

Epigenetic control of planarian stem cell potency limits stem activity and accurately defines differentiation programs

Yuliana Valerieva Mihaylova, MRes.

Thesis Submitted for the Degree of Doctor of Philosophy

December 2015

ABSTRACT

Planarian flatworms are gaining popularity in regenerative medicine research due to the fact that they have unparalleled regeneration capacity. Their tissue recovery abilities are dependent on a pool of adult stem cells (neoblasts). Studies in the recent years have shown that epigenetic mechanisms have an important role in neoblasts' self-renewal and differentiation properties. This thesis focuses on the study of trithorax-related genes and their function in neoblast regulation. Despite the fact that mammalian trithorax-related genes *Mll3* and *Mll4* are among the most frequently mutated genes in cancer, trithorax-related genes are the least well-studies members of the trithorax gene group (TrxG) of histone modifiers. The current study traced the evolutionary history of trithorax-related genes and concluded that they have undergone a number of independent gene fission events across phyla. In planarians, there are three partial orthologue of the mammalian *Mll3* and *Mll4* genes – *Smed-LPT* (corresponding to the N-terminus of *Mll3/4*), *Smed-trr-1* and *Smed-trr-2* (both corresponding to the C-terminus of *Mll3/4*). The three planarian trithorax-related genes are expressed in stem cells and control neoblast differentiation down certain lineages (brain, gut, eyes, pharynx, epidermis). Down-regulation of *Smed-LPT* results in hyperproliferation of stem cells, leading to tumour-like outgrowth formation. It was shown that trithorax-related genes' function in stem cell regulation correlates with histone modification changes, specifically alterations in H3K4me1, H3K4me3 and H3K27me3. Future studies will focus on examining this correlation further via Next-Generation sequencing techniques.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Aziz Aboobaker, who has been extremely supportive and helpful throughout my PhD. His door was never closed when I had a question or needed advice. Equally important in this experience were also my parents who have always been one call away for whatever I needed. I am also extremely grateful for all the amazing colleagues I have had throughout the years in the Aboobaker lab. I am particularly thankful to Dr. Farah Jaber-Hijazi for being so approachable and friendly when I needed it the most. She was also the person who started the *Smed-LPT* project. I would also like to thank my colleague and partner Dr. Damian Kao for all his emotional and scientific support. Dr. Nobuyoshi Kosaka and Dr. Samantha Hughes also deserve a special mention for all the discussions and advice they have provided me with. They have also contributed to experiments presented in this thesis. An important thank you goes to our lab manager Dr. Natasha Poutchkina-Stancheva for being the organisational glue of the lab. Outside of the lab, I would like to thank Dr. Nikola Vlahov for 'converting' me to biology and for always being extremely helpful with scientific issues, as well being a great friend. Dr. Emilia Dimitrova has also been helpful in this experience. I would also like to express my gratitude to all my friends, including Raya, Andrea, Preslav, Molly, Teodora, Ilina. Lastly, my turtles, Rosalind and Ada, deserve a big thank you.

LIST OF PUBLICATIONS

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

Mihaylova, Y., and Aboobaker, A.A. (2013). What is it about “eye of newt”? Genome Biol *14*, 106.

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

Solana, J., Gamberi, C., Mihaylova, Y., Grosswendt, S., Chen, C., Lasko, P., Rajewsky, N., and Aboobaker, A.A. (2013). The CCR4-NOT Complex Mediates Deadenylation and Degradation of Stem Cell mRNAs and Promotes Planarian Stem Cell Differentiation. PLoS Genet *9*, e1004003.

Tejada-Romero, B., Carter, J.-M., Mihaylova, Y., Neumann, B., and Aboobaker, A.A. (2015). JNK signalling is necessary for a Wnt- and stem cell-dependent regeneration programme. Development *142*, 2413–2424.

TABLE OF CONTENTS

Chapter 1. General Introduction 1

1.1. The planarian appeal 1

1.2. Anatomy 2

1.3. Regeneration 5

1.4. Neoblast features and the chromatoid body 9

1.5. Methods for neoblast ablation 12

1.5.1. Gamma-radiation 12

1.5.2. *Smed-H2B*(RNAi) 14

1.5.3. Fluorescence activated cell sorting 15

1.6. Neoblast heterogeneity 16

1.6.1. Variable sensitivity to irradiation 16

1.6.2. The problem of G2-arrested stem cells 18

1.6.3. Neoblast potency 19

1.6.3.1. Neoblasts, germline and germline precursors 19

1.6.3.2. cNeoblasts 20

1.6.3.3. Committed neoblasts 22

1.7. Epigenetics and flatworms 27

1.7.1. DNA methylation 27

1.7.2. The NuRD complex 30

1.7.3. Chromatin structure 32

1.7.4. Histone modifications 35

1.7.4.1. Lysine methylation 37

1.7.4.1.1. Bivalent domains 38

1.7.4.1.2. Polycomb proteins 40

1.7.4.1.3. Trithorax and trhithorax-related proteins 41

1.8. The aim of this thesis 46

Chapter 2. Identification and sequence analysis of *Smed-LPT, Smed-trr-1* and *Smed-trr-2* 47

2.1. Abstract 47

2.2. Introduction 48

2.2.1. The split *Mll3/4* homologue 48

2.2.1.1. Fusion versus fission 52

2.2.2. The PHD-finger 55

2.2.3. The SET domain 59

2.2.4. Aim of chapter 61

2.3. Materials and Methods 62

2.3.1. Reagents 62

2.3.2. *Schmidtea mediterranea* culturing 62

2.3.3. Primers for Rapid Amplification of cDNA Fragments (RACE) and establishing the full-length sequence of *Smed-LPT* 63

2.3.4. Databases used for gene homologue identification 63

2.3.5. Primer design 63

2.3.6. Phusion PCR 64

2.3.7. Gel extraction 64

2.3.8. RNA extraction 64

2.3.9. cDNA preparation 64

2.3.9.1. For 3’ RACE and conventional PCR 64

2.3.9.2. For 5’ RACE 65

2.3.10. 3’ RACE 65

2.3.11. 5’ RACE 66

2.3.12. Protein alignments and phylogenetic trees 66

2.4. Results and Discussion 67

2.4.1. Identifying *Smed-LPT*, Smed-trr-1 and *Smed-trr-2* 67

2.4.2. Finding the full-length sequence of *Smed-LPT* 73

2.4.3. Origin of the split between *LPT* and *trr* 76

2.4.4. Further investigations of the *Mll3/4* split and origin of the duplication of *trr* in planarians 82

2.4.5. Domain conservation in SMED-LPT, SMED-TRR-1 and SMED-TRR-2 86
2.5. Conclusion 92

Chapter 3. Trithorax-related genes are expressed in planarian neoblasts and neoblast progeny 93
3.1. Abstract 93
3.2. Introduction 93
3.2.1. MLL3/4/Trr transcript and protein expression during development, adulthood and regeneration 93
3.2.1.1 Zebrafish 94
3.2.1.2. Mammals 95
3.2.1.3. *Drosophila* 96
3.2.1.4. *Smed-trr-1* and *Smed-trr-2* transcript expression in planarians 98
3.2.2. Aim of chapter 99
3.3. Materials and Methods 100
3.3.1. Reagents 100
3.3.2. Primers used 100
3.3.3. Gene clones used 100
3.3.4. Primer design for RNA probes and dsRNA 100
3.3.5. PCR 101
3.3.6. Cloning of genes 101
3.3.7. RNA probe synthesis 101
3.3.7.1 RNA probes from PCR 102
3.3.7.2. RNA probe from plasmid 102
3.3.8. dsRNA synthesis 103
3.3.9. RNA interference 103
3.3.10. Gamma radiation 104
3.3.11. *In situ* hybridisation protocols 104
3.3.11.1. Carnoy-based *in situ* hybridisation 104

3.3.11.2. King and Newmark (2013) *in situ* hybridisation protocol 105
3.3.12. Preparation of stock fluor-tyramide solutions 106
3.3.13. Imaging 106
3.4. Results and Discussion 106
3.4.1. Trithorax-related genes are expressed in CNS, gut and stem cell progeny 106
3.4.2. Stem cells express trithorax-related genes 111
3.4.3. Trithorax-related genes are expressed in the same cells 116
3.5. Conclusion 118

Chapter 4. Knockdown of trithorax-related genes perturbs regeneration and homeostasis and induces defects in stem cell differentiation 119
4.1. Abstract 119
4.2. Introduction 119
4.2.1. LPT-Trr function in *Drosophila* 119
4.2.2. SET-16 in *C. elegans* 123
4.2.3. MLL3/4 function in mammals 124
4.2.4. *Smed-trr-1* and *Smed-trr-2* in planarians 127
4.2.5. Nuclear receptors in planarians 128
4.2.6. Aim of chapter 130
4.3. Materials and Methods 130
4.3.1. Primers for *in situ* hybridisation markers 130
4.3.2. Gene clones used 130
4.3.3. Antibodies 131
4.3.4. Immunohistochemistry 131
4.3.5. Cell counting 131
4.3.6. RNAi of trithorax-related genes 132
4.3.7. Ectopic serotonin treatment 132
4.3.8. Phenotype scoring and brightfield imaging 132

4.4. Results and Discussion 133
4.4.1. Knockdown of trithorax-related genes 133
4.4.1.1. Knockdown of trithorax-related genes is effective 133
4.4.1.2. Tissue regeneration and homeostasis are impaired following knockdown of trithorax-related genes 135
4.4.2. Defects following trithorax-related genes knockdown are not due to failure in neoblast maintenance or an overall decrease in early neoblast progeny production 142
4.4.3. Trithorax-related genes knockdown impairs anterior structures regeneration 145
4.4.4. Motility defects in trithorax-related genes knockdown animals are due to serotonin deficiency 159
4.4.5. Pharynx, gut and epidermis, but not protonephridia, are affected following knockdown of trithorax-related genes 165
4.5. Conclusion 174

Chapter 5. Cancerous features of *Smed-LPT* knockdown 176
5.1. Abstract 176
5.2. Introduction 176
5.2.1. Models of cancer 179
5.2.1.1. The problem of cancer cell heterogeneity 179
5.2.1.2. Clonal evolution (stochastic) model 180
5.2.1.3. Hierarchical/Cancer Stem Cell (CSC) model 181
5.2.1.4. Combined view 184
5.2.2. Cancer and epigenetics 186
5.2.2.1. DNA methylation 187
5.2.2.2. Histone modifications 188
5.2.3. The MLL family in cancer 190
5.2.3.1. MLL1 190
5.2.3.2. MLL3 191
5.2.4. Cancer in planarians 194

5.2.5. Aim of chapter 198
5.3. Materials and Methods 198
5.3.1. Reagents 198
5.3.2. Staining with anti-H3P antibody 199
5.3.2. Mild *Smed-LPT* knockdown 199
5.3.3. Anaesthetising and partially irradiating planarians 199
5.4. Results and Discussion 200
5.4.1. The appearance of outgrowths in *Smed-LPT*(RNAi) animals is preceded and accompanied by stem increased stem cell proliferation 200
5.4.2. Outgrowths in *Smed-LPT*(RNAi) animals contain stem cells 204
5.4.3. *Smed-LPT*(RNAi) leads to nuclear deformation 208
5.4.4. *Smed-LPT*(RNAi) animals exhibit epidermal abnormalities 210
5.4.5. *Smed-LPT* controls stem cell migration 214
5.5. Conclusion 220

Chapter 6. Cellular dynamics and histone modification changes following knockdown of trithorax-related genes 223
6.1. Abstract 223
6.2. Introduction 224
6.2.1. Histone modification inheritance during replication 224
6.2.1.1. Models of histone modification inheritance 224
6.2.1.2. Histone variants inherited independently of the cell cycle 228
6.2.1.3. Histone lysine acetylation 229
6.2.1.4. Histone lysine methylation 229
6.2.1.5. Histone modifications and different stages of the cell cycle 232
6.2.2. Chromatin states 233
6.2.3. LPT/Trr/MLL3/4 and histone modifications 237
6.2.3.1. Trimethylation of H3K4 on promoters 237

6.2.3.2. Monomethylation of H3K4 on enhancers 238
6.2.3.3. Monomethylation of H3K4 on promoters 240
6.2.3. Aim of chapter 244
6.3. Materials and Methods 244
6.3.1. Reagents 244
6.3.2. RNAi of *Smed-set1* and *Smed-utx* 244
6.3.3. Double RNAi 244
6.3.4. Fluorescence Activated Cell Sorting (FACS) 245
6.3.5. Protein extract from whole worms 245
6.3.6. Whole protein extract from x1 sorted cells 246
6.3.7. Western blot 246
6.4. Results and Discussion 248
6.4.1. Trithorax-related genes control H3K4me1 and H3K27me3 levels during homeostatic conditions 250
6.4.2. Trithorax-related genes regulate histone modifications differently during regeneration compared to homeostasis 259
6.4.3. Functional redundancy between *Smed-trr-1* and *Smed-trr-2* masks trithorax-related proteins' involvement in H3K4me1 270
6.5. Conclusion 275

Chapter 7. Final Discussion 277
7.1. The ancestral Mll3/4 gene homologue underwent multiple gene fission events throughout evolution 277
7.2. Planarian trithorax-related genes are expressed in stem cells and stem cell progeny 281
7.3. Trithorax-related genes control stem cell differentiation of selected lineages 283
7.4. Stem cell overproliferation results in outgrowth formation in *Smed-LPT*(RNAi) animals 286
7.5. Trithorax-related genes and planarian cell dynamics 290

7.6. Trithorax-related genes control H3K4me1, H3K4me3 and H3K27me3 294
7.7. Future directions 296

Chapter 8. Appendices 301
8.1. Appendix A 301
8.2. Appendix B 303
8.3. Appendix C 307
8.4. Appendix D 308
8.5. Appendix E 310
8.6. Appendix F 312
8.7. Appendix G 332
8.8. Appendix H 333
8.9. Appendix I 334
8.10. Appendix J 336
8.11. Appendix K 338

Chapter 9. Bibliography 339

LIST OF FIGURES AND TABLES

Figure 1.1. Planarian anatomy 4

Figure 1.2. Head regeneration timecourse following decapitation 7

Figure 1.3. Neoblast self-renewal and commitment across different lineages 26

Figure 1.4. Histone-modifying TrxG complexes in yeast, flies and mammals 44

Figure 2.1. Protein cleavage of MLL1 versus *Mll3/4* gene homologue split in *Drosophila* 50

Figure 2.2. *Mll3/4* gene homologue split in Diptera and Hymenoptera orders 54

Figure 2.3. Structure of PHD-fingers 58

Figure 2.4. Mechanism of lysine methylation 60

Table 2.1. Tblastn with human MLL3 and MLL4 and *Drosophila* LPT and Trr proteins against planarian transcriptomes 69

Table 2.2. Selected hits and their respective E-values after blastx with Transcript A against NCBI database 70

Figure 2.5. Sequence alignment and phylogenetic relation between the protein translation of Transcript A and proteins from MLL1/Trithorax and MLL3/4 families 72

Figure 2.6. Summary of the 5' and 3' RACE of Transcript A 75

Figure 2.7. Split and un-split *Mll3/4* homologues across phyla. 79

Figure 2.8. Two possible scenarios to explain the evolutionary history of *LPT/trr/Mll3/4* genes 81

Figure 2.9. Phylogenetic tree with MLL3/4 and Trr proteins 85

Figure 2.10. PHD-finger, SET domain and PHD-like zinc-binding domain conservation in trithorax-related proteins 89

Figure 2.11. NR boxes in *D. melanogaster* and *S. mediterranea* trithorax-related proteins 91

Figure 3.1. Expression patterns of trithorax-related genes in planarians 107

Figure 3.2. Expression of trithorax-related genes is radiosensitive 110

Figure 3.3. Trithorax-related genes are down-regulated following *Smed-H2B*(RNAi) and are expressed in x1, x2 and x ins FACS-sorted populations 113
Figure 3.4. Most neoblasts express trithorax-related genes 115
Figure 3.5. Trithorax-related genes co-localise 117
Figure 4.1. RNAi of trithorax-related genes is effective 134
Figure 4.2. Knockdown of trithorax-related genes affects regeneration 139
Figure 4.3. Knockdown of trithorax-related genes affects tissue homeostasis 141
Figure 4.4. Stem cell maintenance and overall early neoblast progeny production is unaffected following trithorax-related genes knockdown 144
Figure 4.5. Trithorax-related genes knockdown affects optic cup regeneration 147
Figure 4.6. GABAergic neuron production is decreased following trithorax-related genes knockdown 149
Figure 4.7. Dopaminergic neuron production is decreased following trithorax-related genes knockdown 151
Figure 4.8. Acetylcholinergic neuron production is decreased following trithorax-related genes knockdown 153
Figure 4.9. Serotonergic neuron production is decreased following trithorax-related genes knockdown 156
Figure 4.10. *Smed-pitx*-positive cell production is increased following trithorax-related genes knockdown 158
Figure 4.11. Cilia regeneration and homeostasis are unaffected following knockdown of trithorax-related genes 161
Figure 4.12. Ectopic serotonin treatment improves mobility of trithorax-related genes knockdown animals 164
Figure 4.13. Pharynx regeneration and homeostatic maintenance are compromised following trithorax-related genes knockdown 166
Figure 4.14. Gut regeneration and homeostatic maintenance are affected following knockdown of trithorax-related genes 169
Figure 4.15. Protonephridia regeneration and homeostasis are unaffected following trithorax-related genes knockdown 171
Figure 4.16. Epidermal progenitor cells regeneration and homeostasis are perturbed following trithorax-related genes knockdown 173

Figure 4.17. Summary of differentiation defects associated with trithorax-related genes knockdown 175
Figure 5.1. Hallmarks of cancer 178
Figure 5.2. A hypothesised symbiosis between the clonal evolution and the CSC model of tumourigenesis 185
Figure 5.3. Overproliferation defects in *Smed-LPT*(RNAi) animals 203
Figure 5.4. Stem cells outside of their usual compartments in *Smed-LPT* knockdown animals 205
Figure 5.5. Mitotic stem cells present in an outgrowth of a *Smed-LPT*(RNAi) animal 207
Figure 5.6. Abnormal nuclear morphology in *Smed-LPT*(RNAi) animals 209
Figure 5.7. Epidermal disorganisation and nuclear hyperplasia in *Smed-LPT*(RNAi) animals 211
Figure 5.8. Renewed *Smed-NB.21.11e* and *Smed-AGAT-1* expression in outgrowth tissue following *Smed-LPT* knockdown 213
Figure 5.9. Brightfield observations following shielded irradiation and amputation of *Smed-LPT*(RNAi) animals 215
Figure 5.10. *Smed-LPT*(RNAi) leads to increased stem cell migration 10 days post-amputation and 11 days post-irradiation 217
Figure 5.11. *Smed-LPT*(RNAi) leads to increased stem cell migration 17 days post-amputation and 18 days post-irradiation 219
Figure 5.12. *Smed-LPT*(RNAi) phenotype in the context of the hallmarks of cancer 222
Figure 6.1. Models of parental and newly synthesised histone dimers deposition at the replication fork 227
Figure 6.2. Kharchenko's 9 chromatin states 235
Figure 6.3. The three proposed modes of action of LPT-Trr/MLL3/4 243
Figure 6.4. *Smed-set1*(RNAi) and *Smed-utx*(RNAi) 249
Figure 6.5. Stem cell and early stem cell progeny dynamics at day 5 homeostasis following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1* 251
Figure 6.6. Early and late epidermal neoblast progeny dynamics at

day 5 homeostasis following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1* 254
Figure 6.7. Histone modification changes at day 5 homeostatic knockdown phenotypes 258
Figure 6.8. Stem cell and early stem cell progeny dynamics at day 3 regeneration following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1* 260
Figure 6.9. Early and late epidermal neoblast progeny dynamics at day 3 regeneration following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1* 262
Figure 6.10. Histone modification changes at day 3 and day 8 regeneration knockdown phenotypes 267
Figure 6.11. Double knockdown of *Smed-trr-1* and *Smed-trr-2* results in fast progressing outgrowth phenotype 272
Figure 6.12. Double *trr* knockdown phenotype results in decreased H3K4me1 at 3 days of regeneration 274
Figure 7.1. Trithorax-related genes' involvement in neoblast self-renewal control, commitment and progenitor differentiation 293
Figure 7.2. Three groups of gene promoter types 298
Table 7.1. The ideal future experiment 299

CHAPTER 1. GENERAL INTRODUCTION

1.1. THE PLANARIAN APPEAL

Planarians are freshwater flatworms that belong to the Platyhelminthes phylum of Lophotrochozoa (Reddien and Sánchez Alvarado, 2004). These animals have captured the imagination of scientists as iconic as the *Drosophila* geneticist T. H. Morgan, who was among the pioneers of the study of planarian biology. Planarians' most attractive feature is their almost limitless regenarative ability, the exploration of which holds promise for regenerative medicine. Flatworms are not at a more disadvantageous position than other model organisms, when it comes to evolutionary distance to humans. Well-studied model systems like *Drosophila melanogaster* and *C. elegans* are at roughly the same evolutionary distance to humans as flatworms are (847 million years), while the distance between *Drosophila* and planarians or planarians and *C. elegans* is 744 million years (Hedges et al., 2006; Hedges et al., 2015). This suggests that planarians could potentially provide insights into mammalian biology that *Drosophila* and *C. elegans* cannot and vice versa.

