarado. SmedGD 2.0: The Schmidtea mediterranea genome database. *Genesis*, 4(1):139–148, 2015.

K. Roder, M.-s. Hung, T.-l. Lee, T.-y. Lin, H. Xiao, K.-i. Isobe, J.-l. Juang, and C. J. Shen. Transcriptional repression by drosophila methyl-CpG-binding proteins. *Molecular Cell*, 20(19):7401–7409, 2000.

T. C. Roloff, H. H. Ropers, and U. A. Nuber. Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics*, 4:1–9, 2003.

B. T. Romero, D. J. Evans, and A. A. Aboobaker. FACS Analysis of the Planarian Stem Cell Compartment as a Tool to Understand Regenerative Mechanisms. *Progenitor Cells: Methods and Protocols*, 916:167–179, 2012.

K. G. Ross, K. C. Omuro, M. R. Taylor, R. K. Munday, A. Hubert, R. S. King, and R. M. Zayas. Novel monoclonal antibodies to study tissue regeneration in planarians. *BMC Developmental Biology*, 15(1), 2015.

L. Rouhana, N. Shibata, O. Nishimura, and K. Agata. Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Developmental Biology*, 341(2):429–443, 2010.

A. Rozanski, H. K. Moon, H. Brandl, J. M. Martín-Durán, M. A. Grohme, K. Hüttner, K. Bartscherer, I. Henry, and J. C. Rink. PlanMine 3.0 - improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research*, 47(D1): D812–D820, 2019.

V. E. A. Russo, R. A. Martienssen, and A. D. Riggs. *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press, Woodbury, 1996.

P. A. Ruzycki, X. Zhang, and S. Chen. CRX directs photoreceptor differentiation by accelerating chromatin remodeling at specific target sites. *Epigenetics and Chromatin*, 11(1):1–16, 2018.

M. Saito and F. Ishikawa. The mCpG-binding Domain of Human MBD3 Does Not Bind to mCpG but Interacts with NuRD / Mi2 Components HDAC1 and MTA2 *. *The Journal of Biological Chemistry*, 277(38):35434–35439, 2002.

R. Santoro and I. Grummt. Epigenetic mechanism of rRNA gene silencing : temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. *Molecular and cellular biology*, 25(7):2539–2546, 2005.

F. Scebba, G. Bernacchia, M. De Bastiani, M. Evangelista, R. M. Cantoni, R. Cella, M. T. Locci, and L. Pitto. Arabidopsis MBD proteins show different binding specificities and nuclear localization. *Plant Molecular Biology*, 53(5):715–731, 2003.

J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 2012.

M. L. Scimone, J. Meisel, and P. W. Reddien. The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian Schmidtea mediterranea. *Development and Setem cells*, 1241:1231–1241, 2010.

T. M. Scott, H. Guo, E. E. Eichler, J. A. Rosenfeld, K. Pang, Z. Liu, S. Lalani, W. Bi, Y. Yang, C. A. Bacino, H. Streff, A. M. Lewis, M. K. Koenig, I. Thiffault, A. Bellomo, D. B. Everman, J. R. Jones, R. E. Stevenson, R. Bernier, C. Gilissen, R. Pfundt, S. M. Hiatt, G. M. Cooper, J. L. Holder, and D. A. Scott. BAZ2B haploinsufficiency as a cause of developmental delay, intellectual disability, and autism spectrum disorder. *Human Mutation*, 41(5):921–925, 2020.

N. Sen, B. Gui, and R. Kumar. Physiological functions of MTA family of proteins Nirmalya. *Cancer Metastasis Rev*, 33(4):869–877, 2014.

B. J. Seo, S. H. Yoon, and J. T. Do. Mitochondrial Dynamics in Stem Cells and Differentiation. *International Journal of Molecular Sciences*, 19, 2018.