Planarians' regenerative capability is driven by a population of adult stem cells called neoblasts (Reddien and Sánchez Alvarado, 2004). Neoblasts also underpin processes such as growth and degrowth depending on nutritional availability (Baguñà et al., 1990). Planarian stem cells have three key properties – unlimited self-renewal capacity, ability to produce all cell types found in the planarian body (pluripotency) and capability to precisely interpret the differentiated cell type production needs

(Aboobaker, 2011; Reddien and Sánchez Alvarado, 2004). Understanding in detail these three characteristics and the relationship between them *in vivo* could advance induced pluripotent stem cells (iPSCs) therapies in regenerative medicine (Aboobaker, 2011). Furthermore, sexual and asexual planarians have proven useful in ageing research, more specifically, in understanding telomere maintenance differences in animals with sexual and asexual evolutionary histories (Tan et al., 2012). Recently, flatworms have been utilised for studying immunity mechanisms (Abnave et al., 2014). Planarians are also suitable for exploring the fine line between controlled highly proliferative stem cell environment and the proliferative disorganisation occurring in cancer. Studies in the field have already shown substantial conservation of genetic cancer-inducing mechanisms between mammals and planarians (Oviedo et al., 2008). Moreover, flatworm studies have informed the search for mammalian tumour suppressor genes (González-Estévez et al., 2012).

1.2. ANATOMY

Planarians possess nervous, digestive, excretory and reproductive (in sexual strains) systems, but lack respiratory, skeletal and circulatory systems (Reddien and Sánchez Alvarado, 2004). Brightfield observations allow the distinction of eyes and a pharynx. *In situ* hybridisation and immunohistochemistry allow a much finer characterisation of planarian structures. Eyes are located dorsally and are comprised of pigmented optic cups and photoreceptor neurons (Sato et al., 2005). The visual axons of the two eyes form an optic chiasm, located on the dorsal side of the brain (Yamamoto and Agata, 2011). The planarian central nervous system (CNS) is situated towards the

ventral side of the animal and is comprised of two cephalic lobes (joined by a commissure; forming an inverted U-shape) and two ventral nerve cords (VNCs) (Agata et al., 1998; Nishimura et al., 2007). The pharynx is used for feeding and waste expulsion and is positioned in the middle part of the body, anteriorly connected to the intestinal duct and opening at the animal's ventral side (Kobayashi et al., 1999). It is comprised mainly of neuronal, muscle and epithelial cells (Kido et al., 1964). The digestive system contains one anterior and two posterior gut branches. The elaborate arborisation of the gut is thought to facilitate the distribution of metabolites throughout the body, reminiscent of the role of the vascular system in more complex organisms (Brønsted, 1969). Histological analyses identified two cell types in the intestine – absorptive phagocytes and secretory goblet cells (Garcia-Corrales and Gamo, 1986). The protonephridial excretory system is composed of flame/tubule cells and two types of tubule-associated cells (Rink et al., 2011; Scimone et al., 2011). Planarian motility depends on ciliated epithelial cells on the ventral side of the worm (Reddien and Sánchez Alvarado, 2004). The main planarian structures, together with the neoblast pool, are shown in **Figure 1.1**.

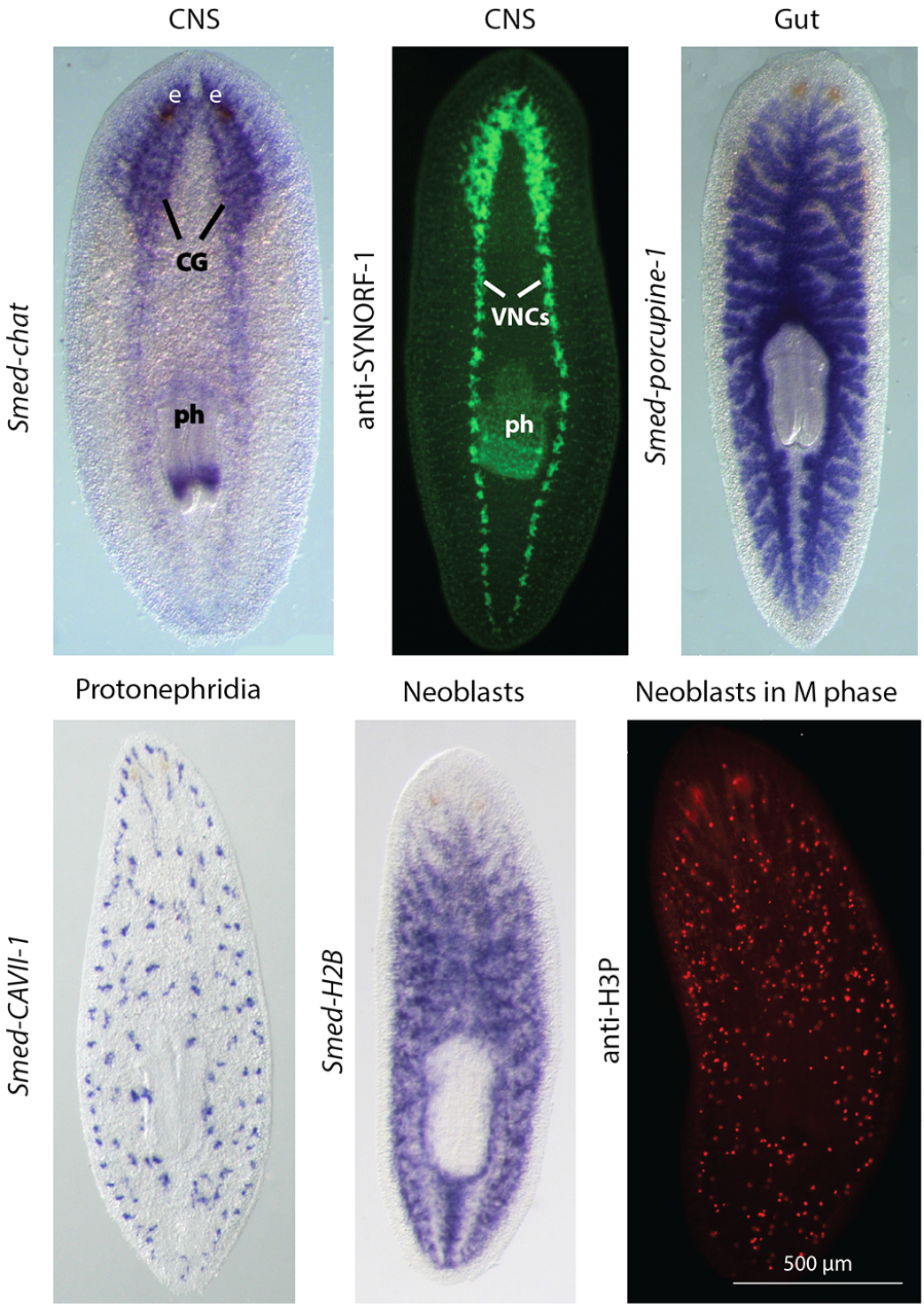


Figure 1.1. Planarian anatomy. Planarian structures labeled with *in situ* hybridisation RNA probes or antibodies. Two stains of the central nervous system (CNS) are shown – *Smed-chat* probe (labeling acetylcholinergic neurons and neuronal progenitors) and anti-SYNORF (anti-synapsin) antibody. 'CG' denotes cephalic ganglia, 'ph' – pharynx, 'VNCs' – ventral nerve cords. The three branches of the gut (one anterior and two posterior) are stained with *Smed-porcupine-1* probe. Protonephridia cells are labeled with *Smed-CAVII-1* (*in situ* hybridisation performed by Dr. Jaber-Hijazi). Neoblasts are detected via *Smed-H2B in situ* hybridisation (performed by Dr. Jaber-Hijazi). The neoblasts undergoing mitosis are labeled with anti-H3P (against phosphorylated serine 10 on histone H3) antibody.

1.3. REGENERATION

Animal regeneration often depends on de-differentiation mechanisms where differentiated cells return to a multipotent progenitor state capable of division, in order to produce more cells of the same differentiated type (Tanaka and Reddien, 2011). Such strategy is employed during zebrafish heart (Jopling et al., 2010; Kikuchi et al., 2010) and fin (Knopf et al., 2011) regeneration, as well as axolotl limb (Kragl et al., 2009) regeneration. A process of trans-differentiation, where one cell type transitions into another (either via reverting back to a pluripotent state or not) was also observed (Tanaka and Reddien, 2011). Examples of this process are seen in silk moths (Selman and Kafatos, 1974), newt cardiomyocytes (Laube et al., 2006) and *Hydra* (Bode et al., 1986; Wenger et al., 2016).

There are a number of studies advocating pluripotent neoblast self-renewal as the main regeneration tool in planarians. Transplantation of neoblast-enriched population of cells from an asexual donor into an irradiated sexual host (devoid of stem cells) restored the regenerative abilities of the irradiated host and converted it into an asexual animal (Baguñà et al., 1989). Transplantation of differentiated cells did not achieve such results (Baguñà et al., 1989). The conclusions from this study were solidified by a more recent experiment where a single neoblast transplanted from a wildtype donor was able to rescue an irradiated host and recover its regenerative abilities (Wagner et al., 2011). These findings strongly suggest that neoblast self-renewal is the primary resource utilised during planarian regeneration. They do not, however, completely exclude the possibility for some contribution of de-differentiation or trans-differentiation to that process.

Planarian regeneration can be separated into two processes – epimorphosis and morphallaxis. Epimorphosis is the process of rebuilding a missing organ from scratch, while morphallaxis is the remodeling of old tissue to accommodate newly regenerated parts (Morgan, 1901). While epimorphosis is associated with regulated burst of stem cell proliferation, morphallaxis is not (Morgan, 1901). In reality, instead of appearing sequentially, epimorphosis and morphallaxis intertwine during tissue recovery (Aboobaker, 2011). It was reported that injury that does not result in tissue loss induces a general neoblast proliferative response 6 hours after wounding (Wenemoser and Reddien, 2010). However, wounding resulting in tissue loss leads to a general (6 hours post-injury) and a regionally specified stem cell proliferative response (at the amputation region; 48 hours post-injury), referred to as the first and the second proliferative peaks respectively (Saló and Baguñà, 1984; Wenemoser and Reddien, 2010). It was demonstrated that the accumulation of stem cells close to the wound is not a result of passive spread of mitotic cells, but of active stem cell migration towards the wound site (Guedelhoefer and Sánchez Alvarado, 2012). The differentiating post-mitotic neoblast progeny cells accumulate to form a regenerative blastema (Eisenhoffer et al., 2008a; Wenemoser and Reddien, 2010), a transparent group of cells at the wound site where recovery of missing tissues happens (Brockes and Kumar, 2008). An example of a *S. mediterranea* planarian regenerating a missing head is shown in **Figure 1.2.**

Anterior regeneration

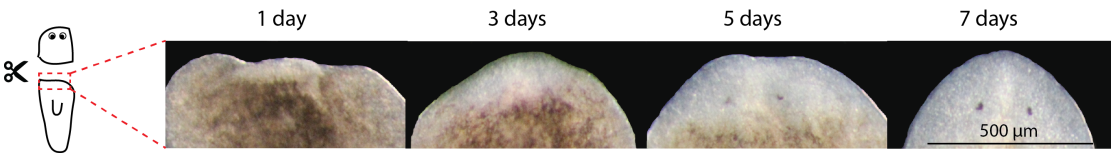


Figure 1.2. Head regeneration timecourse following decapitation. The red dashed box on the schematic shows the imaged brightfield area of the regenerating worm.

Interestingly, the first proliferative peak is preceded by a burst of differentiated cells undergoing apoptosis close to the amputation/wound site, presumably making space for the new cells involved in blastema formation (Pellettieri et al., 2010). The second neoblast proliferation peak is followed by a general second apoptotic peak, possibly important for the ensuing morphallactic process of regeneration (Pellettieri et al., 2010).

The first and most critical part of regeneration is wound response. An attempt was made to classify gene expression patterns in tissues as a function of wound response. Three main categories were identified – 'first wave short' (W1), 'first wave long' (W2) and 'second wave' (W3) (Wenemoser et al., 2012). W1 contained genes expressed within 30 minutes post-injury, but down-regulated within 6-12 hours post-injury. Such genes were expressed near the wound site and included canonical wound response genes like *jun-1, fos-1, egrl1, egr-2, egr-3, egr-4*. This group was hypothesised to have a role in wound closure, defense against pathogens, initiation of proliferation and apoptosis induction. W2 was comprised of genes up-regulated within 30-60 minutes after injury, whose expression persisted 12 hours after wounding. These genes were expressed subepidermally and close to the wound and included genes like *wntless, wnt1, inhibin-1, noggin-like1, follistatin*. A prominent feature of these genes was that they are known to be required for correct patterning during regeneration and homeostatic tissue maintenance. W3 genes were expressed 3-12 hours after wounding (persisting until 23 hours post-injury), mostly in the epidermis, often not in proximity to the wound. This list included *delta-1, plasminogen-1* and some previously unidentified genes, implicating W3 in Notch signaling and extracellular matrix remodeling. A group called W4 was also identified,

but instead of expression in differentiated cells, it comprised of transcripts present in stem cells in response to wounding (ex.: *runt-1* and *runt-2*). *Runt-1* was found to be important for cell type specification during regeneration, promoting wound-proximal neoblast heterogeneity. Overall, these studies demonstrated that, even though different planarian wounds require specifically tailored regenerative programs in the long run, the initial wound healing response is remarkably similar (Wenemoser et al., 2012).

1.4. NEOBLAST FEATURES AND THE CHROMATOID BODY

Neoblasts are widely cited as being the only proliferative stem cells in planarians, comprising 20-30% of the cells in the body (Baguna et al., 1989). They are situated in the mesenchymal tissue, not present in the pharynx or the region anterior to the eyes (Reddien and Sánchez Alvarado, 2004). Planarian stem cells have a diameter of less than 10 µm, large nucleus, scant cytoplasm, very few mitochondria and no endoplasmic reticulum (Higuchi et al., 2007). Neoblasts also possess electron-dense particles comprised of large protein-RNA complexes called chromatoid bodies located in their cytoplasm (Morita et al., 1969). The chromatoid bodies are hypothesised to be of great importance to the process of stem cell maintenance (Higuchi et al., 2007). It was observed that, as neoblast differentiation progresses, chromatoid bodies decrease in both size and abundance (Hori, 1997).

Research outside of the planarian field has shown that chromatoid bodies are an accumulation of male germ cells in haploid spermatids (Kotaja and Sassone-Corsi,

2007), or the mammalian structure analogue to germ plasm (Parvinen, 2005). A popularly held view is that chromatoid bodies are germline cells' equivalent to the Processing bodies (P bodies) found in somatic cells (Kotaja and Sassone-Corsi, 2007). Evidence for this hypothesis was provided by Kotaja et al. (2006) who demonstrated that P body markers are abundant in chromatoid bodies. P bodies are viewed as a storage station for Argonaute proteins, miRNA and mRNA destined to be degraded (Cougot et al., 2004). Chromatoid bodies are thought of as the structures where miRNA and RNA-decay processes meet, making chromatoid bodies the organisers of RNA metabolism (Kotaja and Sassone-Corsi, 2007).

In planarians, chromatoid bodies are present in neoblasts as well as germline cells (Teshirogi and Ishida, 1987) and neurons (Fernandéz-Taboada et al., 2010). As differentiation progresses, the chromatoid bodies disappear in all structures but the germline and neurons (Coward, 1974; Yoshida-Kashikawa et al., 2007).

A great amount of genes classically used as neoblast markers are associated with the chromatoid body. One such example is *piwi,* part of the Piwi/Argonaute/Zwille (PAZ) RNA-binding protein family, involved in the production of piRNAs, known to be involved in transposon gene silencing (Ishizu et al., 2012). In the planarian species *Dugesia japonica* there is one *piwi* orthologue – *DjPiwi-1* (Rossi et al., 2006), while *Schmidtea mediterranea* has three – *Smedwi-1, Smedwi-2, Smedwi-3* (Palakodeti et al., 2008; Reddien et al., 2005b). *DjPiwi-1*(RNAi) led to compromised formation of blastema in only 10% of the animals, only in anterior pieces regenerating posterior structures (Rossi et al., 2006). This weak phenotype might be related to the fact that *Djpiwi-1* was only expressed in a subpopulation of *D. japonica'* stem cells (Rossi et

al., 2006). Knockdown of *Smedwi-1* produced no observable defects, while *Smedwi-2*(RNAi) and *Smedwi-3*(RNAi) led to severe defects in homeostatic tissue renewal as well as impaired regeneration of missing structures, correlated with a decrease in piRNAs (Palakodeti et al., 2008; Reddien et al., 2005b). Neoblast number was still intact in the described phenotypes, meaning that *Smedwi-2* and *Smedwi-3* play a role mainly in neoblast progeny cell production (Palakodeti et al., 2008; Reddien et al., 2005b).

It was demonstrated that planarian chromatoid bodies contain histone mRNA (Rouhana et al., 2014). SMEDWI-1 and SMEDWI-3 were found to be important specifically for the localisation of *germinal histone H4* mRNA to the chromatoid bodies (Rouhana et al., 2014). Furthermore, while only 20-30% of identified planarian piRNAs bind to transposable elements (Resch and Palakodeti, 2012), piRNAs with homology to loci on all core histones' mRNA were found (Rouhana et al., 2014). These results suggest that PIWI proteins and chromatoid bodies are involved in histone mRNA processing in neoblasts (Rouhana et al., 2014).

The TUDOR protein is associated with piRNA regulation and transposon silencing via its interaction with PIWI (Siomi et al., 2010). The *tudor* gene homologue in the planarian *Schmidtea polychroa* is expressed in the chromatoid body and was found to have a function in neoblast maintenance (Solana et al., 2009).

Another chromatoid body-related neoblast gene is *Smed-bruli* (an RNA-binding protein belonging to the Bruno-like family), which is expressed in stem cells and the CNS (Guo et al., 2006). It was hypothesised that SMED-BRULI binds to 3'UTR regions of transcripts, preventing their translation (Guo et al., 2006). *Smed-bruli* is involved in

neoblast maintenance (Guo et al., 2006). Similar involvement in stem cell self-renewal was also reported for the PUF protein PUMILIO's *D. japonica* homologue (Salvetti et al., 2005). PUMILIO had previously been associated with germline stem cell maintenance and embryonic patterning (Forbes and Lehmann, 1998). Two more genes expressed in the planarian chromatoid body were identified by Shibata et al. (1999) – the *vasa(vas)*-related genes – *DjvlgA* and *DjvlgB.*

It was proposed that *piwi, nanos, tudor, vasa, bruno* and *pumilio* in other organisms are involved in the germline multipotency program, i.e. they are expressed in both germline cells and multipotent precursors of the germline (Juliano et al., 2010). This program appears to also be active in planarian somatic stem cells (Aboobaker, 2011).

1.5. METHODS FOR NEOBLAST ABLATION AND ISOLATION

Understanding planarian neoblast biology has been aided immensely by the development and adaptation of methods for stem cell isolation and selective destruction.

1.5.1. GAMMA-RADIATION

Gamma irradiation leads to double strand breakage of DNA, which, depending on its severity, either results in activation of DNA repair mechanisms or death of the affected cell via apoptosis or necrosis (Pawlik and Keyomarsi, 2004). P53, being the guardian of the two crucial DNA damage checkpoints (G1/S and G2/M), is one of the

most important coordinators of cell cycle progression, apoptosis and DNA repair (Pawlik and Keyomarsi, 2004). Experimentally, cells in G2/M transition are the most sensitive to radiation damage, followed by cells in G1 and end of S phase (Pawlik and Keyomarsi, 2004).