S. Shao, H. Cao, Z. Wang, D. Zhou, C. Wu, S. Wang, D. Xia, and D. Zhang. CHD4/NuRD complex regulates complement gene expression and correlates with CD8 T cell infiltration in human hepatocellular carcinoma. *Clinical Epigenetics*, 12 (1):1–13, 2020.

T. Shimbo, Y. Du, S. A. Grimm, A. Dhasarathy, D. Mav, R. R. Shah, H. Shi, and P. A. Wade. MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genetics*, 9(12), 2013.

D. Shlyueva, G. Stampfel, and A. Stark. Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4):272–286, 2014.

T. Simsek, F. Kocabas, J. Zheng, R. J. Deberardinis, I. Ahmed, E. N. Olson, J. W. Schneider, C. C. Zhang, and A. Hesham. The Distinct Metabolic Profile of Hematopoietic Stem Cells Reflects Their Location in a Hypoxic Niche. *Cell Stem Cell*, 7(3):380–390, 2014.

J. Solana. Closing the circle of germline and stem cells: The Primordial Stem Cell hypothesis. *EvoDevo*, 4(1):1, 2013.

J. Solana, D. Kao, Y. Mihaylova, F. Jaber-Hijazi, S. Malla, R. Wilson, and A. Aboobaker. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome biology*, 13(3):R19, 2012.

J. Solana, M. Irimia, S. Ayoub, M. R. Orejuela, V. Zywitza, M. Jens, J. Tapial, D. Ray, Q. Morris, T. R. Hughes, B. J. Blencowe, and N. Rajewsky. Conserved functional antagonism of CELF and MBNL proteins controls stem cell-specific alternative splicing in planarians. *eLife*, 5(AUGUST):1–29, 2016.

M. Y. Son, H. Choi, Y. M. Han, and Y. S. Cho. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem Cells*, 31(11):2374–2387, 2013.

M. Y. Son, Y. Kwon, M. Y. Son, B. Seol, H. S. Choi, S. W. Ryu, C. Choi, and Y. S. Cho. Mitofusins deficiency elicits mitochondrial metabolic reprogramming to pluripotency. *Cell Death and Differentiation*, 22(12):1957–1969, 2015.

K. Song, Y. J. Nam, X. Luo, X. Qi, W. Tan, G. N. Huang, A. Acharya, C. L. Smith, M. D. Tallquist, E. G. Neilson, J. A. Hill, R. Bassel-Duby, and E. N. Olson. Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature*, 485(7400):599–604, 2012.

X. Song, F. Huang, J. Liu, C. Li, S. Gao, W. Wu, M. Zhai, X. Yu, W. Xiong, J. Xie, and B. Li. Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in Tribolium castaneum. *DNA Research*, 24(5):445–458, 2017.

X. Song, Y. Zhang, Q. Zhong, K. Zhan, J. Bi, J. Tang, J. Xie, and B. Li. Identification and functional characterization of methyl-CpG binding domain protein from Tribolium castaneum. *Genomics*, 112(3):2223–2232, 2020.

S. Spicuglia and L. Vanhille. Chromatin signatures of active enhancers. *Nucleus*, 3(2): 126–131, 2012.

B. Spurlock, J. M. Tullet, J. L. Hartman, and K. Mitra. Interplay of mitochondrial fission-fusion with cell cycle regulation: Possible impacts on stem cell and organismal aging. *Experimental Gerontology*, 135(November 2019):110919, 2020.

R. Stark and G. Brown. DiffBind: differential binding analysis of ChIP-Seq peak datatle. Technical report, 2016.

P. A. Steffen and L. Ringrose. What are memories made of? How polycomb and trithorax proteins mediate epigenetic memory. *Nature Reviews Molecular Cell Biology*, 15(5):340–356, 2014.

D. Strepetkaitė, G. Alzbutas, E. Astromskas, A. Lagunavičius, R. Sabaliauskaitė, K. Arbačiauskas, and J. Lazutka. Analysis of DNA methylation and hydroxymethylation in the genome of crustacean Daphnia pulex. *Genes*, 7(1):1–14, 2015.