Since neoblasts are the only cells going through the cell cycle in asexual planarians, it stands to reason that they are the most susceptible to irradiation damage. Not surprisingly, exposure of planarians to gamma radiation leads to depletion of proliferating cells and increase in apoptosis (Dubois, 1949; Rossi et al., 2006). Amputation immediately after a strong dose of gamma irradiation (80–100 Gy) resulted in the formation of a blastema, which regressed soon after and the animal died (Chandebois, 1976; Reddien and Sánchez Alvarado, 2004). Alternatively, if the animals were cut 7 days after radiation exposure, blastema never formed and death followed (Dubois, 1949). The reason for this discrepancy was illustrated by Eisenhoffer et al. (2008). Neoblast post-mitotic progeny cells survive after irradiation and proceed to reach terminal differentiation eventually, but there are no neoblasts left to replace them. Therefore, if expression of neoblast transcripts takes 24-48 hours to disappear, transcripts expressed in stem cell progeny could take 4-7 days (or more) to completely disappear after irradiation (Eisenhoffer et al., 2008). If planarians were to be amputated immediately after irradiation, the available stem cell progeny would be recruited to form a regenerative blastema, which later would regress due to the discontinued production of stem cells (Reddien and Sánchez Alvarado, 2004). If there was a delay between irradiation and amputation, the stem cell progeny available after irradiation would have already been used up for

homeostatic maintenance and, with post-irradiation neoblast depletion, no blastema would be formed (Reddien and Sánchez Alvarado, 2004).

Studies demonstrated that planarian neoblast populations can recover completely after sublethal doses of irradiation (5 Gy) (Salvetti et al., 2009). The studied species, *D. japonica*, showed an initial sharp decrease in stem cell number, but after 4 weeks, it had recovered its stem cell population to levels comparable to an untreated worm (Salvetti et al., 2009). Such findings show that planarians are capable of DNA repair caused by low doses of irradiation.

1.5.2. SMED-H2B(RNAI)

Knockdown of histone H2B is an alternative way of stem cell ablation in planarians. Histone H2B is part of the histone octamer core of nucleosomes (Jenuwein and Allis, 2001). All four histones' (H2A, H2B, H3, H4) mRNA is produced during S phase of the cell cycle (Hewitson et al., 2006). The *H2B* gene homologue in the planarian *S. mediterranea* (*Smed-H2B*) was shown to have a radiosensistive neoblast expression, very similar to that of *Smedwi-1* (Solana et al., 2012). *Smed-H2B*(RNAi) animals had a regeneration phenotype not unlike that of lethally irradiated worms – drastically decreased blastema size, compared to control knockdown, followed by death (Solana et al., 2012). It was observed that 5 days following the last day of *Smed-H2B*(RNAi) treatment, neoblast markers like *Smedwi-1*, *Smedwi-2* and *Smed-tud-1* (the last two also expressed in CNS) were down-regulated to levels comparable to those at 2 days after lethal irradiation (Solana et al., 2012). *Smedwi-2* and *Smed-tud-1*'s expression in the CNS, however, was better preserved after *Smed-H2B*(RNAi) compared to post-

irradiation (Solana et al., 2012). Such findings suggest that *Smed-H2B*(RNAi) is, in fact, a more specific/targeted way of neoblast gene expression depletion than irradiation (Solana et al., 2012).

Two hypotheses explaining *Smed-H2B*(RNAi) phenotype could be formulated. The assumption underlying both is that *Smed-H2B* mRNA is depleted before the cell division cycle has finished. Having encountered this problem, the neoblast could either choose to divide and produce another sub-optimal stem cell, or choose not to undergo mitosis. If division is preferred, nucleosome formations might be compromised due to the inadequate amounts of newly synthesised H2B protein. This could result in an increased number of unbound histones, a known differentiation stimulus in embryonic stem cells (ESCs) (Meshorer et al., 2006). Alternatively, if the cell decides not to undergo division, it could instead go into differentiation mode where no newly made histone proteins are required. Both of these explanations could account for the increase in neoblast progeny transcripts at the early stages of the phenotype (Solana et al., 2012).

1.5.3. Fluorescence activated cell sorting

Fluorescence Activated Cell Sorting (FACS) has proven very useful for isolation of planarian neoblasts and other radiosensistive cells. The most common methods include staining dissociated cells with Hoechst 33342 (nuclear dye) and calcein AM, as a measure of nuclear size and cytoplasmic complexity respectively (Hayashi et al., 2006; Higuchi et al., 2007; Romero et al., 2012). Three distinct populations of planarian cells can be observed – x1 (G2/M neoblasts), x2 (mixture of G1 neoblasts

and neoblast progeny); x ins (radiation insensitive population; terminally differentiated cells) (Hayashi et al., 2006; Higuchi et al., 2007; Romero et al., 2012; Shibata et al., 2010; Eisenhoffer et al., 2008). Studies have shown that 90% of the cells in x1 and 10-20% of those in x2 are *Smedwi-1*-positive (Eisenhoffer et al., 2008; Labbé et al., 2012; Önal et al., 2012; Resch et al., 2012; Zhu et al., 2015). The x2 fraction is enriched for known stem cell progeny markers (Eisenhoffer et al., 2008b; Labbé et al., 2012; Önal et al., 2012; van Wolfswinkel et al., 2014; Zhu et al., 2015). These findings make the x1 fraction the preferred group of cells when a clean neoblast population is required.

1.6. NEOBLAST HETEROGENEITY

A long-standing question in planarian biology is whether the population of neoblasts is homogenous or heterogenous. This section presents studies clearly identifying planarian adult stem cells as a heterogenous pool.

1.6.1. VARIABLE SENSITIVITY TO IRRADIATION

One way in which the neoblast population exhibits heterogeneity is associated with gamma-radiation sensitivity. The primary example is *D. japonica's piwi* expression. Interestingly, neoblasts located along the midline of the animal, positive for both *DjPiwi-1* and *DjMCM2* (another neoblast gene) markers, were noticeably more radiosensitive than the dorso-laterally located *DjMCM2*-positive/*DjPiwi-1*-negative parenchymal cells (Rossi et al., 2006). These findings are in line with the suggestion

that neoblasts situated towards the animal' periphery are less sensitive to gamma-radiation in *D. japonica* (Salvetti et al., 2009).

One possible explanation for the observed differences in radiation sensitivity among stem cell populations is related to the cell cycle. Salvetti et al. (2009) speculated that *D. japonica*'s neoblasts have different cycling rates, resulting in faster cycling cells exhibiting less radiation tolerance. In support of this, Rossi et al. (2006) reported that there was a greater amount of mitotic cells in the midline stem cell population compared to the lateral populations. It is possible that faster cycling stem cells spend less time at the cell cycle checkpoints and, therefore, fail to initiate DNA repair when needed (Kastan and Bartek, 2004). In support of this interpretation, it was reported that mammalian embryonic stem cells had a mostly non-functional G1/S checkpoint, which helped them progress faster through the cell cycle (Ballabeni et al., 2011; White and Dalton, 2005).

Furthermore, regions with clustered neoblasts seem to be less sensitive to radiation than areas where neoblasts are more spread out (Salvetti et al., 2009). Such observations led Salvetti et al. (2009) to propose that there is a hierarchy of stem cells in *D. japonica* comprised of: (1) 'true' stem cells – highly sensitive to DNA damage; (2) more resistant to DNA damage neoblasts, possibly in a state of early lineage commitment; (3) radiosensitive proliferating transit stem cell progeny.

Interestingly, in another planarian species, *S. mediterranea* (the organism of choice for the experiments reported in this thesis), *Piwi* transcripts are expressed in all neoblasts, which exhibit a much more uniform response to irradiation when compared to *D. japonica* (Palakodeti et al., 2008; Reddien et al., 2005b). A possible

explanation is that the stem cell population in *S. mediterranea* is better synchronised than that of *D. japonica*. Unlike in *D. japonica*, no regions of the *S. mediterranea* body had been observed to undergo more mitoses than others. Therefore, *S. mediterranea* stem cells might not have the cell cycle heterogeneity observed in *D. japonica* and could be going through the cell cycle at equal speed and have comparable irradiation damage as a result.

1.6.2. THE PROBLEM OF G2- ARRESTED STEM CELLS

It was speculated that, in order for the first mitotic peak to happen as soon as 6 hours after injury (or feeding), there must be a population of neoblasts arrested in G2 ready to quickly respond to injury (or feeding) (Saló and Baguñà, 1984). This hypothesis was tested by Newmark and Sánchez Alvarado (2000) via observing co-localisation between the S phase marker BrdU (Bromodeoxyuridine; thymidine analogue) and the M phase marker anti-H3P. The authors hypothesised that if a G2-arrested cell population existed, the first cells labeled by anti-H3P (M phase) after injury (or feeding) would not be positive for BrdU (would not have gone through S phase). Thirty percent of M phase cells were co-labeled with BrdU as early as 4-6 hours after feeding. Furthermore, the median length of G2 was calculated to be around 6 hours, deemed to be sufficient to explain the proliferation peak seen after injury or feeding, without resorting to a specifically G2-arrested population. Finally, continuous BrdU-labeling experiments (reaching almost 100% labeled cells in the animal after 3 days) indicated that the presence of slow-cycling and/or G2-arrested

population of stem cells is unlikely in *S. mediterranea* (Newmark and Sánchez Alvarado, 2000).

Wenemoser and Reddien (2010) further explored the question of neoblast cell cycle changes in response to injury. Their supposition was that one mechanism that could account for the first proliferation peak in response to injury/feeding would be related to either shortening of S phase or shortening of G2 phase (Wenemoser and Reddien, 2010). Provided that the S phase was shortened, the authors expected to observe an increase in the S phase neoblasts prior to the proliferation peak. Alternatively, if G2 phase were shortened, this would not be the case. The team failed to observe a drastic increase in S phase neoblasts, therefore concluding that G2 shortening was a more likely feature of the first proliferation peak (Wenemoser and Reddien, 2010). It was previously observed that mammalian embryonic stem cells had significantly decreased duration of the gap phases of the cell cycle (G1 and G2), allowing them to cycle quicker compared to differentiating or differentiated cells (Savatier et al., 1994; White and Dalton, 2005). Therefore, it is not surprising that similar mechanisms might be at place during planarian stem cell-dependent wound response.

1.6.3. NEOBLAST POTENCY

1.6.3.1. Neoblasts, germline and germline precursors

Germline stem cells appear in infant sexual planarians, meaning that neoblasts can fall under two categories – somatic neoblasts, important for regeneration and tissue homeostasis and germline stem cells – responsible for germ cell production during

sexualisation (Sato et al., 2006). The great majority of studies investigating planarian neoblast potency, including the work presented in this thesis, were performed on asexual flatworms that do not contain germline. Neoblasts in asexual flatworms, however, share a significant portion of their gene expression patterns with the germline cells in sexually reproducing flatworm strains (Newmark et al., 2008; Solana et al., 2009). One crucial difference is the germline expression of NANOS, a CCHC zinc-finger protein involved in germline maintenance and development (Shibata et al., 2010). *Nanos* expression is present in the sexual strain's germline, as well as in the germline precursor cells of asexual planarians (Sato et al., 2006; Wang et al., 2007). The asexual's germline precursor cells got their name due to the fact that, upon sexualisation of an asexual animal, they convert into a functional germline (Sato et al., 2006). Interestingly, it was shown that both the germline cells and the germline precursors co-expressed neoblast markers (Wang et al., 2007). Furthermore, *nanos* expression was found to be radiosensitive (Salvetti et al., 2009). Knockdown of *nanos* led to disappearance of germline precursor cells (asexuals), as well as ovaries and testes (sexuals), but the stem cell-dependent regenerative capacity of the worm remained intact (Wang et al., 2007). These findings are among the earliest examples of the existence of neoblasts with different potency in planarians.

1.6.3.2. cNeoblasts

The first significant stride in answering the question of whether planarian neoblasts are a uniformly pluripotent population or a heterogeneously potent pool of cells was

made by Wagner et al. (2011). The authors subjected *S. mediterranea* to a dose of irradiation (17.5 Gy), which led to 78% of the animals losing *Smedwi-1* expression completely, while 22% of animals started forming colonies of *Smedwi-1*-positive cells a week after irradiation. Wagner et al. (2011) called these specific colony-forming stem cells clonogenic neoblasts or cNeoblasts. An even greater feat of the study was that it demonstrated that transplantation of a single neoblast (a cNeoblast) was capable of rescuing a lethally irradiated animal.

The rescue success rate was around 20%, which could either indicate that only a small proportion of neoblasts were capable of repopulating an irradiated animal with pluripotent stem cells, or that there were technical limitations of the procedure, or both. What the experiment clearly showed, however, is that, at least some part of the planarian neoblast population, is truly pluripotent. Nevertheless, the results do not exclude the possibility of a transplanted committed neoblast undergoing de-differentiation and becoming pluripotent again. Previous studies in mice showed that, upon transplantation, a multipotent cell could revert back to pluripotency (Blanpain et al., 2007).

A follow-up cNeoblast study revealed the genes involved in colony expansion, and those involved in differentiation of the stem cells in the colony (Wagner et al., 2012). Zinc finger containing genes (*fhl-1, zmym-1 and zfp-1*), as well as transcription factors (*soxP-1, p53, junl-1*), chromatin modifying genes (*setd8-1, ezh, sz12-1, eed-1*) and genes coding for RNA-binding proteins (*Smedwi-2, Smedwi-3, bruli, khd-1, cip29, vasa-1*) were found to be involved in clonal expansion processes (Wagner et al., 2012). The same classes were also important for clonal differentiation –RNA-binding

(*vasa-1*), zinc finger (*zfp-1*), transcription factor (*p53*), chromatin modifying (*CHD4*) (Wagner et al., 2012).

1.6.3.3. Committed neoblasts

A recent study suggested that cNeoblasts might actually belong to the so-called 'sigma' class of stem cell genes (van Wolfswinkel et al., 2014). A single cell qPCR-based method on FACS-sorted neoblasts revealed the existence of two main classes of stem cells – 'sigma' and 'zeta' (van Wolfswinkel et al., 2014). Neither of these classes were defined by changes in cell cycle stages. Zeta class was associated with neoblasts committed to epidermal fate and express genes like *Smed-zfp-1*, *Smed-soxP-3* and *Smed-egr-1*. Sigma class, on the other hand, showed high expression of genes like *Smed-fgfr-4* and the Sox transcription factors *Smed-soxP-1* and *Smed-soxP-2*. The sigma cells were shown to be the primary population of stem cells responsible for wound response. They were also demonstrated to be able to self-renew in the absence of the zeta pool and to even recover the depleted zeta pool following *Smed-zfp-1*(RNAi). These findings led to the interpretation that the collective sigma population is truly pluripotent. Furthermore, the zeta class was shown to be dependent on the sigma population for its maintenance. A sigma cell could transition into a zeta during the S phase, but once this 'fate'-ing cell cycle was complete, the zeta cell could not undergo any more mitoses (van Wolfswinkel et al., 2014).

These results suggest that cNeoblasts belong to the sigma class (van Wolfswinkel et al., 2014). However, the possibility of a transplanted zeta neoblast rescuing an

irradiated host via re-entering a pluripotent state cannot be excluded by the experimental techniques available in the field.

An important point made by van Wolfswinkel et al. (2014) was that the sigma and zeta classes were not entirely homogenous and it was possible to see distinct subpopulations within them. The most prominent example was the 'gamma' population, a subclass of sigma. The authors observed that the gamma pool of cells was enriched for genes like the endoderm-promoting transcription factors *Smed-hnf-4* and *Smed-gata4/5/6*, often associated with expression in non-stem cell gut progenitors. Therefore, it is likely that the gamma population was mainly comprised of stem cells fated to become gut cells (van Wolfswinkel et al., 2014).

Van Wolfswinkel et al.'s (2014) study was not the first one to identify neoblasts committed to different lineages. In most such studies, sufficient evidence for a committed neoblast is considered when a cell positive for a canonical neoblast marker (usually *Smedwi-1),* is also positive for a marker associated with differentiated tissues. In 2011, Lapan and Reddien identified neoblasts double positive for *Smedwi-1/Smed-H2B* and differentiated eye cup markers like the transcription factors *Smed-sp6-9, Smed-dlx, Smed-six1-2, Smed-eya*. These cells were located anterior to the pharynx, at the base of a 'trail' going from committed stem cells through neoblast progeny to differentiated eye cup cells as the trail approached the region of the eyes (Lapan and Reddien, 2011).

Lapan and Reddien answered another pressing question – namely, are the committed neoblasts only present during regeneration, or are they also involved in homeostatic tissue turnover as well? *Smed-ovo* was found to be expressed in

differentiated eyes, as well as neoblast eye precursor cells (*Smedwi-1+)* (Lapan and Reddien, 2012). Knockdown of *Smed-ovo* revealed that the gene was important for homeostatic maintenance of the eye. Such findings suggest that committed neoblasts are not just a feature emerging under the stress of regeneration, but rather a general phenomenon existing in uninjured planarians (Lapan and Reddien, 2012). However, whether this is the case for all committed neoblasts described in the current text, remains to be established. Another unknown is to what extent recovery of missing tissues depends on already existing committed neoblasts versus the pluripotent sigma population producing the required upon injury committed stem cells. It is possible that, similar to what had been discovered about the zeta population (van Wolfswinkel et al., 2014), committed neoblasts across all lineages do not have unlimited self-renewal capacity and depend mainly on sigma for their maintenance. In this case, the immediate response to tissue amputation could be performed by the already committed stem cells of the respective relevant lineages, followed by the sigma population producing variable amounts of committed stem cells as guided by the gravity of the injury.

Neoblasts committed to the excretory cell lineage were identified by Scimone et al. (2011). These cells were positive for both *Smedwi-1/Smed-H2B* and *Smed-Six1/2-2* (transcription factor)*, Smed-Sall* (transcription factor) and *Smed-POU2/3* (POU family of homeobox genes). Similar reports exist for pharyngeal neoblast precursors, double positive for *Smedwi-1* and the forkhead transcription factor *Smed-FoxA* (Adler et al., 2014). Cowles et al. (2013) described the possibility of pan-neuronal progenitor stem cell that was triple positive for *Smedwi-1* and the helix-loop-helix neurogenesis factors *Smed-sim* and *Smed-coe*. It was shown that there were also populations of

cells that were positive for *Smedwi-1* and *Smed-coe* OR *Smed-sim*, presumably indicating a step further in the neoblast commitment (Cowles et al., 2013). Furthermore, the homeodomain transcription factors *Smed-pitx* and *Smed-lhx1/5-1* were also co-expressed with *Smedwi-1* in serotonergic neuronal precursors (Currie and Pearson, 2013).

Interestingly, SMEDWI-1 protein is expressed in stem cells, but it also persists in early stem cell progeny cells (Guo et al., 2006; Wagner et al., 2011). Therefore, cells positive for the SMEDWI-1 protein, but negative for the *Smedwi-1* transcript are representative of early stem cell progeny that is post-mitotic (Guo et al., 2006; Wagner et al., 2011).

A summary of the research and an emerging model on planarian stem cell potency is presented in **Figure 1.3**.

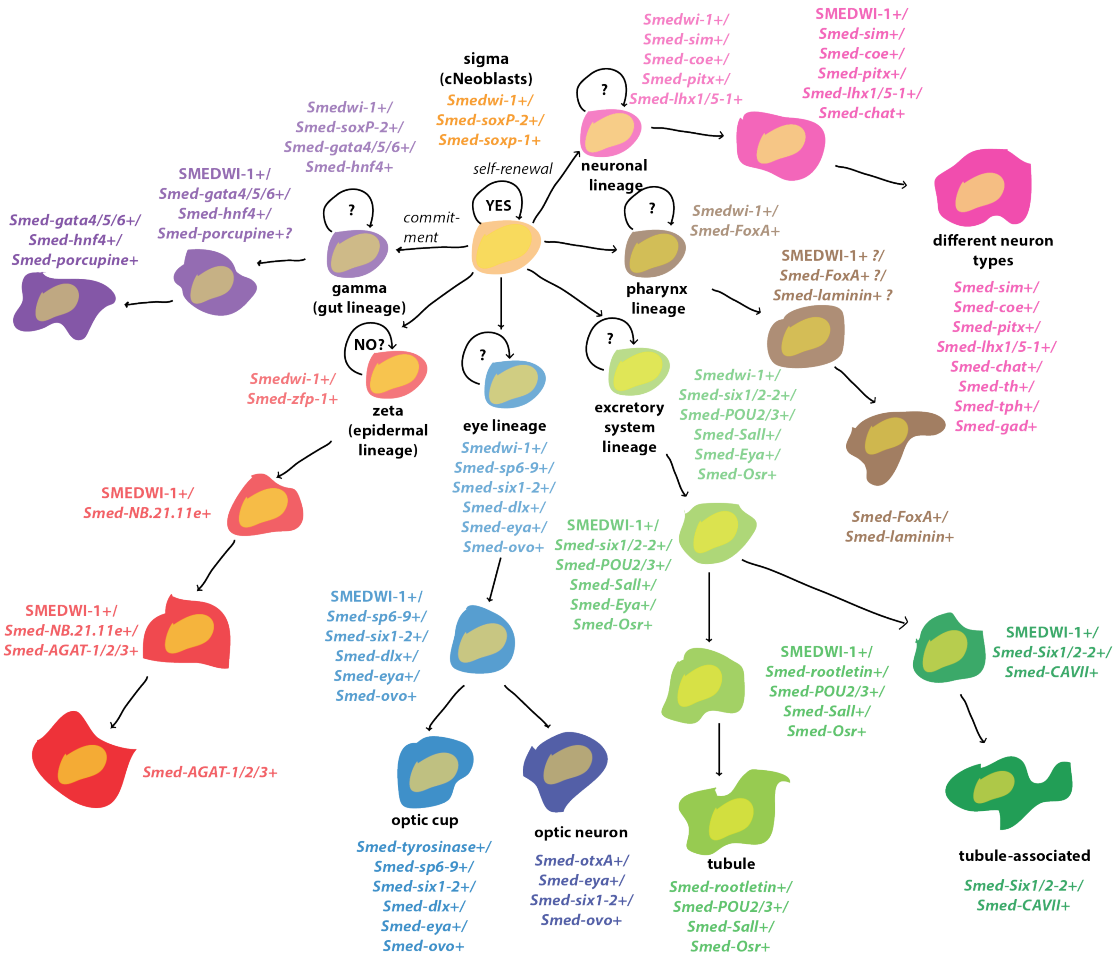


Figure 1.3. Neoblast self-renewal and commitment across different lineages. The darker the shade and the more elaborate the shape, the more differentiated the cell. Circular arrows indicate self-renewal, while straight arrows represent commitment and differentiation. Characteristic gene (and SMEDWI-1 protein) expression for each stage is presented with a matching colour for each lineage described. The neuronal lineage is too diverse, hence, for the sake of succinctness, only one terminally differentiated neuron is presented with genes expressed throughout different neuronal lineages. Data on germline specification is not included in this figure. The above figure does not claim to be an exhaustive list of all the markers identified for each differentiation stage in each lineage.