R. Strohner, A. Nemeth, P. Jansa, U. Hofmann-rohrer, R. Santoro, G. La, and I. Grummt. NoRC-a novel member of mammalian ISWI-containing chromatin remodeling machines. *The EMBO journal*, 20(17), 2001.

T. Suda, K. Takubo, and G. L. Semenza. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell*, 9(4):298–310, 2011.

L. S. Swapna, A. M. Molinaro, N. Lindsay-Mosher, B. J. Pearson, and J. Parkinson. 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):1–22, 2018.

M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324 (5929):930–935, 2009.

K. Takahashi and S. Yamanaka. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4):663–676, 2006.

Z. Tang, O. J. Luo, X. Li, M. Zheng, J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, B. Ruszczycki, P. Michalski, E. Piecuch, P. Wang, S. Z. Tian, M. Penrad-mobayed, L. M. Sachs, X. Ruan, C.-l. Wei, E. T. Liu, G. M. Wilczynski, and D. Plewczynski. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, 163(7):1611–1627, 2015.

S. Thomas, X. Y. Li, P. J. Sabo, R. Sandstrom, R. E. Thurman, T. K. Canfield, E. Giste, W. Fisher, A. Hammonds, S. E. Celniker, M. D. Biggin, and J. A. Stamatoyannopoulos. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biology*, 12(5), 2011.

J. D. Thompson, J. Ou, N. Lee, K. Shin, V. Cigliola, L. Song, G. E. Crawford, J. Kang, and K. D. Poss. Identification and requirements of enhancers that direct gene expression during zebrafish fin regeneration. *Development (Cambridge)*, 147(14), 2020.

K. C. Tu, L. C. Cheng, H. T. Vu, J. J. Lange, S. A. McKinney, C. W. Seidel, and A. Sánchez Alvarado. Egr-5 is a post-mitotic regulator of planarian epidermal differentiation. *eLife*, 4(OCTOBER2015):1–27, 2015.

M. Van Den Hurk, G. Kenis, C. Bardy, D. L. Van Den Hove, F. H. Gage, H. W. Steinbusch, and B. P. Rutten. Transcriptional and epigenetic mechanisms of cellular reprogramming to induced pluripotency. *Epigenomics*, 8(8):1131–1149, 2016.

J. C. van Wolfswinkel and R. F. Ketting. The role of small non-coding RNAs in genome stability and chromatin organization. *Journal of Cell Science*, 123(11):1825–1839, 2010.

J. C. van Wolfswinkel, D. E. Wagner, and P. W. Reddien. Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. *Cell Stem Cell*, pages 326–339, 2014.

C. Vásquez-Doorman and C. P. Petersen. The NuRD complex component p66 suppresses photoreceptor neuron regeneration in planarians. *Regeneration (Oxford, England)*, 3 (3):168–78, 2016.

A. Visel, M. J. Blow, Z. Li, T. Zhang, J. A. Akiyama, I. Plajzer-frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E. M. Rubin, and L. A. Pennacchio. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231):854–858, 2009.

E. Vizcaya-Molina, C. C. Klein, F. Serras, R. K. Mishra, R. Guigó, and M. Corominas. Damage-responsive elements in Drosophila regeneration. *Genome Research*, 28(12): 1841–1851, 2018.

P. Voigt, W. W. Tee, and D. Reinberg. A double take on bivalent promoters. *Genes and Development*, 27(12):1318–1338, 2013.

T. von Zelewsky, F. Palladino, K. Brunschwig, H. Tobler, A. Hajnal, and F. Müller. The C. elegans Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. *Development*, 127(24):5277–5284, 2000.

C. H. Waddington. The epigenotype. *Endeavour*, 1(0):18–20, 1942.

P. A. Wade, A. Gegonne, P. L. Jones, E. Ballestar, F. Aubry, and A. P. Wolffe. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nature Genetics*, 23(SEPTEMBER):1–5, 1999.