1.7. EPIGENETICS AND FLATWORMS

The definition of epigenetics pertains to the heritability (both mitotic and meiotic) of DNA sequence-independent patterns of gene expression (Jeltsch, 2006; Vaissière et al., 2008). Epigenetic mechanisms have a crucial role in the development of a fertilised egg into a mature organism (Waddington, 1956). The most prominent epigenetic processes involve post-translational histone modifications and DNA methylation, which can often act together (Vaissière et al., 2008).

1.7.1. DNA METHYLATION

The most common form of DNA methylation is 5-methylcytosine, deposited by DNA methyltransferases (DNMTs) (Bird, 1984; Jeltsch, 2002). There are four known DNA methyltransferases in mammals – DNMT1, DNMT2, DNMT3a and DNMT3b (Jeltsch, 2006). DNMT1 maintains methylation following replication via depositing methyl groups on hemimethylated CpG dinucleotides (Jeltsch, 2006). DNMT2 is not directly involved in DNA methylation (Jeltsch, 2002; Schaefer and Lyko, 2010). It was identified as a $tRNA^{Asp}$ methylator in mouse, *A. thaliana* and *Drosophila* (Goll et al., 2006). DNMT3a and DNMT3b perform *de novo* methylation (Okano et al., 1999).

In funghi only repetitive sequences have 5-methylcytosine deposition, an occurrence associated with transposon silencing (Selker et al., 2003). *A. thaliana*'s DNA methylation is observed in regions containing transposable elements, as well as some promoter regions and gene bodies (Zhang et al., 2006). In mammals, DNA methylation is seen throughout the entire genome with the exception of short

unmethylated regions (usually associated with gene promoters) termed CpG islands (Bird, 2002, 1986). However, in cancer it is common to see an inverse pattern – decreased overall methylation, but increased promoter methylation on specific genes (Berman et al., 2012). The methylated sequences in mammals include genes, transposons and intergenic DNA (Eckhardt et al., 2006). One of the most important functions of DNA methylation in mammalian genomes is transcriptional silencing, a key example of which is promoter methylation of genes located on the inactivated X chromosome (Mohandas et al., 1981; Venolia et al., 1982). It has been postulated that DNA methylation of gene bodies is associated with prevention of spurious initiation of transcription, thus reducing transcriptional noise (Simmen et al., 1999; Zilberman et al., 2007). The cited DNA methylation studies lead to the interpretation that DNA methylation has a mostly genome protective role (transposon silencing) in plants and funghi, while the focus has shifted more towards gene regulation in mammals (Suzuki and Bird, 2008). Histone modifications and DNA methylation converge on silencing pluripotency-related genes during cell differentiation (Reik, 2007). Hypomethylation of CpG-rich regions is evident in stem cells, but during differentiation, CpG-rich promoter and non-promoter regions often become hypermethylated, resulting in silencing (Meissner et al., 2008). DNA methylation is also involved in chromatin structure, X chromosome inactivation, imprinting and chromosome stability (Robertson, 2005). As a consequence, aberrant DNA methylation is implicated in a plethora of diseases like Beckwith-Wiedemann, Prader-Willi, Angelman, Fragile X syndromes and many cancers (Robertson, 2005). Because of these reasons, DNA methylation has received a lot of attention in biological research. However, the most widely studied form of DNA methylation – 5-

methylcytosine, is not universally present in popular animal model organisms. In *Drosphila*, the only conserved DNA methyltransferase – DNMT2, was found to deposit low amounts of 5-methylcytosine, preferentially on CpT and CpA dinucleotides in embryos (Kunert et al., 2003). *C. elegans* lacked any detectable DNMTs and 5-methylcytosine methylation (Gutierrez and Sommer, 2004; Simpson et al., 1986). Interestingly, recently, methylation of N^6-Adenine was reported in both *Drosophila* (Zhang et al., 2015) and *C. elegans* (Greer et al., 2015).

In flatworms, DNA methylation studies have been controversial. Parasitic flatworms were initially found to lack DNA methylation (Regev et al., 1998). Surprisingly, Geyer et al. (2011) reported the existence of DNMT2-mediated cytosine methylation in the parasite *Schistosoma mansoni.* It was also demonstrated that DNA methylation in this organism was responsible for egg production and maturation (Geyer et al., 2011). A more extensive study of DNA methylation in Platyhelminthes revealed that representative species from all four Platyhelminth classes (Monogenea, Trematoda, Turbellaria and Cestoda) had a *Dnmt2* orthologue and detectable levels of cytosine methylation (Geyer et al., 2013). The Turbellarian species examined by Geyer et al. (2013) was *Polycelis nigra*. Importantly, another Turbellarian species – *S. mediterranea* also was found to have a conserved *Dnmt2*, but no detectable levels of cytosine-dependent DNA methylation (Jaber-Hijazi et al., 2013). This finding demonstrates that not all Platyhelminthes rely on DNA methylation in maintaining their physiological processes. Alternatively, it is possible that the *S. mediterranea* genome possesses some other type of DNA methylation, unrelated to 5-methylcytosine. One suitable candidate for investigation, especially in the light of

findings in *C. elegans* and *Drosophila* (Zhang et al., 2015; Greer et al., 2015) is methylation of N^6-Adenine.

1.7.2. THE NURD COMPLEX

DNA methylation is 'read' by methyl-CpG binding domain proteins (MBD), which include MeCP2, MBD1, MBD2, MBD3, MBD4 in vertebrates, the first four of which are involved in transcriptional repression (Hendrich and Bird, 1998; Jaenisch and Bird, 2003). Contrary to expectation, MBD3 is not known to bind methylated DNA (Hendrich and Bird, 1998). Both MBD2 and MBD3 are a part of the nucleosome remodelling and histone deacetylase (NuRD) complex, involved in epigenetic regulation of cell fate (Denslow and Wade, 2007; Xue et al., 1998; Zhang et al., 1999). NuRD was found to be responsible for differentiation control via chromatin-compaction-induced transcriptional repression (Crook et al., 2006). Since MBD3 is the only member of NuRD that is not known to be involved in any other complexes, MBD3 functional studies could shed light on NuRD-specific properties (Jaber-Hijazi et al., 2013).

The existence of an ancestral bilaterian MBD protein, homologous to MBD2 and MBD3, had been suggested before (Gutierrez and Sommer, 2004; Tweedie et al., 1999) and it was identified in *S. mediterranea* (Jaber-Hijazi et al., 2013). *Smed-mbd2/3* did not have conserved residues allowing it to bind methylated DNA (Jaber-Hijazi et al., 2013; Ohki et al., 2001). Functional studies demonstrated that *Smed-mbd2/3* was required for proper stem cell differentiation, without affecting self-renewal capabilities (Jaber-Hijazi et al., 2013). The authors concluded that ancestral

MBD2/3 proteins had a role in maintaining stem cell pluripotency in a manner independent of DNA methylation (Jaber-Hijazi et al., 2013).

The NuRD complex also contains HDAC1 and 2, RbAp46 and 48, MTA1/2 and Mi-2/CHD4 (Denslow and Wade, 2007; Xue et al., 1998). Planarian *HDAC-1* (*histone deacetylase 1*) was found to be expressed in neoblasts and its knockdown led to failure in stem cell maintenance (Eisenhoffer et al., 2008; Zhu and Pearson, 2013). Mi-2/CHD4 (chromodomain helicase DNA binding protein 4) is a chromatin remodelling protein, found to be required for neoblast maintenance and differentiation during regeneration and homeostasis (Scimone et al., 2010). The *D. japonica* RbAp48 (Retinoblastoma-associated protein 48) homologue was also involved in neoblast self-renewal and differentiation function (Bonuccelli et al., 2010). Mi-2/CHD4 and RbAp46 are known to be involved in other chromatin-modifying complexes, such as Sin3, CAF-1 and dMec (Kunert and Brehm, 2009; McDonel et al., 2009), meaning that their role in stem cell maintenance could be associated with these complexes (Aboobaker, 2011). The proteins' involvement in stem cell differentiation, however, seems to be related to their participation in the NuRD complex (Aboobaker, 2011). This view is supported by the fact that the only gene exclusive to the NuRD complex, *Smed-mbd2/3*, led to perturbed neoblast differentiation, but intact neoblast self-renewal following knockdown (Jaber-Hijazi et al., 2013). Furthermore, knockdown of other CAF-1 complex members – *Smed-p150, Smed-p60* and *Smed-p48* resulted in failure of neoblast maintenance (Zeng et al., 2013). CAF-1 is involved in histone deposition during S phase of the cell cycle (Smith et al., 2014), so it is possible that perturbation of this complex might lead to defects similar to *Smed-H2B*(RNAi), i.e. stem cell depletion.

1.7.3. CHROMATIN STRUCTURE

The main structural unit of chromatin is the nucleosome, first discovered by Kornberg and Thonmas (1974). The nucleosome contains roughly 150 base pairs of DNA wrapped around a protein octamer comprised of two of each of the histone proteins: H2A, H2B, H3, H4 (Kornberg and Thonmas, 1974). Chromatin organisation roughly falls under two main categories – heterochromatin and euchromatin. Heterochromatin is characterised by a condensed structure unfavourable for transcription, while euchromatin's organisation is more loose and transcription-friendly (Murakami, 2013). These states of chromatin were first described by Heitz (1928), as regions that are decondensed during interphase (euchromatin) or remain condensed and densely stained throughout the cell cycle (heterochromatin). The boundaries between heterochromatin and euchromatin are demarcated by shifts in the balance of post-translational histone modifications (Murakami, 2013). The most famous example of the consequences of crossing the boundaries between hetero- and euchromatin was described in *Drosophila*, where suppression of the red eye pigmentation gene was seen when the gene was in close proximity to (or within) a heterochromatic region (Muller and Altenburg, 1930). This observation was termed position effect variegation (PEV) and constitutes the first recorded association between transcriptional activity and chromatin state (Trojer and Reinberg, 2007).

Heterochromatin is divided into two functionally different types – facultative and constitutive. In lower eukaryote organisms, most of the genome is euchromatic, with the exception of telomeric and centromeric regions, kept constantly condensed and with minimal levels of transcription, referred to as constitutive chromatin (Grewal

and Jia, 2007). In higher eukaryotes, telomeric, centromeric, noncoding and repetitive sequences all fall under the constitutive heterochromatin definition (Craig, 2005; Grewal and Jia, 2007). The main difference between constitutive and facultative heterochromatin is that the latter retains the ability to switch between hetero- and euchromatin (Trojer and Reinberg, 2007). Facultative heterochromatin can revert back to euchromatic state depending on developmental, spatial or parental/heritable (monoallelic gene expression) conditions (Trojer and Reinberg, 2007).

One of the important proteins in heterochromatic organisation (specifically pericentric regions) is heterochromatin protein 1 (HP1) (James and Elgin, 1986). HP1 is involved in *Drosophila* PEV-dependent gene regulation (Eissenberg et al., 1990). The three mammalian isoforms – HP1-alpha, HP1-beta and HP1-gamma (Singh et al., 1991) were all found to be mainly associated with centromeric localisation, but HP1-beta and HP1-gamma could also be localized in euchromatic regions (Minc et al., 2000; Nielsen et al., 2001), possibly suggesting a transcription regulatory function.

HP1-alpha's primary binding partner is histone H3 trimethylated at lysine 9 (Fischle et al., 2003; Lachner et al., 2001). The histone methyltransferase Suv39h1 performs H3K9me3 and also binds HP1 (Aagaard et al., 1999), which led to the idea that heterochromatin maintenance might be related to a feedback loop where HP1 binds H3K9me3 and, at the same time, attracts more histone methyltrasherases (Grewal and Moazed, 2003).

The planarian genome *(S. mediterranea*) contains two HP1 orthologues named *Smed-HP1-1* and *Smed-HP1-2* respectively (Zeng et al., 2013). HP1-1 and HP1-2

localised in the nuclei and co-localised with H3K9me3. These results suggest that the tested planarian proteins were indeed HP1 homologues. Following knockdown, only *Smed-HP1-1*(RNAi) produced a regeneration phenotype, exhibiting symptoms of failure in neoblast maintenance, which was in agreement with the gene's stem cell-specific expression pattern. However, the neoblast depletion was due to increased differentiation, as demonstrated by the elevated number of cells positive for early neoblast progeny of the epithelial lineage. Since HP1 is known to be involved in heterochromatin maintenance, it was surprising to find that *Smed-HP1-1*(RNAi), apart from gene up-regulation, led to considerable gene down-regulation as well. The most substantially down-regulated genes were associated with proliferation and DNA replication (neoblasts), while the up-regulated ones included membrane proteins, proteases and metabolic proteins, mostly associated with differentiated cells. These results are consistent with SMED-HP1-1 promoting stem cell proliferation and restricting differentiation (Zeng et al., 2013). The study also demonstrated that *Smed-SSRP1* and *Smed-Spt16*, components of the FACT chromatin complex (involved with polymerase Pol II transcription fidelity (Mason and Struhl, 2003)), co-localised with *Smed-HP1-1* and led to the down- and up-regulation of similar targets after knockdown (Zeng et al., 2013). These findings raise the possibility that the HP1-1 and FACT-dependent mechanisms of neoblast proliferation and differentiation control might be (partially) functionally redundant in planarians.

1.7.4. HISTONE MODIFICATIONS

Chromatin alterations associated with transcriptional regulation depend on three main mechanisms – covalent post-translational modification (PTM) of histone N-terminal tails, histone variant replacement and physical movement of chromatin (ATP-dependent nucleosome remodeling) (Dutnall and Ramakrishnan, 1997; Henikoff et al., 2004; Luo and Dean, 1999). This section will focus on histone tail modifications. There has been a wide variety of PTMs on histone tails discovered so far: acetylation, methylation, ubiquitylation, phosphorylation, sumoylation, proline isomerisation, ADP-ribosylation, citrullination, propionylation, butyrylation and glycosylation (Chen et al., 2007; Kouzarides, 2007; Sakabe et al., 2010). An attempt was made to rationalise the complexity of these modifications into the 'histone code hypothesis', which posits that histone modifications act in concert with each other, forming specific patterns in order to lead to a particular downstream function (Strahl and Allis, 2000). 'Writer' proteins deposit a PTM on a designated histone tail residue, after which 'Reader' proteins interpret (bind to) the mark and evoke the downstream processing (for example binding of transcription factors) (Gardner et al., 2011). When the process is no longer needed, 'Eraser' proteins are recruited to the site in order to remove the PTM and discontinue the effect it had produced (Gardner et al., 2011). The same type of histone tail residue is sometimes capable of acquiring different modifications (i.e. lysine residue 27 on H3 can be acetylated of methylated) and Writers, Erasers and Readers are not always exclusive to one type of modification (Gardner et al., 2011). Furthermore, the same PTM can sometimes recruit different effector proteins (Ruthenburg et al., 2007; Sims and Reinberg, 2006). These findings have led to many researchers debating the suitability of the

term ‘histone code’ because the interpretation appears more flexible than an actual code, where one occurrence has exactly one invariable meaning (Henikoff and Shilatifard, 2011). Some researchers suggest that PTMs on chromatin should be perceived as a ‘language’ instead, since often histone modifications have effects depending on context and require a more complex evaluation (Lee et al., 2010). One of the greatest debates regarding histone tails PTMs is whether they are causative for transcriptional changes or just associated with different transcriptional states (Gardner et al., 2011). The first hypothesis assumes that histone modifications are necessary for the recruitment and assembly of the transcription machinery (Gardner et al., 2011). In support of this interpretation, TAF subunits of the TFIID transcription factor complex were found to bind to histone modifications in a direct fashion, possibly meaning that this is a requirement for transcription initiation (Jacobson et al., 2000; Vermeulen et al., 2007). The second hypothesis is that transcriptional activator/repressor complexes recruit histone modifying complexes, whose products aim to further facilitate/hinder transcription (Gardner et al., 2011). In line with this idea of the secondary importance of histone modifications, Henikoff and Shilatifard (2011) suggested that a more useful way of thinking about histone tail PTMs might be under the umbrella of the DNA accessibility paradigm. This theory poses that nucleosome positioning is the main determinant of transcriptional processes, i.e. nucleosomes act as blockage for some DNA-protein interactions crucial for transcription initiation, but, depending on their location, allow others (Weintraub and Groudine, 1976). A study that supports this interpretation was done by Dion et al. (2005). The researchers exchanged the lysine residues on H4 tail with arginine in various combinations. These manipulations resulted in changes in gene expression,

which did not depend on specific positional combinations of lysine and arginine, but rather on the total number of remaining lysines. This result is consistent with the idea that acetylation-dependent neutralisation of basic lysines leads to reduced interactions of histone tails with nucleosomes, in which case the amount of acetylated lysines matters more than their position (Henikoff and Shilatifard, 2011). The DNA accessibility paradigm could also suggest that histone tail PTMs do not recruit Reader proteins per se, but rather bridge proteins to join neighbouring nucleosomes (Henikoff and Shilatifard, 2011).

There is, most likely, no absolute truth encompassing all histone tail PTMs when it comes to their role in transcriptional regulation and chromatin organisation. An avenue that seems worthy of further explorations is their relationship with nucleosome positioning. As Next-Generation sequencing is becoming more accessible and procedures like histone modifications ChIP-seq (Chromatin immunoprecipitation sequencing) and MNase-seq (reveals nucleosome occupancy in the genome) are gaining popularity, it appears that the answers would be soon to come.

1.7.4.1. Lysine methylation

Unlike histone tail acetylation, methylation does not alter the charge of the modified histone tail residue (Gardner et al., 2011). Specialised domains within effector proteins recognise and bind methylated locations in order to (potentially) orchestrate downstream effects (Gardner et al., 2011). Such Reader domains include tudor, PHD fingers, Ankyrin repeats, PWWP, MBT, HEAT and WD40 repeats (Collins

et al., 2008; Liu et al., 2010; Taverna et al., 2007; Vezzoli et al., 2010; Wang et al., 2009b). The two most influential histone lysine methylating families are the Polycomb (PcG) and the Trithorax (trxG) groups of proteins. PcG proteins are mainly associated with maintaining the silenced state of gene promoters via trimethylating lysine residue 27 on histone H3 (H3K27), while trxG are mostly involved in maintaining a transcriptionally active state via trimethylating lysine residue 4 on histone H3 (H3K4) on promoters (Allis, 2007). These two groups of proteins are usually thought of as agents of cellular memory, i.e. preserving the cellular identity throughout challenging circumstances such as cell division (Allis, 2007).

1.7.4.1.1. Bivalent domains

One of the most intriguing discoveries in epigenetic stem cell gene regulation are the so-called bivalent domains, maintained by PcG and trxG proteins (Voigt et al., 2013). Bivalent domains are promoters characterised by large regions of H3K27me3 containing a smaller region of H3K4me3, initially identified in mouse embryonic stem cells (Bernstein et al., 2006). Developmental transcription factors are often under bivalent domain transcriptional control in ES cells, resulting in very low levels of expression. The authors argued that this state is silencing genes, but also keeping them poised for activation if necessary. They demonstrated that bivalent domains reach a resolution during differentiation and development where genes either lose H3K4me3 completely and become silenced or lose H3K27me3 and become activated (Bernstein et al., 2006). A similarly timed, but independent from the Bernstein et al.'s

(2006) study also demonstrated the existence of bivalent domains in pluripotent cell lines (Azuara et al., 2006).

One of the pioneering endeavors to expand the bivalent domain exploration to a whole-genome level using ChIP-seq, was performed by Mikkelsen et al. (2007). The study demonstrated that all CpG-rich promoters had the H3K4me3 mark in mouse ES cells. Twenty-two percent of these promoters also had the H3K27me3 mark and showed low expression levels. Mikkelsen et al. (2007) were also the first to show that promoter bivalency could also be encountered in differentiated lineages (to a much lesser degree than ES cells). The presence of bivalent promoters was subsequently demonstrated in human ES cells (Pan et al., 2007). A significant overlap was found to exist between genes normally controlled by bivalent promoters and genes pathologically affected in cancer (Bapat et al., 2010; Rodriguez et al., 2008).

Interestingly, bivalent domains do not appear to be universal throughout the animal kingdom. While zebrafish was found to have similar bivalent patterns to mammals (Vastenhouw et al., 2010), *Xenopus* and *Drosophila* lacked promoters with co-occurrence of repressive and activation-associated histone marks (Akkers et al., 2009; Schuettengruber et al., 2009). One potential reason for these findings could be that mammals have a much higher prevalence of CpG island promoters (Voigt et al., 2013).

1.7.4.1.2. Polycomb proteins

PcG proteins were first discovered in the context of HOX gene regulation in *Drosophila* (Lewis, 1978). More recently, PcG proteins were also described in X chromosome inactivation, stem cell differentiation and tumour onset and development (Lanzuolo and Orlando, 2012). PcG complexes are recruited to Polycomb response elements (PRE) on gene promoters in *Drosophila* (Allis, 2007). The search for PREs in mammals has proven difficult, but the prevalent hypothesis is that CpG islands, particularly in pluripotent stem cells, serve this purpose (Ku et al., 2008; Lynch et al., 2012). The two main Polycomb repressive complexes are PRC1 (only in *Drosophila* and vertebrates) and PRC2 (in both plants and animals) (Allis, 2007).

The PRC1 group contains proteins associated with variable histone mark recognition and modification functions like binding to H3K27me3 and RNA, as well as performing H2AK119 ubiquitination, H3K4 demethylation and H3K36 demethylation (Lanzuolo and Orlando, 2012). Interestingly, the PRC1 component RING1B is able to also navigate nucleosome compaction (Eskeland et al., 2010).

The main components of the PRC2 group include Enhancer of zeste [E(z)], Extra sex combs (Esc), Suppressor of zeste 12 [Su(z)12] and nucleosome-remodeling factor Nurf 55 (Margueron and Reinberg, 2011). PRC2 proteins are mainly involved in H3K27me2/3 recognition and deposition (Lanzuolo and Orlando, 2012; Margueron and Reinberg, 2011). An intriguing finding was that mammalian Ezh1 is capable of condensing chromatin independently of its histone methyltransferase activity (Margueron et al., 2008). This finding is reminiscent of Henikoff and Shilatifard's

(2011) suggestion that nucleosome positioning could be the primary factor in transcriptional regulation.

It is not entirely clear how PRC1 and PRC2 interact with PRE. A popular hypothesis is described by Allis (2007). According to this theory, the PcG binding to PRE depends on whether the PRE is currently transcribed or not (i.e. whether transcriptional repressor or activator proteins are bound to it). If PRE is not transcribed, it is bound by a DNA-binding protein (PHO, PHOL, GAF, etc.), which then recruits both PRC1 and PRC2 components. This leads to histone lysine methylation (H3K27me3) dependent on [E(z)]. Once the original transcriptional repressor is gone, maintenance of the repression relies on PRC1 being stably bound to H3K27me3. Thus, PRC1 performs further chromatin compaction.

The mammalian homologue of Nurf 55 is RbAp46/48 (Lanzuolo and Orlando, 2012). In planarians, as discussed previously, the *DjRbAp48* gene was implicated in stem cell maintenance (Bonuccelli et al., 2010). More planarian genes homologous to PRC2 components were also identified. *Smed-ezh*, *Smed-sz12-1*, and *Smed-eed-1* (*Embryonic ectoderm development*) were found to be expressed in neoblasts and to be required for long-term stem cell proliferation and neoblast colony expansion properties (Wagner et al., 2012).

1.7.4.1.3. Trithorax and trithorax-related proteins

The functional 'nemesis' of PcG proteins' activity is the family of trithorax proteins Allis (2007). They perform this antagonistic activity via trimethylating H3K4,

dimethylating H3K36, acetylating H3K27 and performing nucleosome-remodeling, all hallmarks of active transcription (Schuettengruber et al., 2011). TrxG family are also involved in general transcription machinery, making them more diverse than just PcG antagonists (Schuettengruber et al., 2011). Apart from maintenance of active transcriptional states, they had been implicated in ES cells self-renewal (Ang et al., 2011), fate determination and proliferation (Bagchi et al., 2007), X chromosome inactivation (Pullirsch et al., 2010), apoptosis (Tyagi and Herr, 2009), regeneration, growth plasticity and stress response-dependent senescence (Greer et al., 2010; Siebold et al., 2010). TrxG proteins can be divided in two main groups: SET domain-containing histone methyltransferases (COMPASS - complex of proteins associated with Set1, COMPASS-like, TAC1, ASH1 complexes) and ATP-dependent chromatin remodelers (SWI/SNF, ISWI, CHD1-4, CHD6-8 complexes) (Schuettengruber et al., 2011).

Of specific relevance to this thesis are the histone methyltransferase COMPASS and COMPASS-like complexes. Yeast contains only one COMPASS complex, performing H3K4 mono-, di- and trimethylation as a primary function (Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001). In *Drosophila*, COMPASS has three daughter complexes – Set1 (COMPASS), trithorax (COMPASS-like) and trithorax-related (COMPASS-like) with substantial amount of functional divergence between them (Mohan et al., 2011). Set1 was reported to perform the bulk of H3K4 di- and trimethylation on gene promoters (Mohan et al., 2011). Trithorax (Trx) was associated with H3K4me3 deposition specifically on promoters of HOX gene clusters *Antennapedia* and *Bithorax* (Chinwalla et al., 1995; Schuettengruber et al., 2009). Mammalian COMPASS-like complex MLL1 (homologous to trx) was also implicated in

the trimethylation of H3K4 of a selected number of genes, among which some HOX genes (Wu et al., 2008). The other mammalian complex homologous to *Drosophila* Trx (MLL2) wasfound to be instrumental for the establishment of bivalent promoter domains (Hu et al., 2013a).

The second *Drosophila* COMPASS-like complex (Trithorax-related) was associated with hormone (specifically ecdysone) dependent H3K4 trimethylation of gene promoters and subsequent transcriptional activation of these genes (Chauhan et al., 2012; Sedkov et al., 2003). More recently, the complex was also implicated in enhancer function (Herz et al., 2012). An interesting feature of *Drosophila* trithorax-related complex is that the core subunit is split in two – Trr (Trithorax-related) and LPT (Lost PHD-fingers of Trithorax-related), where LPT contains DNA-binding domains and Trr contains the catalytic SET domain (Mohan et al., 2011).

The mammalian homologues of Trithorax-related – MLL3 and MLL4 were also found to be involved in hormone-dependent gene activation (Lee et al., 2008), enhancer activation (Herz et al., 2012; Hu et al., 2013b) and, interestingly, in gene repression in facultative chromatin (Cheng et al., 2014). Trithorax-related, MLL3 and MLL4 complexes also contain H3K27 demethylase, which is hypothesised to relieve PcG-based repression and facilitate subsequent gene activation (Schuettengruber et al., 2011).

COMPASS and COMPASS-like proteins and complexes in yeast, *Drosophila* and mammals are summarised in **Figure 1.4**. These organisms were chosen for depiction in the figure due to the fact that trithorax and trithorax-related genes have been studied most extensively in them.

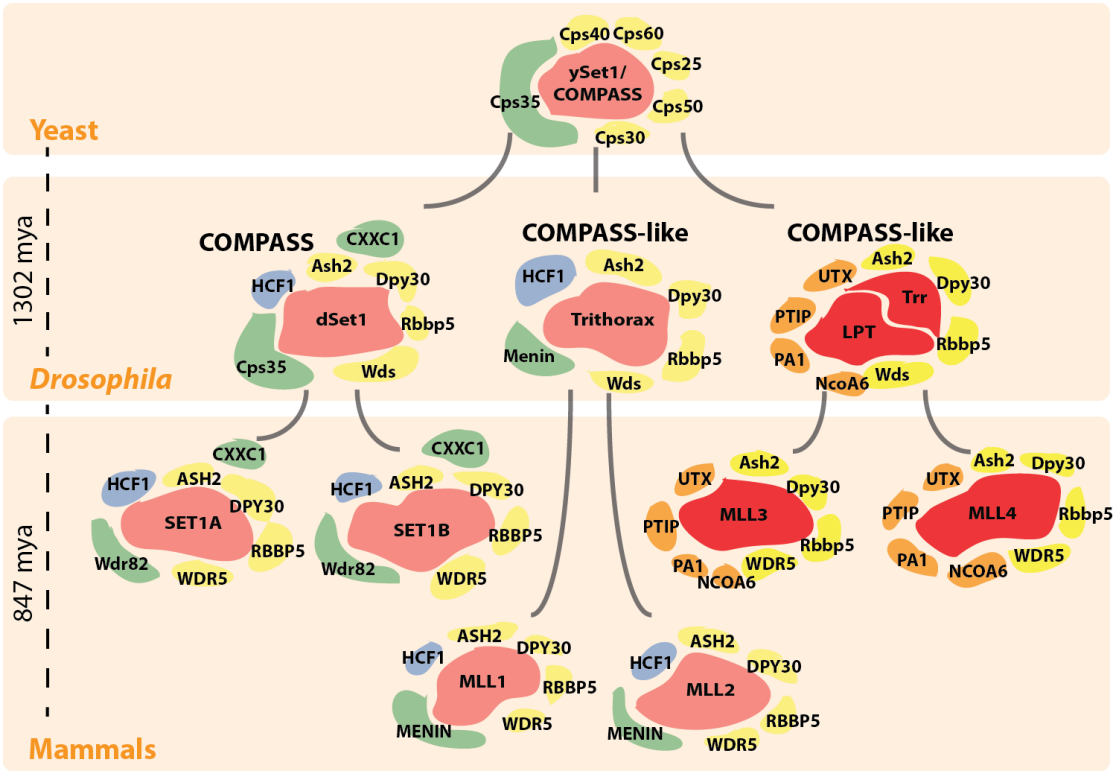


Figure 1.4. Histone-modifying TrxG complexes in yeast, flies and mammals. The ancestral yeast COMPASS complex performs all of the H3K4 methylation, while 3 specialised complexes perform this function in *Drosophila.* Each of these three complexes has two daughter complexes in mammals. The COMPASS-like complexes most relevant to this thesis are shown in bright red. The figure is modified from (Shilatifard, 2012). Years of evolutionary distance indicating that yeast and *Drosophila* shared a common ancestor 1302 million years ago (mya) and *Drosophila* and mammals diverged 847 mya are also shown (Hedges et al., 2006; 2015).

Recently, the Zayas laboratory looked for COMPASS and COMPASS-like genes in planarians. As a result, one homologue of *set1*, one of *mll1/2* (or *trx*) and two of *trr* were identified (Hubert et al., 2013). The existence of an *LPT* homologue was hypothesised, but not further investigated. Homologues for other complex subunits were also found. Genes characteristic for all three COMPASS and COMPASS-like complexes were identified in planarians – *Smed-ash2, Smed-dpy30-1, Smed-dpy30-2, Smed-rbbp5* and *Smed-wdr5* (Hubert et al., 2013). Genes coding for proteins specific to each complex were also found – *Smed-hcf1, Smed-cxxc1, Smed-wdr82-1* and *Smed*-wdr82-2 (COMPASS), *Smed-menin* and *Smed-hcf1* (trithorax COMPASS-like), *Smed-ptip* and *Smed-utx* (trithorax-related COMPASS-like) (Hubert et al., 2013). All of these planarian genes were expressed in the CNS, neoblast/neoblast progeny and gut tissues (Hubert et al., 2013). The authors reported that *Smed-set1*(RNAi) and *Smed-trr-1*(RNAi) led to difficulties with regeneration, producing smaller blastema and delaying eye formation (Hubert et al., 2013). These two genes were also found to have an effect on stem cell proliferation, with *Smed-set1*(RNAi) leading to stem cell depletion eventually (Hubert et al., 2013). Knockdown of *Smed-ash2, Smed-wdr82-1, Smed-rbbp5, Smed-wdr5, Smed-hcf1, Smed-ptip* and *Smed-utx* also resulted in smaller regenerative blastema upon amputation, accompanied with delayed eye formation (Hubert et al., 2013). Interestingly, *Smed-utx*(RNAi) exhibited a pigmentation phenotype, not seen in any of the other knockdowns in the study (Hubert et al., 2013). These findings could mean that some of the investigated genes have a role outside of the COMPASS and COMPASS-like context (Hubert et al., 2013).

1.8. THE AIM OF THIS THESIS

The general goal of this thesis is to explore the role of trithorax-related genes in planarian stem cell function. Despite its involvement in various pathologies, including cancer, the Trithorax-related complex and its mammalian orthologues (MLL3 and MLL4) have received significantly less attention than COMPASS and Trx/MLL1/2 COMPASS-like complex. This work builds on Hubert et al.'s (2013) findings, aiming to characterise in greater depth the function of trithorax-related genes.

The specific aims are to:

1) Investigate whether the planarian orthologue of *Mll3/4* is indeed split into two parts (*LPT* and *trr*), as suggested by Hubert et al. (2013); place planarian trithorax-related genes in a phylogenetic context and expand on the evolutionary history of *LPT/trr/Mll3/4* genes across phyla
2) Examine the cells and tissues in which the planarian trithorax-related genes are expressed and whether they co-localise with each other
3) Study the defects in stem cell differentiation emerging from perturbation of trithorax-related gene function
4) Characterise the outgrowth formation phenotype resultant from knockdown of the trithorax-related gene *Smed-LPT*
5) Connect trithorax-related genes' function in planarians to histone tail modifications

CHAPTER 2. IDENTIFICATION AND SEQUENCE ANALYSIS OF SMED-LPT, SMED-TRR-1 AND SMED-TRR-2

2.1. ABSTRACT

Based on existing sequence analysis of arthropods, nematodes and vertebrates, the characterisation of the evolutionary history of the *Mll3/4/LPT/trr* genes across phyla is fragmentary at best. This chapter presents a comprehensive analysis of the evolutionary history of *LPT/trr/Mll3/4* genes in an attempt to establish when and if gene fission or gene fusion had occurred. One of the main conclusions drawn was that basal species with radial symmetry had one un-split copy of the gene, which appears to have undergone independent fission events in separate protostome and deuterostome phyla. The split observed in *S. mediterranea* genome data was confirmed experimentally. Further sequence analyses of Platyhelminth species revealed that this was the only phylum that had multiple duplications of the C-terminus corresponding part of the split gene (*trr*). Additionally, the amino acid composition of the conserved domains of the LPT/TRR planarian proteins revealed that PHD-fingers in SMED-LPT could potentially bind non-methylated or monomethylated H3K4 histone residues and that the SET domains in SMED-TRR-1 and SMED-TRR-2 could possibly perform monomethyltion of the same residue.

2.2. INTRODUCTION

2.2.1. THE SPLIT MLL3/4 HOMOLOGUE

It had been well-established that the *Drosophila* homologue of the genes *Mll3* and *Mll4* is split into two parts – *LPT* and *trr* (Mohan et al., 2011). *LPT* possesses plant homeodomains (PHD-fingers), which are thought to bind chromatin (Mohan et al., 2011). This 'docking' via LPT is hypothesised to direct the H3K4 methyltransferase activity of the Trr protein and the H3K27me3 demethylase activity of UTX (Ubiquitously transcribed TPR gene on the X chromosome), proteins found in the same COMPASS-like complex as LPT (Mohan et al., 2011). This means that LPT has a role in both histone methylation and demethylation, as confirmed by functional studies in *Drosophila* by Chauhan et al. (2012). If this model is correct, then the combined function of LPT and Trr should equal that of MLL3/MLL4.

A split of one biological function into two structural units is not a novelty for the COMPASS-like family of proteins. MLL1 protein is cleaved post-translationally by a threonine protease called Taspase 1 (Hsieh et al., 2003). Taspase 1 cleaves the mature protein at two cleavage sites – 2666/2667 and 2718/1719 (Hsieh et al., 2003). This leads to the reassociation of the N- and the C-termini of the cleaved protein via the interaction between the FYRN (Phenylalanine/tyrosine-rich N-terminal) and FYRC (Phenylalanine/tyrosine-rich C-terminal) domains (Hsieh et al., 2003). Thus, the N- and C-termini form a heterodimer, which stabilises the complex (Hsieh et al., 2003). Knockdown of Taspase 1 resulted in mature MLL1 not being cleaved, which in turn led to HOX gene expression perturbation (Hsieh et al., 2003).

Since MLL1-MLL4 proteins all belong to COMPASS-like complexes, it would not be hard to imagine that the split between LPT and Trr in *Drosophila* might be related to the example of MLL1 cleavage described above. The split site between LPT and Trr in the MLL3/4 homologue, however, precedes the PHD-like zinc finger and the FYRN domain, while the proteolitic cleavage of MLL1 happens between the FYRN and FYRC domains (**Figure 2.1**).

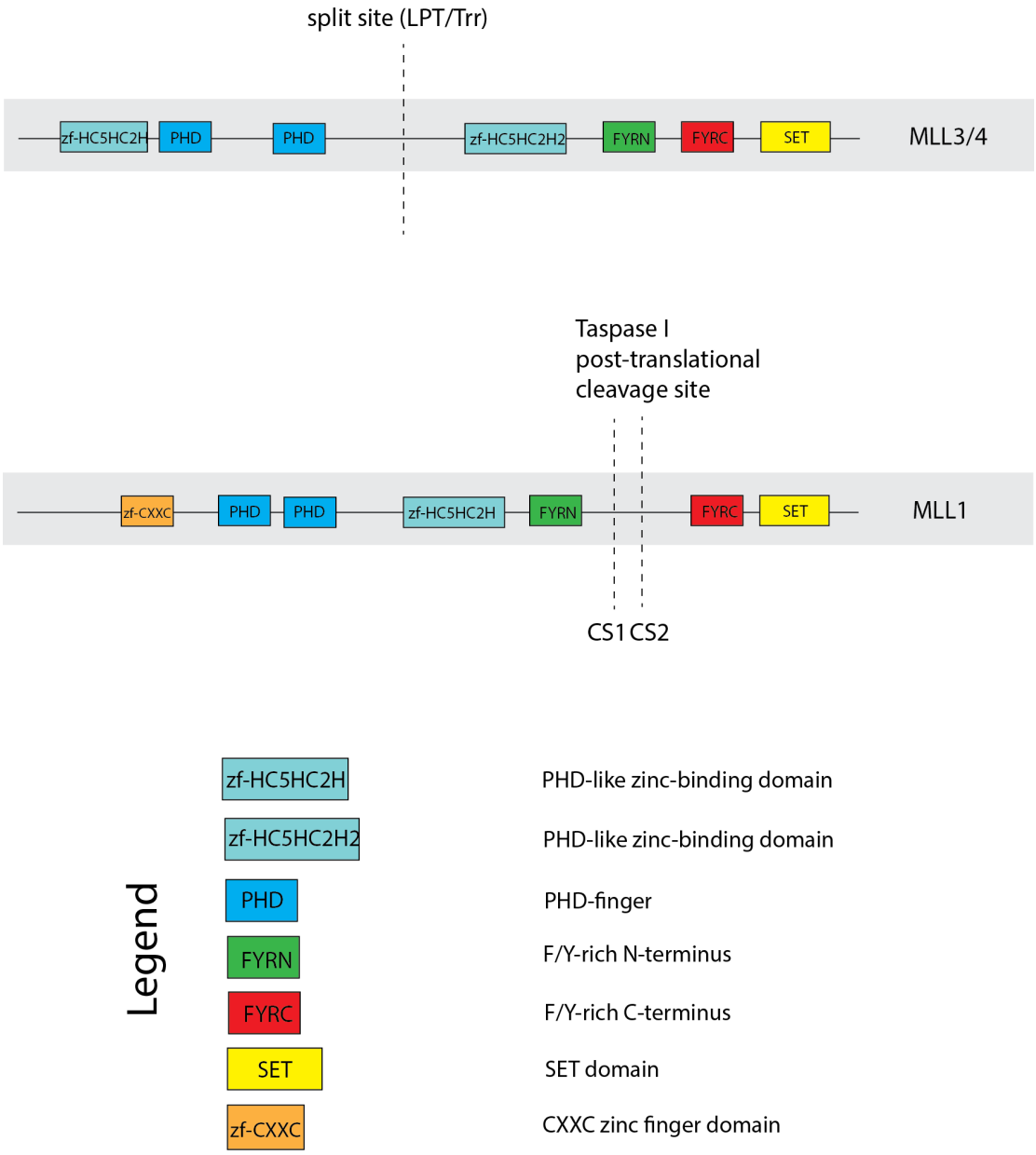


Figure 2.1. Protein cleavage of MLL1 versus *Mll3/4* gene homologue split in *Drosophila*. A comparison between the protein cleavage site in mammalian MLL1 and the split *Mll3/4* gene orthologue in *Drosophila*. 'CS1' and 'CS2" stand for cleavage site 1 and 2 respectively.