D. E. Wagner, I. E. Wang, and P. W. Reddien. Clonogenic neoblasts are pluripotent adult Stem cells that underlie planarian regeneration. *Science*, 332(May):811–817, 2011.

J. Wang, Y. Zhao, X. Zhou, S. W. Hiebert, Q. Liu, and Y. Shyr. Nascent RNA sequencing analysis provides insights into enhancer-mediated gene regulation. *BMC Genomics*, 19(1):1–18, 2018.

Y. Wang, H. Zhang, Y. Chen, Y. Sun, F. Yang, W. Yu, J. Liang, L. Sun, X. Yang, L. Shi, R. Li, Y. Li, Y. Zhang, Q. Li, X. Yi, and Y. Shang. LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell*, 138(4): 660–672, 2009.

Y. Wang, J. M. Stary, J. E. Wilhelm, and P. A. Newmark. A functional genomic screen in planarians identifies novel regulators of germ cell development. *Genes and Development*, 24(18):2081–2092, 2010.

K. Wasik, J. Gurtowski, X. Zhou, O. M. Ramos, M. J. Delás, G. Battistoni, O. El Demerdash, I. Falciatori, D. B. Vizoso, A. D. Smith, P. Ladurner, L. Schärer, W. R. McCombie, G. J. Hannon, and M. Schatz. Genome and transcriptome of the regeneration-competent flatworm, Macrostomum lignano. *Proceedings of the National Academy of Sciences*, 112(40):12462–12467, 2015.

C. M. Weber and S. Henikoff. Histone variants: Dynamic punctuation in transcription. *Genes and Development*, 28(7):672–682, 2014.

D. Wenemoser and P. W. Reddien. Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Developmental Biology*, 344(2):979–991, 2010.

J. A. West, A. Cook, B. H. Alver, M. Stadtfeld, A. M. Deaton, K. Hochedlinger, P. J. Park, M. Y. Tolstorukov, and R. E. Kingston. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nature Communications*, 5:1–12, 2014.

W. A. Whyte, S. Bilodeau, D. A. Orlando, H. A. Hoke, G. M. Frampton, C. T. Foster, S. M. Cowley, and R. A. Young. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*, 482(7384):221–225, 2014.

C. J. Williams, T. Naito, P. G.-d. Arco, J. R. Seavitt, S. M. Cashman, B. D. Souza, X. Qi, P. Keables, U. H. V. Andrian, and K. Georgopoulos. The Chromatin Remodeler Mi-2β Is Required for CD4 Expression and T Cell Development. *Immunity*, 20:719–733, 2004.

M. J. Wu, Y. S. Chen, M. R. Kim, C. C. Chang, S. Gampala, Y. Zhang, Y. Wang, C. Y. Chang, J. Y. Yang, and C. J. Chang. Epithelial-Mesenchymal Transition Directs Stem Cell Polarity via Regulation of Mitofusin. *Cell Metabolism*, 29(4):993–1002.e6, 2019.

Y. Xue, J. Wong, G. T. Moreno, M. K. Young, J. Côté, and W. Wang. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular Cell*, 2(6):851–861, 1998.

T. Yamada, Y. Yang, M. Hemberg, T. Yoshida, H. Y. Cho, J. P. Murphy, D. Fioravante, W. G. Regehr, S. P. Gygi, K. Georgopoulos, and A. Bonni. Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron*, 83(1):122–134, 2014.

S. Yamanaka and H. M. Blau. Nuclear reprogramming to a pluripotent state by three approaches. *Nature*, 465(7299):704–712, 2010.

F. Yan, D. R. Powell, D. J. Curtis, and N. C. Wong. From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, 21(1):1–16, 2020.

K. H. Yang and J. Kang. Tissue Regeneration Enhancer Elements: A Way to Unlock Endogenous Healing Power. *Developmental Dynamics*, 248(1):34–42, 2019.