Interestingly, mammalian MLL2 and *Drosophila* Trithorax also undergo cleavage. However, they are only cleaved at CS2 since they lack CS1 (Hsieh et al., 2003). Nevertheless, the result after their cleavage and reassembly is ultimately the same as in the cleaved MLL1 – the interaction between FYRN and FYRC domains bringing the PHD-like zinc-binding domain and the SET domain closer together. Hsieh et al. (2003) speculated that this outcome mimics what we see in other COMPASS-like proteins, such as the vertebrate MLL3, the *Drosophila* Trr and the *C. elegans* SET-16 (homologous to mammalian MLL3/4). In MLL3, Trr and SET-16 the four domains at the C-terminus – PHD-like zinc-binding domain, FYRN, FYRC and SET are much closer together than in the uncleaved MLL1. The authors hypothesised that the existence of the Taspase 1-mediated MLL1, MLL2 and Trithorax cleavage is a means to achieve the potentially functionally important model of proximity of these four domains (Hsieh et al., 2003). This hypothesis is also supported by the fact that in *Drosophila* the split between LPT and Trr has left the region with the PHD-like zinc-binding, FYRN, FYRC and SET domains intact (all are located in Trr, not split between LPT and Trr). It was suggested (but not experimentally proven) that a *Drosophila*-like split in the *Mll3/4* gene homologue had happened in planarians as well (Hubert et al., 2013). The study points out that the N-terminus corresponding part of the gene was on a different contig from the two C-terminus corresponding genes (Hubert et al., 2013). This does not, however, conclusively prove that there is a split. This is one of the problems the current chapter addresses via conducting Rapid Amplification of cDNA Ends (RACE) in order to identify the full-length sequence of the N-terminus corresponding gene.

2.2.1.1. Fusion versus fission

A question worth investigating is what led to some *Mll3/4* gene homologues being split and others not. Possibly, the ancestral version of *Mll3/4* was split and underwent fusion in mammals. Alternatively, the ancestral *Mll3/4* homologue could have started off as one gene and undergone fission in different descendent species.

The most common cause for fusion or fission events are indels at the N- or C-terminus of genes (Pasek et al., 2006). Fusion/fission events have contributed substantially to the evolution of multi-domain proteins (Pasek et al., 2006). Kummerfeld and Teichmann (2005) investigated the gene fusion and fission rates across the kingdoms Archaea, Bacteria and Eukarya. They found that fusion was much more prevalent than fission; the fusion to fission ratios across the three kingdoms were 5.1, 3.9 and 4.2 respectively (Kummerfeld and Teichmann, 2005). The same study explained why that might be the case. Fusion is easier to achieve since it requires the loss of a terminal region of gene A and the loss of initial regulatory region of gene B. Fission, on the other hand, assumes the appearance of these regions, which would be much more difficult. Furthermore, the physical joining of two genes performing the same function is, most often, beneficial to the system (Marcotte et al., 1999). Kummerfeld and Teichmann (2005) also observed that, in over 70% of the examined cases, the change in gene architecture across phyla could be traced back to one common ancestor, i.e. fusions or fissions rarely happen more than once. The authors' functional analysis suggests that in the cases where gene fission/split has occurred, the two proteins either perform the same function via forming a heterodimer or fulfill consecutive steps of the same molecular process (Kummerfeld and Teichmann, 2005).

According to Mohan et al. (2011), the split between *LPT* and *trr* in *Drosophila* was the result of a single ancestral gene that had undergone fission in the orders of Diptera and Hymenoptera, rather than at the base of arthropod evolution (see **Figure 2.2**).

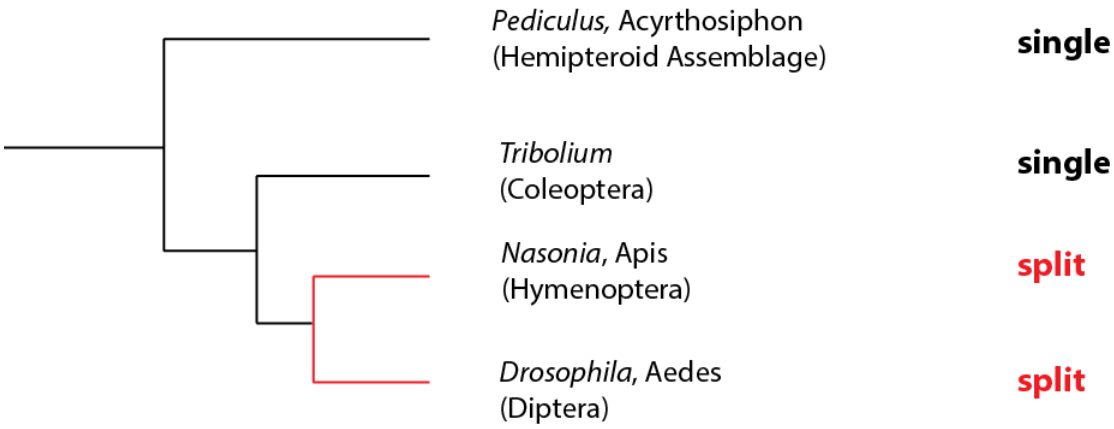


Figure 2.2. ***Mll3/4*** **gene homologue split in Diptera and Hymenoptera orders** (recreated from Mohan et al. (2011)).

Mohan et al.'s (2011) interpretation, that LPT binds to chromatin and directs the activity of Trr, fits Kummerfeld and Teichmann's (2005) conclusion that split genes often act together to fulfill a common molecular goal.

2.2.2. THE PHD-FINGER

In order to learn more about what LPT and Trr proteins do, it would be helpful to look at their domain composition. The most important domain type in LPT is the PHD-finger. It contains 60 amino acids and is a part of the C4HC3-type zinc finger family, with a notable conservation of tryptophan (W) preceding the last cysteine pair (Bienz, 2006). The PHD-finger is a Zn^{2+}-binding domain, most closely related to the E2 ligase-binding RING-finger (Bienz, 2006). The PHD-finger is a domain prevalent throughout Eukarya (Bienz, 2006). Letunic et al. (2015) and Schultz et al. (1998) identified 181 human proteins containing more than 300 PHD-fingers (Zhang, 2006). Proteins with PHD-fingers are nuclear and bind to chromatin (Aasland et al., 1995). It was speculated that they also bind to proteins (Aasland et al., 1995) and thus serve as a docking station between chromatin and the bound protein (Bienz, 2006).

PHD-fingers in ING2 (Inhibitor of Growth protein 2) and BPTF (Bromodomain and PHD-finger-containing Transcription Factor; the largest subunit of the Nucleosome remodeling factor (NURF) complex) proteins were found to have H3K4me3-binding capabilities (Peña et al., 2006; Shi et al., 2006; Li et al., 2006; Wysocka et al., 2006). Two features were related to this property in BPTF. First, a conserved tryptophan (W32) separates two binding grooves formed by the PHD-finger (Li et al., 2006; Wysocka et al., 2006). The histone tail's R2 and H3K4me3 fit into those grooves (Li et

al., 2006; Wysocka et al., 2006). Second, it was found that in BPTF protein, the PHD-finger groove that binds to H3K4me3 is an aromatic cage comprised of tyrosine residues (Y10, Y17, Y23) and the already mentioned W32 (Li et al., 2006; Wysocka et al., 2006). This is demonstrated in the crystal structures in **Figure 2.3A** and **Figure 2.3B**. The mechanism was found to be similar in the PHD-finger of ING2 protein. Again, two grooves are formed – one for H3K4me3, the other one for R2 on the histone tail (Peña et al., 2006). The two grooves are also separated by tryptophan, which binds T3 (threonine) on the histone tail (Peña et al., 2006). The groove recognising H3K4me3 contains Y (tyrosine), S (serine), M (methionine) and the groove recognising R2 contains G (glycine), C (cysteine), D (aspartate), E (glutamate), as indicated in **Figure 2.3C** and **Figure 2.3D** (Peña et al., 2006). Shi et al. (2006) discovered that ING2's PHD-finger is capable of binding H3K4me2, as well as H3K4me3.

Binding H3K4me3, however, is not a universal property of PHD-fingers – Peña et al. (2006) found only 11 proteins that contained a PHD-finger and did not belong to the ING family that had the residues necessary for binding H3K4me3.

Some PHD-fingers are capable of binding unmethylated or monomethylated H3K4. It was reported that the first PHD-finger of the autoimmune regulator (AIRE) had the strongest preference for binding H3K4me0 and exhibited weaker binding to H3K4me1, H3K4me2 and H3K4me3 (Org et al., 2008). The main difference between AIRE's first PHD finger and those of ING2 and BPTF was that in AIRE the aromatic residues needed for the formation of the groove detecting H3K4me3 were

substituted by the negatively charged aspartate (D) and hydrophobic alanine (A) (Org et al., 2008). See **Figure 2.3E** and **Figure 2.3F** for details.

In summary, the PHD-finger binds to chromatin, recognising H3K4me0/1/2/3.

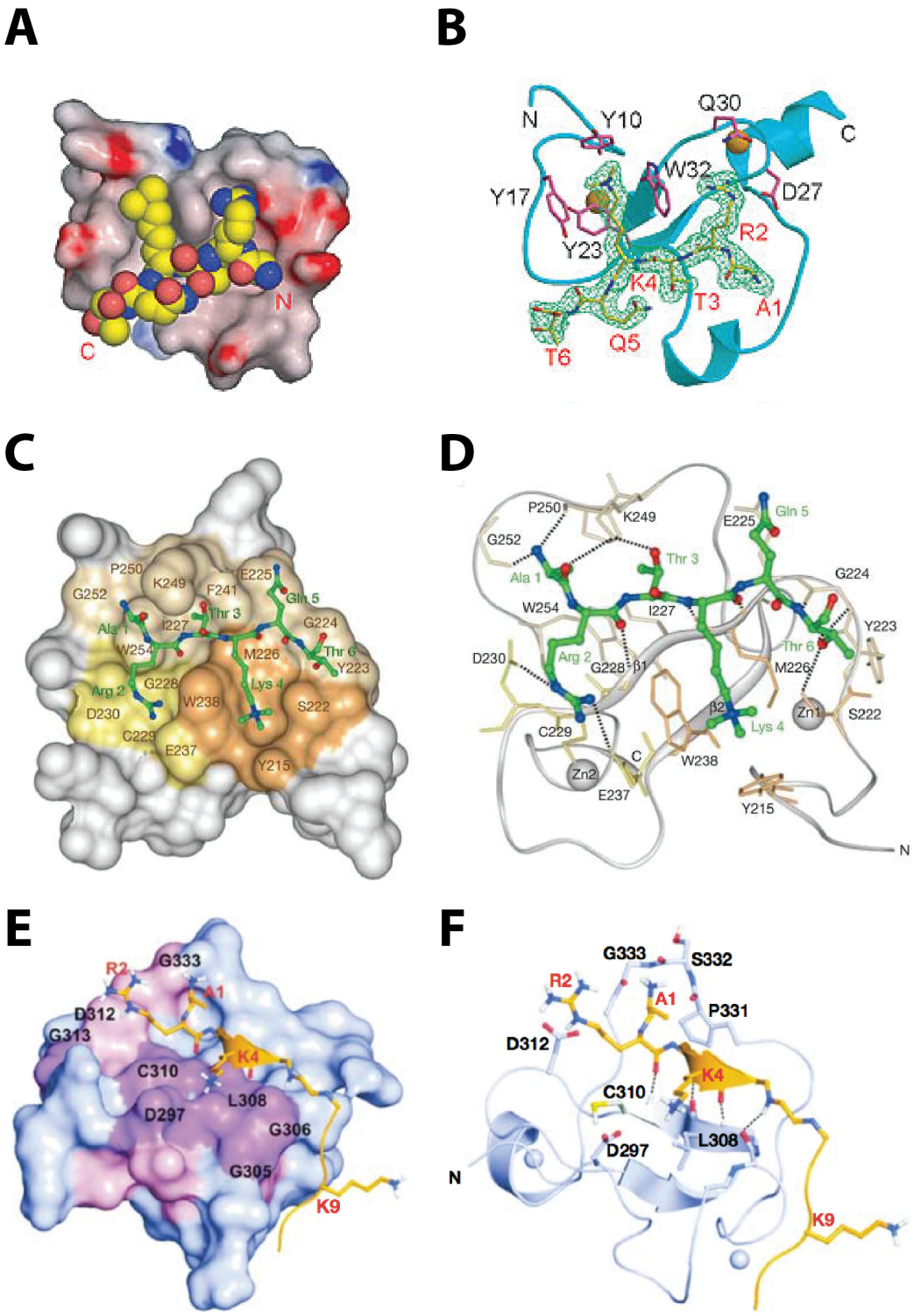


Figure 2.3. Structure of PHD-fingers. (**A**, **B**) Structure of BPTF's PHD-finger bound to H3K4me3 with red indicating positively charged surface and blue – negatively charged surface in the 3D representation. (**A**) also emphasises the formation of groves where R2 and H3K4me3 from the histone tail fit in the PHD-finger. Adapted from Li et al. (2006). (**C**, **D**) Structure of ING2's PHD-finger (grey) bound to H3K4me3 (green). Adapted from Peña et al. (2006). (**E**, **F**) AIRE's first PHD-finger (blue) bound to H3K4me0 (orange). Adapted from (Org et al., 2008).

2.2.3. THE SET DOMAIN

The other important domain in the homologues of mammalian MLL3/4 is the catalytic SET domain. The domain was first described in *Drosophila* proteins – Suppressor of variegation 3-9 (Tschiersch et al., 1994), Enhancer of zeste (Jones and Gelbart, 1993) and Trithorax (Stassen et al., 1995). The abbreviation of the name of the domain comes from these three proteins. The SET domain is found in all histone lysine methyltransferase families, except DOT1 (van Leeuwen et al., 2002; Feng et al., 2002). SET domain-containing proteins are capable of methylating K4, K27, K36 and K79 on histone H3, K20 and K59 on histone H4 (Zhang et al., 2003), as well as K26 on histone H1B (Kuzmichev et al., 2004). SET1, SET2, RIZ, SMYD, EZ, SUV39 and SUV4-20 are the main SET protein families, while SET8 and SET7/9 proteins are not classified as belonging to any of these families (Dillon et al., 2005). The MLL homologues are part of the SET1 family (Dillon et al., 2005). The catalytic activity of the SET domain consists of the transfer of a CH_3 group from S-adenosyl-L-methionine (also known as AdoMet) to the amino group of a lysine residue (Dillon et al., 2005). This results in the lysine residue being methylated and the byproduct S-adenosyl-L-homocysteine (AdoHcy) being made (Dillon et al., 2005). The chemistry of this process relies on pH-dependent deprotonation of the lysine sidechain, which produces a lone pair of electrons (Xiao et al., 2003). This is when the SET domain-containing lysine methyltransferase can add a methyl group, which makes the lysine positively charged (Xiao et al., 2003). In order to add more methyl groups, deprotonation would have to happened again (Xiao et al., 2003). A summary of this process, culminating in trimethylation of the lysine residue, is shown in **Figure 2.4:**

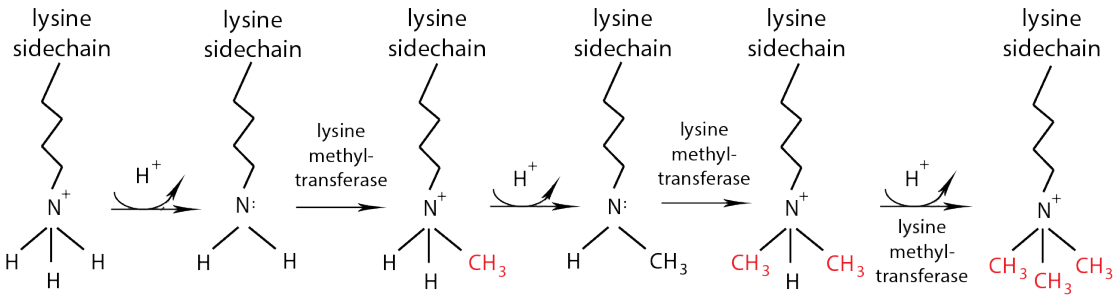


Figure 2.4. Mechanism of lysine methylation. Red indicates the subsequently added methyl groups after each round of deprotonation. Recreated from Xiao et al. (2003).

In the case where more than one methyl group is to be attached to the same lysine, the AdoHcy byproduct has to be detached from the ternary complex and replaced by AdoMet (Xiao et al., 2003). For this to happen, the C-terminus part of the methyltransferase has to partially dislodge, thus exposing the site of interest and potentially leading to the next deprotonation event (Xiao et al., 2003).

Two conserved sequence motifs are responsible for the formation of the SET domain's active site - RFINHXCXPN and ELXFDY (Dillon et al., 2005). A F/Y change in the second part of the active site (ELXF/YDY) was found to be related to the enzyme's mono-, di-, or tri-methylating ability (Collins et al., 2005b). In the human G9a H3K9 methyltransferase, a phenylalanine (F) at this position produced di- and tri-methylation, while a tyrosine (Y) resulted in mono- and di-methylation (Collins et al., 2005b).

Interestingly, some SET domain-containing proteins were also found to methylate non-histone proteins. One example of this is SET7/9's ability to methylate p53 (Chuikov et al., 2004).

To sum up, the SET domain performs a methyltransferase function on histone tails, as well as on some proteins. A F/Y switch is important for the decision whether to perform mono-, di- or trimethylation.

2.2.4. AIM OF CHAPTER

This chapter aims to investigate whether the *Mll3/4* gene homologue is indeed split in the planarian *S. mediterranea*. This question is addressed via RACE (Rapid

Amplification of cDNA Ends) technique. After the split is confirmed, the problem of whether it had originated separately from the split in *Drosophila* is addressed. This is achieved via phylogenetic comparative analysis of *Mll3/4* homologues across different clades. Furthermore, the origin of the *trr* duplication in Platyhelminthes is explored within a phylogenetic tree context. Another goal of this chapter is to assess whether the conserved domains in SMED-LPT, SMED-TRR-1 and SMED-TRR-2 are likely to be capable of binding and modifying histone tails based on sequence analyses. Finally, the presence of nuclear receptor binding motifs in planarian trithorax-related proteins is explored.

2.3. MATERIALS AND METHODS

2.3.1. REAGENTS

The reagents used for experiments in this chapter are listed in **Appendix A**.

2.3.2. SCHMIDTEA MEDITERRANEA CULTURING

Asexual *S. mediterranea* (originally from Barcelona, Spain) were kept in 1x Montjuic salts dissolved in distilled water. The recipe for the preparation of the water came from Alejandro Sánchez Alvarado's and Phillip Newmark's laboratories (Cebrià and Newmark, 2005). The components were as follows:

1.6 mM NaCl; 1 mM $CaCl_2$; 1 mM $MgSO_4$; 0.1 mM $MgCl_2$; 0.1 mM KCl;

1.2 mM $NaHCO_3$

The pH of the water was kept at 7.2-7.4. The animals were fed once or twice a week with organic calf liver. Water was changed after every feeding. The planarians were kept primarily in the dark at 20°C

2.3.3. Primers for Rapid Amplification of cDNA Fragments (RACE) and establishing the full-length sequence of Smed-LPT

Primer sequences are listed in **Appendix B**.

2.3.4. Databases used for gene homologue identification

Databases used for analyses in this chapter are listed in **Appendix C**.