O. Yildirim, R. Li, J.-h. H. Hung, P. B. Chen, X. Dong, L. S. Ee, Z. Weng, O. J. Rando, and T. G. Fazzio. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, 147(7):1498–1510, 2011.

T. Yoshida, I. Hazan, J. Zhang, S. Y. Ng, T. Naito, H. J. Snippert, E. J. Heller, X. Qi, L. N. Lawton, C. J. Williams, and K. Georgopoulos. The role of the chromatin remodeler Mi-2β in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes and Development*, 22(9):1174–1189, 2008.

M. D. Young, T. A. Willson, M. J. Wakefield, E. Trounson, D. J. Hilton, M. E. Blewitt, A. Oshlack, and I. J. Majewski. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Research*, 39(17):7415–7427, 2011.

E. Zacharioudaki, J. F. Sanjuan, and S. Bray. Mi-2/NuRD complex protects stem cell progeny from mitogenic Notch signaling. *eLife*, pages 1–24, 2018.

A. Zemach and D. Zilberman. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current Biology*, 20(17):R780–R785, 2010.

A. Zemach, I. E. Mcdaniel, P. Silva, and D. Zilberman. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 11928(MAY):916–920, 2010.

A. Zeng, H. Li, L. Guo, X. Gao, S. McKinney, Y. Wang, Z. Yu, J. Park, C. Semerad, E. Ross, L. C. Cheng, E. Davies, K. Lei, W. Wang, A. Perera, K. Hall, A. Peak, A. Box, and A. Sánchez Alvarado. Prospectively Isolated Tetraspanin+Neoblasts Are Adult Pluripotent Stem Cells Underlying Planaria Regeneration. *Cell*, pages 1593–1608, 2018.

Y. Zhang, H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes and Development*, 13(15):1924–1935, 1999.

Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9):R137, 2008.

A. Zhao, H. Qin, and X. Fu. What determines the regenerative capacity in animals? *BioScience*, 66(9):735–746, 2016.

X. Zhao, T. Ueba, B. R. Christie, B. Barkho, M. J. Mcconnell, K. Nakashima, E. S. Lein, B. D. Eadie, A. R. Willhoite, A. R. Muotri, R. G. Summers, J. Chun, K.-f. Lee, and F. H. Gage. Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proceedings of the National Academy of Sciences*, 100(11):6777–6782, 2003.

Y. Zhao, D. Zheng, and A. Cvekl. Profiling of chromatin accessibility and identification of general cis-regulatory mechanisms that control two ocular lens differentiation pathways. *Epigenetics and Chromatin*, 12(1):1–23, 2019.

X. M. Zheng, V. Moncollin, J. M. Egly, and P. Chambon. A general transcription factor forms a stable complex with RNA polymerase B (II). *Cell*, 50(3):361–368, 1987.

W. Zhou, M. Choi, D. Margineantu, L. Margaretha, J. Hesson, C. Cavanaugh, C. A. Blau, M. S. Horwitz, D. Hockenbery, C. Ware, and H. Ruohola-Baker. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO Journal*, 31(9):2103–2116, 2012.

D. Zhu, J. Fang, Y. Li, and J. Zhang. Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. *PLoS ONE*, 4(11), 2009.

L. J. Zhu, C. Gazin, N. D. Lawson, H. Pagès, S. M. Lin, D. S. Lapointe, and M. R. Green. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11, 2010.

S. J. Zhu and B. J. Pearson. The Retinoblastoma pathway regulates stem cell proliferation in freshwater planarians. *Developmental Biology*, 373(2):442–452, 2013.

S. J. Zhu, S. E. Hallows, K. W. Currie, C. Xu, and B. J. Pearson. A mex3 homolog is required for differentiation during planarian stem cell lineage development. *eLife*, 4 (JUNE 2015):1–23, 2015.