2.3.5. Primer design

The primers used for the RACE experiments were designed according to the instructions in the Ambion's FirstChoice® RLM-RACE Kit (http://tools.thermofisher.com/content/sfs/manuals/cms_056070.pdf). The software Primer3Plus (Untergasser et al., 2007) was used.

For non-RACE primers (D_F and D_R), Primer3Plus was also used. Primers were 20-25 bp. GC content was >40%. Self-complementarity and cross-complementarity were avoided.

2.3.6. PHUSION PCR

For RACE and non-RACE PCR, manufacturer's instructions (NEB) were followed for reaction set-up.

The PCR cycling parameters were as follows: Initial denaturation - 98°C for 30 sec; 35-40 cycles (98°C – 12 seconds; 60-65°C – 30 seconds; 72°C – 2.5 minutes); Final extension - 72°C for 15 minutes

2.3.7. GEL EXTRACTION

The Qiagen MinElute kit was used for gel extraction and the manufacturer's guidelines were followed.

2.3.8. RNA EXTRACTION

Ten-fifteen worms were used for Trizol-based RNA isolation. Both intact and regenerating worms at different stages were present. RNA was precipitated with isopropanol and washed with Ethanol. The sample was DNase-treated and cleaned with phenol (pH 4.5)/chloroform and prefipitated with sodium acetate and ethanol.

2.3.9. cDNA PREPARATION

2.3.9.1. For 3' RACE and conventional PCR

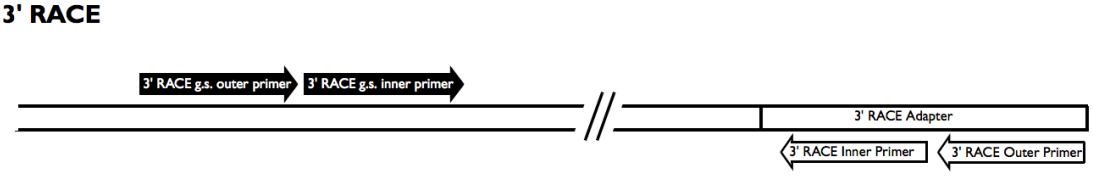
For reverse transcription, the protocol from Ambion's FirstChoice® RLM-RACE Kit (http://tools.thermofisher.com/content/sfs/manuals/cms_056070.pdf) was followed. Instead of M-MLV Reverse Transcriptase, Superscript III was used.

2.3.9.2. For 5' RACE

Ambion's FirstChoice® RLM-RACE kit manual (http://tools.thermofisher.com/content/sfs/manuals/cms_056070.pdf) was followed for the CIP (Calf Intestine Alkaline Phosphatase) treatment, TAP (Tobacco Acid Pyrophosphatase) treatment and 5' RACE Adapter ligation. The specifications for 'small rxn' were used. Random decamers, combined with 5' Outer RACE custom-made primer (Eo) were used during the Reverse Transcription.

2.3.10. 3' RACE

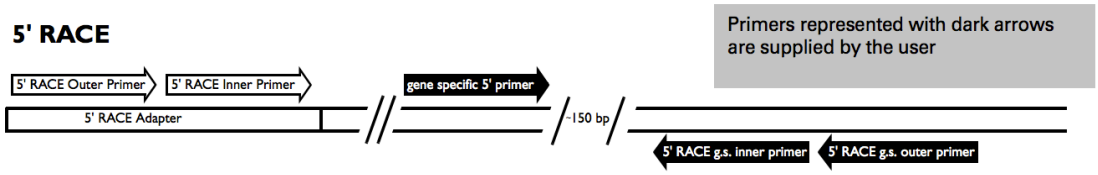
The procedure relies on a nested PCR principle. The custom-designed 3' Outer RACE primer together with the 3' Outer RACE primer in the FirstChoice® RLM-RACE Kit amplify the template used for the 3' Inner RACE reaction (performed with 3' Inner RACE kit primer and 3' Inner RACE custom-made primer). The snapshot below is taken from the FirstChoice® RLM-RACE Kit manual:



The primers in black are the custom-made primers. All PCR reactions were performed with Phusion High fidelity polymerase in a manner described earlier in this chapter. Products were gel-extracted, sequenced and analysed.

2.3.11. 5' RACE

Like 3' RACE, the procedure relies on a nested PCR principle. The custom-designed 5' Outer RACE primer together with the 5' Outer RACE primer in the FirstChoice® RLM-RACE Kit amplify the template used for the 3' Inner RACE reaction (performed with 5' Inner RACE kit primer and 5' Inner RACE custom-made primer). The snapshot below is taken from the FirstChoice® RLM-RACE Kit manual:



2.3.12. PROTEIN ALIGNMENTS AND PHYLOGENETIC TREES

All protein alignments were done using MAFFT software (Katoh and Standley, 2013) with default settings, unless otherwise stated. In the cases where the alignments were used in a figure, they were visualized via *MEGA 5.2.2* (Tamura et al., 2011). Phylogenetic trees were done by first trimming sequences according to Pfam (Finn et al., 2014) domain annotations, aligning the trimmed sequences via MAFFT and constructing the tree using *MEGA 5.2.2*.

2.4. Results and Discussion

2.4.1. Identifying Smed-LPT, Smed-trr-1 and Smed-trr-2

A tblastn algorithm with the *Homo sapiens* MLL3 (accession number: AAK00583.1) and MLL4 (accession number: NP_003473.3) proteins against three planarian transcriptomes (Kao et al., 2013; Labbé et al., 2012; Önal et al., 2012) generated identical top hits. This suggested that in planarians, *Mll3* and *Mll4* were not present as separate genes, but rather had the same homologue. One interesting observation was the fact that Transcript A (Labbé et al., 2012) hit the N-terminus of the query sequence (MLL3/MLL4), while two transcripts – Transcript B (Labbé et al., 2012) and Transcript C (Kao et al., 2013) hit the C-terminus of these proteins. It was found that the reverse complement of Transcript B corresponded to *Smed-trr1* (accession number: KC262345) and Transcript C corresponded to *Smed-trr2* (accession numbers: DN309269, HO004937), previously described in Hubert et al. (2013). Homology searches with the tblastn algorithm using *Drosophila melanogaster*'s longest Trr isoform - D (accession number: AAN09063.2) and *Drosophila melanogaster*'s LPT longest isoform - A (accession number: AAF47094.2) resulted in Trr hitting Transcripts B and C and LPT hitting Transcript A. See **Table 2.1** for summary of these results and **Appendix D** for the transcript sequences.

Upon reciprocal blastx against the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi), Transcript A hit exclusively MLL3 and MLL3-like proteins from a variety of clades, as shown in **Table 2.2.**

Queary ID	**Subject ID**	**Queary length(aa)**	**Sequence length(nt)**	**% identity**	**queary start**	**queary end**	**e-value**
MLL3 (*Homo sapeins*)	Transcript C	4911	4598	46.42	4368	4911	4.00E-144
	Transcript B	4911	4686	43.18	4398	4911	6.00E-118
	Transcript A	4911	7048	34.41	928	1379	2.00E-63
	Transcript A	4911	7048	32.44	134	497	3.00E-59
	Transcript A	4911	7048	36.08	955	1095	3.00E-23
	Transcript A	4911	7048	31.03	344	563	1.00E-19
	Transcript A	4911	7048	39.74	1630	1707	4.00E-11
MLL4 (*Homo sapiens*)	Transcript C	5537	4598	48.5	5021	5537	5.00E-148
	Transcript B	5537	4686	45.45	5028	5537	3.00E-124
	Transcript A	5537	7048	55.33	1363	1557	3.00E-50
	Transcript A	5537	7048	43.18	159	364	7.00E-

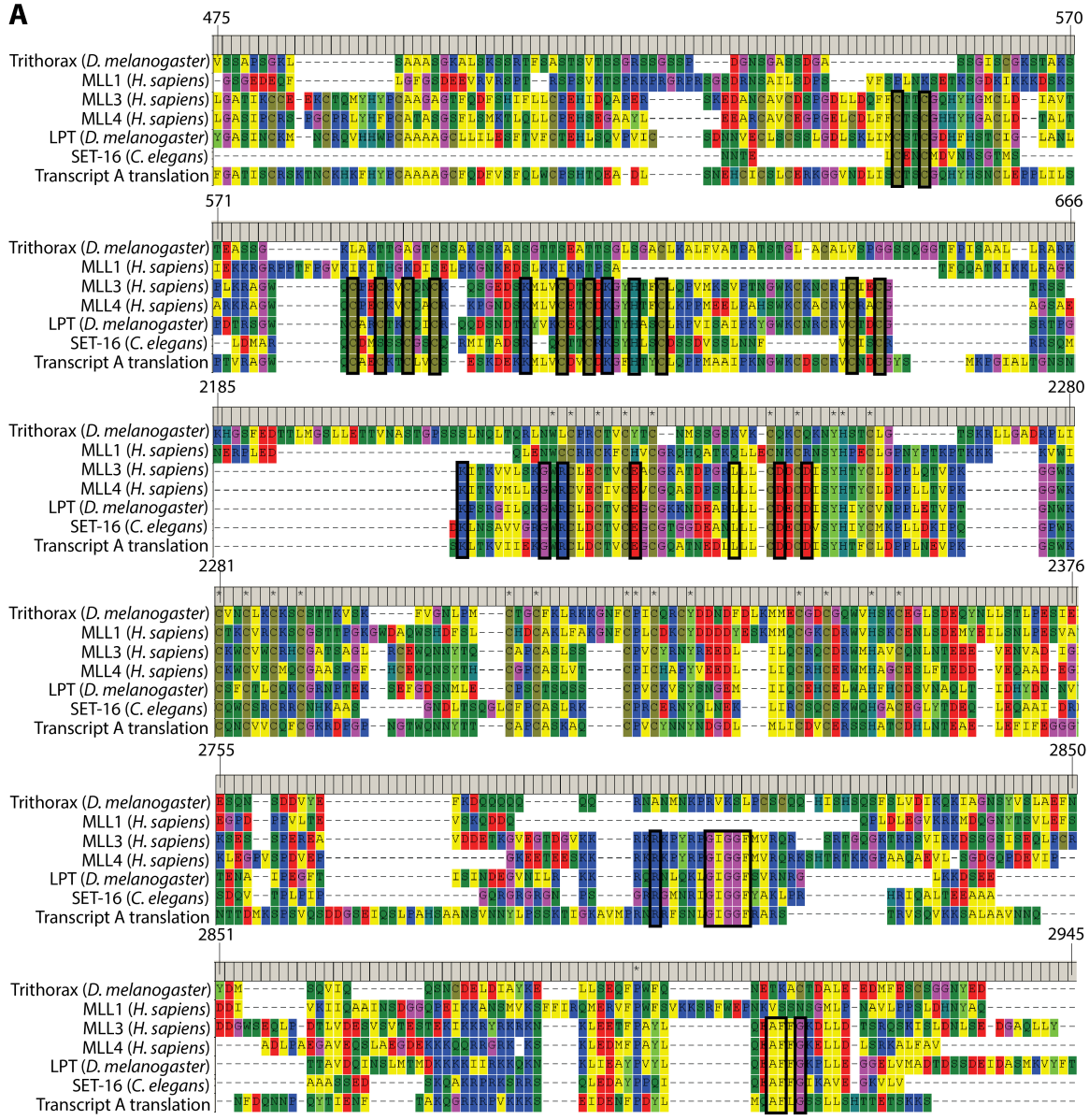
							50
	Transcript A	5537	7048	35.37	1368	1502	1.00E-20
	Transcript A	5537	7048	39.85	224	353	7.00E-16
LPT (*Drosophila melanogaster*)	Transcript A	1482	7048	35.23	439	801	8.00E-50
	Transcript A	1482	7048	39.22	111	367	9.00E-50
Trr (*Drosophila melanogaster)*	Transcript C	2441	4598	45.96	1904	2441	4.00E-142
	Transcript B	2441	4686	45.12	1908	2441	2.00E-133

Table 2.1. Tblastn with human MLL3 and MLL4 and *Drosophila* LPT and Trr proteins against planarian transcriptomes.

Hit	**E-value**
MLL3 (*Clonorchis sinensis*)	4E-114
MLL3 (*Echinococcus multilocularis*)	2E-107
MLL3 (*Hymenolepis microstoma*)	1E-94
Predicted: MLL3-like (*Limulus polyphemus*)	6E-73
MLL3 (*Crassostrea gigas*)	5E-68
Predicted: MLL3-like (*Gavia stellata*)	4E-66
MLL3 (*Struhio camelus australis*)	3E-64
Predicted: MLL3 isoform X9 (*Homo sapiens*)	2E-63

Table 2.2. Selected hits and their respective E-values after blastx with Transcript A against NCBI database.

In order to further verify that Transcript A was indeed homologous to *Mll3/4* genes, the transcript was translated into a protein sequence and aligned with other species' homologues of MLL3/4 proteins (human MLL3, MLL4, *Drosophila* LPT and *C. elegans* SET-16). Sequences from the Trithorax/MLL1 family were also included in the alignment (human MLL1 and *Drosophila* Trithorax) for comparison. Sequences were aligned using MAFFT (Katoh and Standley, 2013) alignment tool with the option E-INS-I (for sequences with multiple conserved domains and long gaps) selected. The results are shown in **Figure 2.5A**, where an asterisk marks conservation in all seven proteins and black boxes mark conservation only between proteins in the MLL3/4 family. Because of space concerns, only representative parts of the alignment are shown. A maximum likelihood phylogenetic tree with these sequences was also constructed (**Figure 2.5B**). Based on the results in **Table 2.2** and **Figure 2.5**, we can conclude that Transcript A does, in fact, share closest homology with the *Mll3/4* family of genes.


A
475
570
571
666
2185
2280
2281
2376
2755
2850
2851
2945
Trithorax (D. melanogaster)
MLL1 (H. sapiens)
MLL3 (H. sapiens)
MLL4 (H. sapiens)
LPT (D. melanogaster)
SET-16 (C. elegans)
Transcript A translation

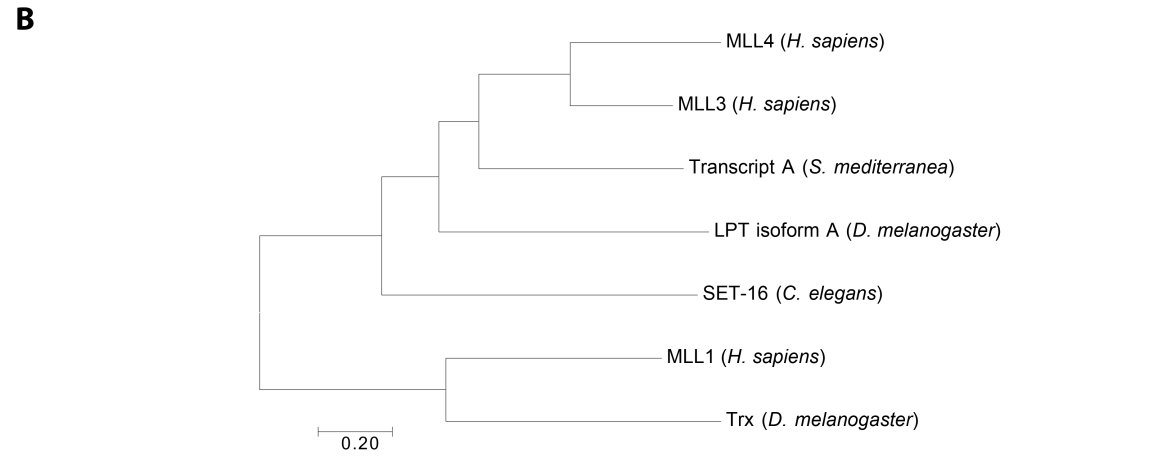

B
MLL4 (H. sapiens)
MLL3 (H. sapiens)
Transcript A (S. mediterranea)
LPT isoform A (D. melanogaster)
SET-16 (C. elegans)
MLL1 (H. sapiens)
Trx (D. melanogaster)
0.20

Figure 2.5. Sequence alignment and phylogenetic relation between the protein translation of Transcript A and proteins from the MLL1/Trithorax and MLL3/4 families. (**A**) The MAFFT alignment tool was used. Asterisks label conserved residues across all proteins, while black boxes indicate conservation only seen in the MLL3/4 family of proteins. Different amino acids are in different colours. Numbers at the top of each alignment table indicate the number of the respective amino acid in the whole protein alignment. Alignment visualisation was performed on *MEGA5.2.2* (Tamura et al., 2011). (**B**) The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-5540.1357) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 348 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2015).

2.4.2. Finding the full-length sequence of Smed-LPT

Five and 3' RACE was performed in order to get the full-length sequence of Transcript A. This was done to answer whether the gene was indeed split like in *Drosophila* (Mohan et al., 2011), as predicted by (Hubert et al., 2013), or it merged with *Smed-trr1* or *Smed-trr2*.

First, 3' RACE was performed (**Figure 2.6A**). The product of the 3' Outer RACE PCR reaction (Ao primer + 3' Outer RACE kit primer) was used as a template for the subsequent nested 3' Inner RACE PCR reactions. Due to the large size of this gene and the size limitations of Sanger sequencing, three inner PCR reactions were required (with primers Ai, B and C respectively, partnered with the 3' Inner RACE kit primer). A transcript from the Graveley transcriptome (Resch et al., 2012) – Transcript D, was predicted to overlap with the sequenced 3' Inner RACE PCR product generated with primer C and 3' inner RACE kit primer and to further extend

the product, reaching a stop codon. In order to confirm this prediction, two new primers were used – primer D_F designed in the overlapping region and primer D_R (reverse) designed immediately after the stop codon. The PCR confirmed the overlap with the RACE-d sequence and the predicted stop codon.

In order to find the start codon of Transcript A, 5' Outer and Inner RACE were performed. For the 5' Outer RACE the primer Eo and the 5' Outer RACE kit primer were used. This reaction resulted in a 400 bp product, which was used as a template for the 5' Inner RACE PCR done using the Ei and the 5' Inner RACE kit primer. This product was 367 bp, included a start codon and entered the 5' UTR. Details are shown in **Figure 2.6B**.

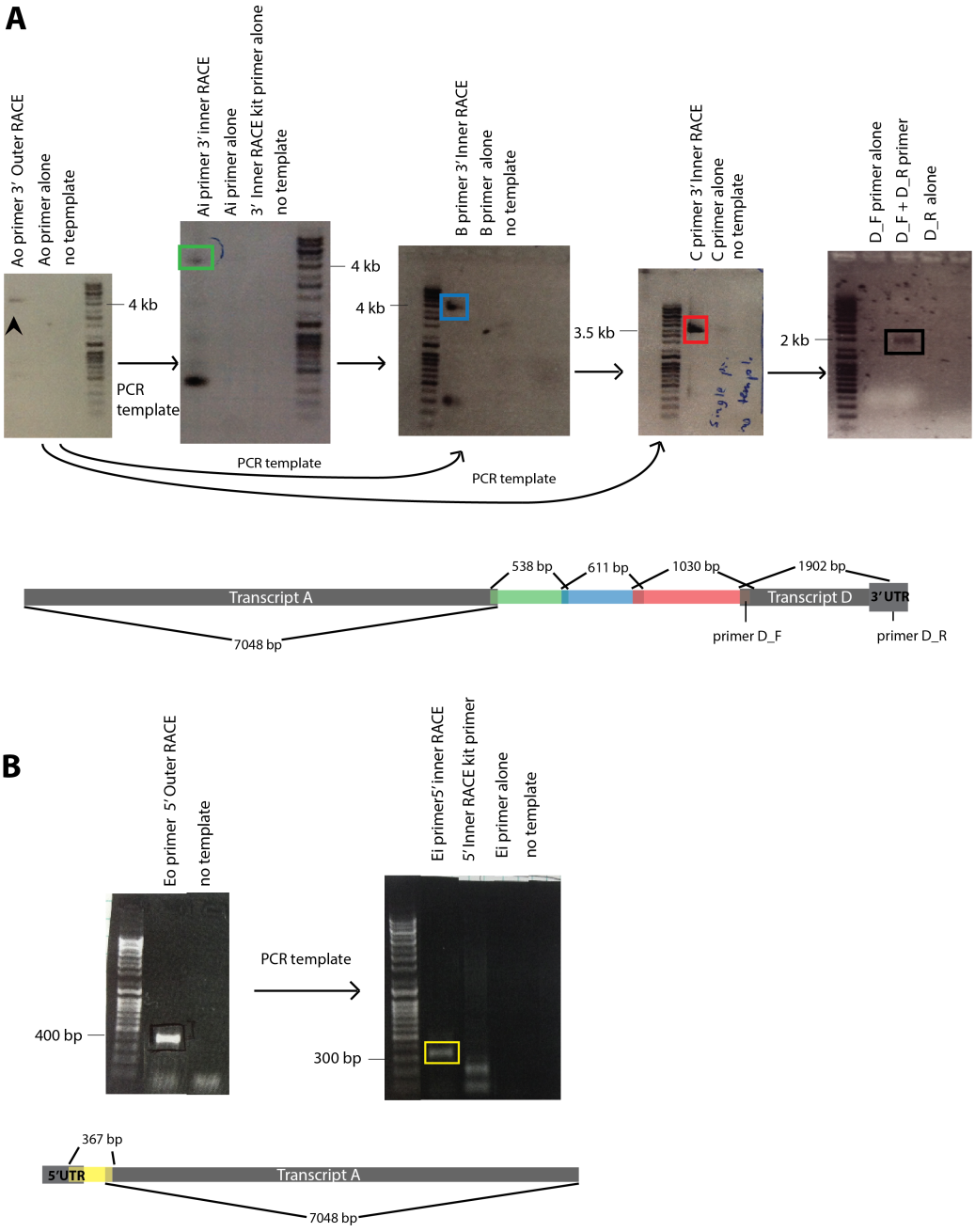


Figure 2.6. Summary of the 5' and 3' RACE of Transcript A. 3' RACE (**A**) and 5' RACE (**B**) schematic and results for Transcript A (corresponding to the N-terminus of the *Mll3/4* homologue). (**A**) 3' RACE with a series of nested Inner 3' RACE PCRs using the product of the 3' Outer RACE reaction as a template. Three nested PCRs were performed with different custom-designed primers (primers Ai, B, C). The reactions with Ai and B primers also produced smaller molecular weight bands, which upon sequencing did not match Transcript A. Transcript D overlapped with the end of the sequenced C primer + 3' Inner RACE kit primer product (red). Primers D_F and D_R confirmed the stop codon predicted by Transcript D. (**B**) 5' RACE with one nested Inner 5' RACE PCR using the product of the 5' Outer RACE reaction as a template. Custom-designed primer Ei was used together with the 5' Inner RACE kit primer. The product contained a start codon and reached the 5' UTR.

The RACE experiments resulted in a full-length sequence comprised of 11119 nucleotides (**Appendix E**). This gene was named *Smed-LPT*, since it was demonstrated that in planarians, like in *Drosophila*, the homologue of *Mll3/4* is split in 2 (*LPT* and *trr*) and *Smed-LPT* corresponds to the N-terminus of these genes. The RACE procedure did not elucidate the presence of any new domains. SMED-LPT protein contains PHD-fingers and a PHD-like zinc-binding domain towards its N-terminus, but no recognisable motifs or domains towards its C-terminus.

2.4.3. Origin of the split between LPT and TRR

Evolutionary history of *Mll3/4* homologous genes might aid putting their functional significance in context. In order to better understand the evolution of *Mll3/4* genes in animals and to trace the origin of the split in planarians, protein predictions from available genomic databases were analysed. The homology to MLL3/4 of each chosen sequence was confirmed by a reciprocal blastp against NCBI database. Sequences were only included in this analysis if they had mammalian MLL3 or MLL4 as one of their top hits. All the sequences used in this analysis are in **Appendix F**. For domain and motif detection Pfam (Finn et al., 2014) was used with the default e-value threshold setup (http://pfam.xfam.org/search). Other domain prediction softwares, like NCBI's CDD (Marchler-Bauer et al., 2015), on average, detected more protein domains than Pfam, but the evidence for the newly detected domains was not as convincing (e-value > E-05). Hence, a decision was made that Pfam's more stringent predictions were to be used.

It was already known that the ancestral yeast protein SET1 is encoded by one gene (Nislow et al., 1997). For the purposes of this analysis, however, it was necessary to find basal species that have a defined MLL3/4 homologue, rather than a ubiquitous H3K4 methyltransferase. It was found that Cnidaria and Ctenophora (belonging to the superphylum Radiata) represented by *Hydra vulgaris, Nematostella vectensis (*Cnidaria) and *Mnemiopsis leidyi* (Ctenophora), had only one gene homologous to *Mll3/4* (see **Figure 2.7***)*. The single *Mll3/4* homologue had the architecture of mammalian *Mll3/4* with PHD-fingers at the N-terminus and PHD-like zinc-binding domain (apart from *Nematostella*), FYRN, FYRC and set domains towards the C-terminus. Therefore, based on the presented findings, the superphylum Radiata seems to harbor an un-split *Mll3/4* homologue.

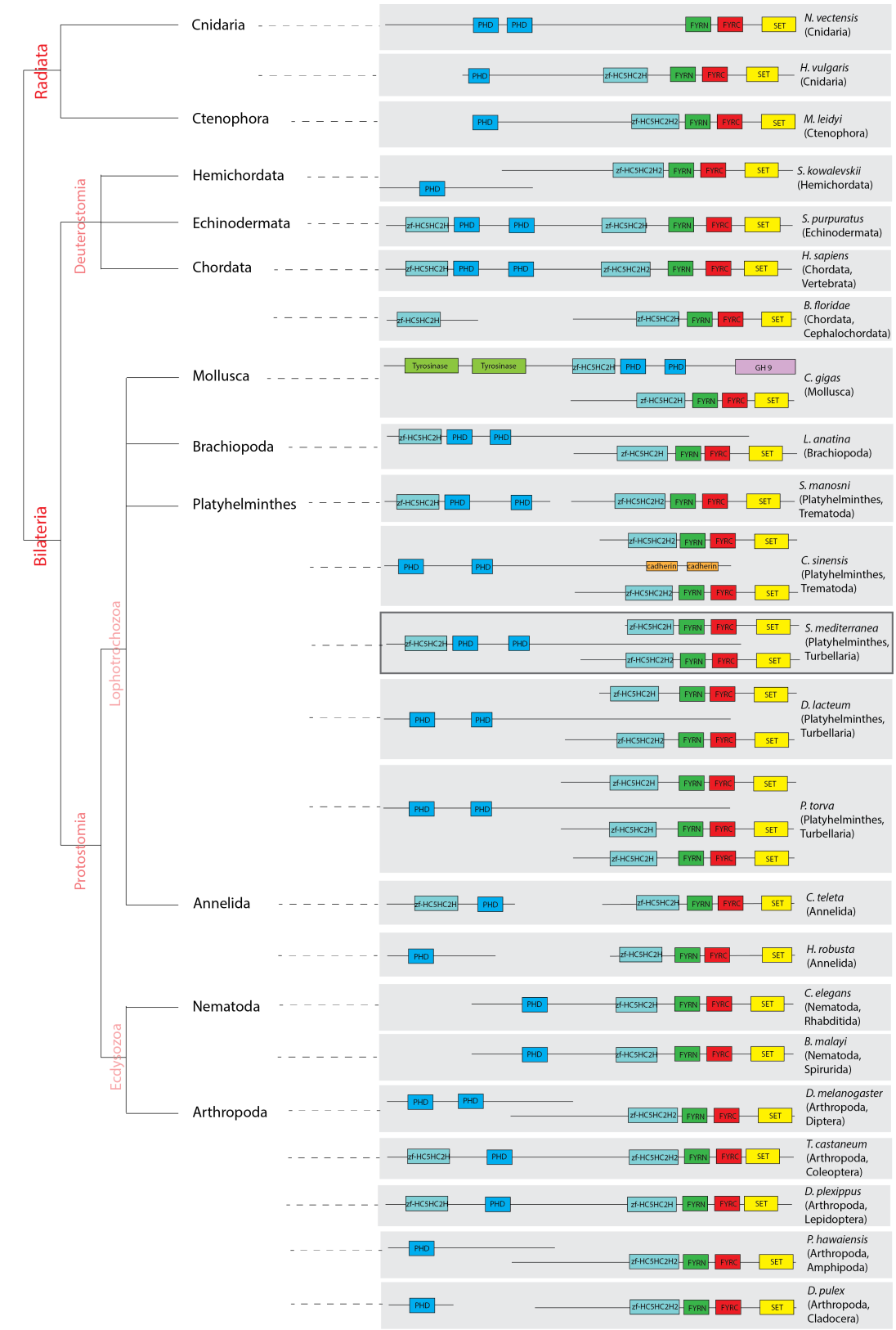
When we looked at the deuterostomes (belonging to Bilateria) the Echinodermata representative *S. purpuratus* and human (representative of Chordata Vertebrata) had one un-split *Mll3/4* gene. Another species belonging to phylum Chordata – the cephalochordate *B. floridiae*, however, had a split gene. The same pattern was observed for the hemichordate *S. kowalevskii*. These findings suggest that two independent fission events happened in deuterostomes – one in hemichordates and another one in cephalochordates.

In Ecdysozoa (protostomes), as previously mentioned, some arthropod species, like those in the orders of Diptera and Hymenoptera, had a split gene, while others, like the order Coleoptera, did not (Mohan et al., 2011). The analysis presented in this chapter demonstrated that Amphipoda and Cladocera also had a split gene, while Lepidoptera did not. Furthermore, nematodes were also found to possess an un-split

Mll3/4 gene homologue. These findings indicate that the gene fission event had affected crustaceans (*P. hawaiensis* and *D. pulex*) and some insects (*D. melanogaster*), but not others (*T. castaneum, D. plexippus*).

In the queried lophotrochozoan (protostomes) species (Platyhelminthes, molluscs, brachiopods and annelids), *Mll3/4* gene homologue was always split (**Figure 2.7**). Therefore, this split pattern is potentially basal to lophotrochozoans.

A curious observation was that some of the analysed LPT proteins had domains not usually present in MLL3/4 proteins. Examples of this phenomenon were the cadherin domains in *C. sinensis* and the common central domain of tyrosinase and glycosyl hydrolase family 9 domains in *C. gigas*. The presence of these domains might be indicative that the LPT protein is taking on some novel functions, not necessarily related to Trr.



Legend

zf-HC5HC2H	PHD-like zinc-binding domain
zf-HC5HC2H2	PHD-like zinc-binding domain
PHD	PHD-finger
FYRN	F/Y-rich N-terminus
FYRC	F/Y-rich C-terminus
SET	SET domain
cadherin	cadherin
Tyrosinase	Common central domain of tyrosinase
GH 9	Glycosyl hydrolase family 9

Figure 2.7. Split and un-split *Mll3/4* homologues across phyla. Protein predictions homologous to MLL3/4 proteins (*Homo sapiens*) across species were scanned with Pfam to identify conserved domains. The species from the superphylum Radiata have one homologous sequence. In the superphylum of Bilateria, deuterastomes have a split homologous sequence in Hemichordata and Chordata (Cephalochordata), but not in Chordata (Vertebrata) and Echinodermata. In Ecdysozoa (protostomes), the gene is split in some species like *D. melanogaster* (Arthropoda, Diptera), *P. hawaiensis* (Artropoda, Amphipoda) and *D. pulex* (Arthropoda, Cladocera), but not others like *T. castaneum* (Artropoda, Coleoptera), *D. plexippus* (Artropoda, Lepidoptera) and the nematodes *C. elegans* and *B. Malayi.* In the lophotrchozoan branch of protostomes all four studied clades (Platyhelminthes, molluscs, branchiopods and annelids) have a split gene. In some Platyhelminthes, the C-terminus part of the gene is duplicated (*S. mediterranea*, *C. sinensis, P. torva, D. lacteum*).

Overall, the gene architecture observations point to two possible interpretations of equal parsimony (illustrated in **Figure 2.8**):

1) The ancestral homologue of *Mll3/4* underwent 2 independent splits in deuterostomes – one in cephalochordates and the other in hemichordates. There were also two other independent splits in protostomes – basal to lophotrochozoans and affecting only selected arthropod species from Ecdysozoa (**Figure 2.8A**).
2) The *Mll3/4* gene homologue was split at the level of the bilateral common ancestor. From there, some deuterostomes (Hemichordata and Cephalochordata) kept it split, while others (vertebrates and echinoderms) fused it back together. In lophotrochozoans no subsequent fusion happened, while in Ecdysozoa, the gene got fused again in nematodes and some arthropods (ex: Coleoptera, Lepidoptera), but not others (Amphipoda, Cladocera, Diptera, Hymenoptera) (**Figure 2.8B**).

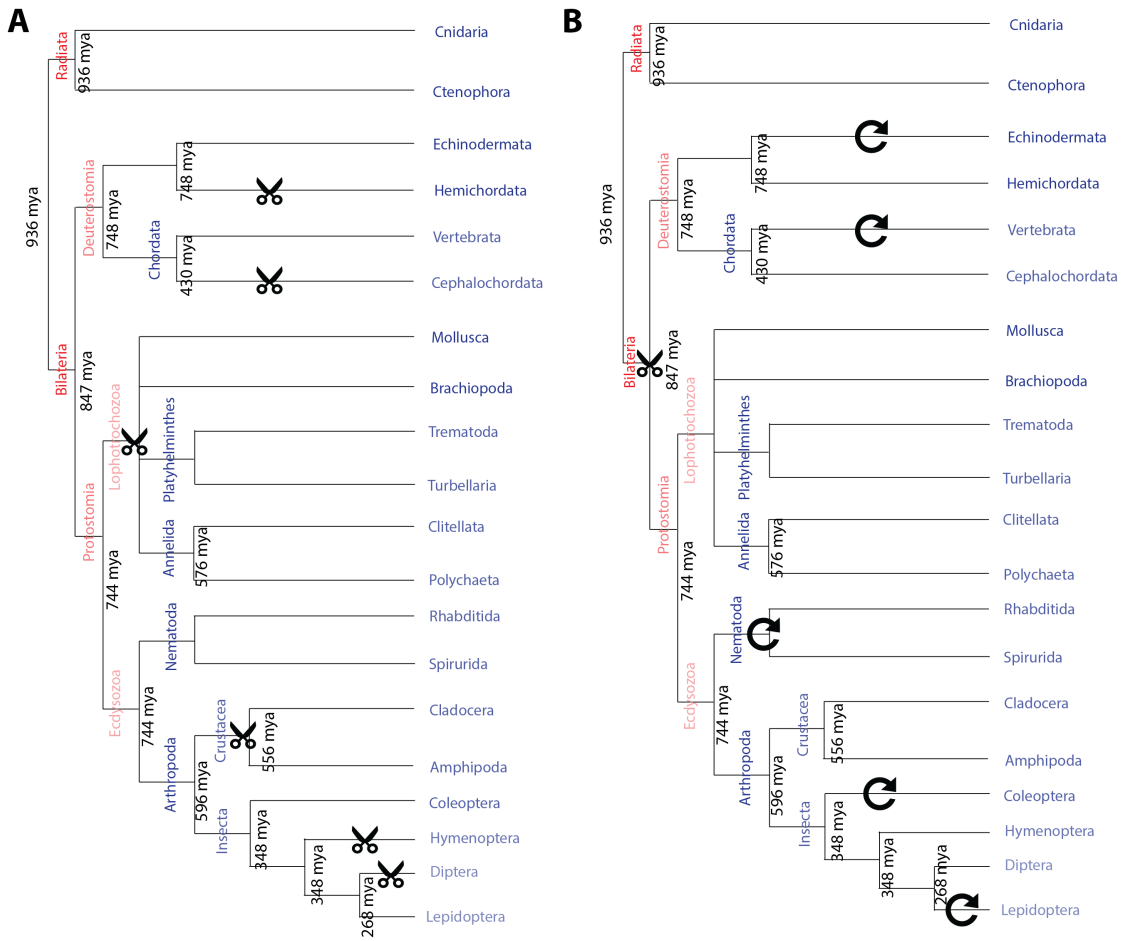


Figure 2.8. Two possible scenarios to explain the evolutionary history of *LPT/trr/Mll3/4* genes. (**A**) The gene underwent six independent splits – two in deuterostomes, one basal to Lophotrochozoa, one basal for crustacean arthropods and two in insects (Diptera and Hymenoptera). Scissors indicate gene fission event. (**B**) The gene underwent a fission event basal to all Bilateria, following which the two split genes got fused back together independently in echinoderms, vertebrates, nematodes, coleopterans and lepidopterans. Circular arrows indicate re-fusion of the split genes. This figure is not an exhaustive representation of all possible evolutionary scenarios, just the most parsimonious ones. In both (**A**) and (**B**), evolutionary distance is noted where data was available (Hedges et al., 2006; 2015). For example, Bilateria and Radiata shared a common ancestor 936 million years ago (mya).

Both scenarios involve 6 independent events. As mentioned previously, Eukarya gene fusion events are four times more likely than fission events, since fusion is less costly to perform (Kummerfeld and Teichmann, 2005). However, this holds true for first-time fusions. No such data is available on secondary fusion events (genes that had already split once). Sequence analysis in the context of a phylogenetic tree could help to resolve the issue and confirm one of the two interpretations (section **2.4.4**).

2.4.4. Further investigations of the Mll3/4 split and origin of the duplication of Trr in planarians

In order to further illuminate the evolutionary history of the *Mll3/4* homologue across species, a phylogenetic tree was constructed. For this purpose, MLL1, MLL2, MLL3, MLL4 and Trr protein orthologues from different phyla were analysed (sequences in **Appendix F**). Pfam (Finn et al., 2014) was used in order to establish domain coordinates and excise only the region from the beginning of the PHD-like zinc-binding domain to the end of the protein. This was done so that the same regions containing all the important C-terminus domains across proteins could be aligned. The phylogenetic tree resulting from this alignment is shown in **Figure 2.9.** MOZ/SAS regions from MYST acetyltransferase protein sequences were used for rooting the tree. The group of MLL1/MLL2 homologues is shown in yellow, while that of MLL3/MLL4 is shown in orange. The most ancestral MLL3/4 homologue found is that of the Ctenophore species *M. leidyi* where the gene is not split. The nematode species *B. malayi* and *C. elegans* cluster with another basal species – the Cnidarian *H.*

vulgaris. This suggests that the gene had most likely 'travelled' from the basal phylum of Cnidaria to the protostome nematodes unaffected by splitting and subsequent fusion. This hypothesis led to the conclusion that the split gene architecture occurring in some arthropods was probably a result of independent fission events, affecting specific orders in arthropods (Amphipoda, Diptera, Hymenoptera, Cladocera) and not the common ecdysozoan ancestor. Further support for this interpretation was provided by the fact that the arthropod Trr sequences did not cluster with each other, but rather, *Drosophila* Trr grouped with *T. castaneum'* and *D. plexippus'* MLL3/4, while *Parhyale hawaiensis'* and *Daphnia pulex*' Trrs were more distantly related to the cluster. This showed that the fission event was so recent in evolutionary terms that the split and un-split arthropod homologues hadn't properly diverged from each other.

Not unexpectedly, the TRR sequences (highlighted) clustered with the MLL3/4 group. Interestingly, the MLL3/4 homologue seemed to have diverged substantially, between annelid species (*C. teleta* and *H. robusta*), as illustrated by the fact that *C. teleta's* homologue was closer to molluscs and brachiopods, than to *H. robusta's*.

Molluscs, brachiopods and the annelid *C. teleta* had a Trr homologue closest to deuterostomes. Arthropods, Platyhelminthes and the annelid *H. robusta,* on the other hand, formed a cluster of their own. These seemingly unintuitive clusters could be reflective of similar functional adaptations and convergent evolution within each cluster. More specifically, the gene split could have happened in the lophotrochozoan common ancestor, but the split genes within lophotrochozoans could have diverged from each other and developed similarities with either

arthropod sequences or deuterostome sequences. The most interesting example of this were the two annelid species falling in the two separate clusters.

In summary, these observations suggest that the un-split ancestral homologue seen in Radiata species had undergone six independent splits (**Figure 2.8A**):

1) deuterostomes – affecting separately hemichordates and cephalochordates
2) basal split in Lophotrochozoa – affecting all phyla
3) basal split in arthropod crustaceans
4) independent splits in the orders of Hymenoptera and Diptera

The *trr* gene duplications probably emerged within the Platyhelminthes phylum. *S. mansoni* was the only one of the chosen flatworm species that did not have the duplication. One explanation for this could be that the duplication affected some species (*P. torva*, *S. mediterranea*, *D. lacteum*, *C. sinensis*), but not others (*S. mansoni*). Alternatively, it is possible that the *trr* duplication happened in all Platyhelminthes, but was lost in some species (*S. mansoni*). The argument that this effect might be related to differences between parasitic and free-living flatworms, does not seem plausible since both *S. mansoni* and *C. sinensis* are parasitic, yet showed divergent patterns.

The TRR-2 orthologue of the C-terminus MLL3/4 homologue of *S. mediterranea*, *P. torva*, *D. lacteum* and *C. sinensis* clustered together. Conversely, the TRR-1 orthologue of *P. torva*, *S. mediterranea*, *D. lacteum, C. sinensis* and the only TRR of *S. mansoni* formed another group. The arthropod cluster was not phylogentically closer to either of the two Platyhelminth TRR groups.

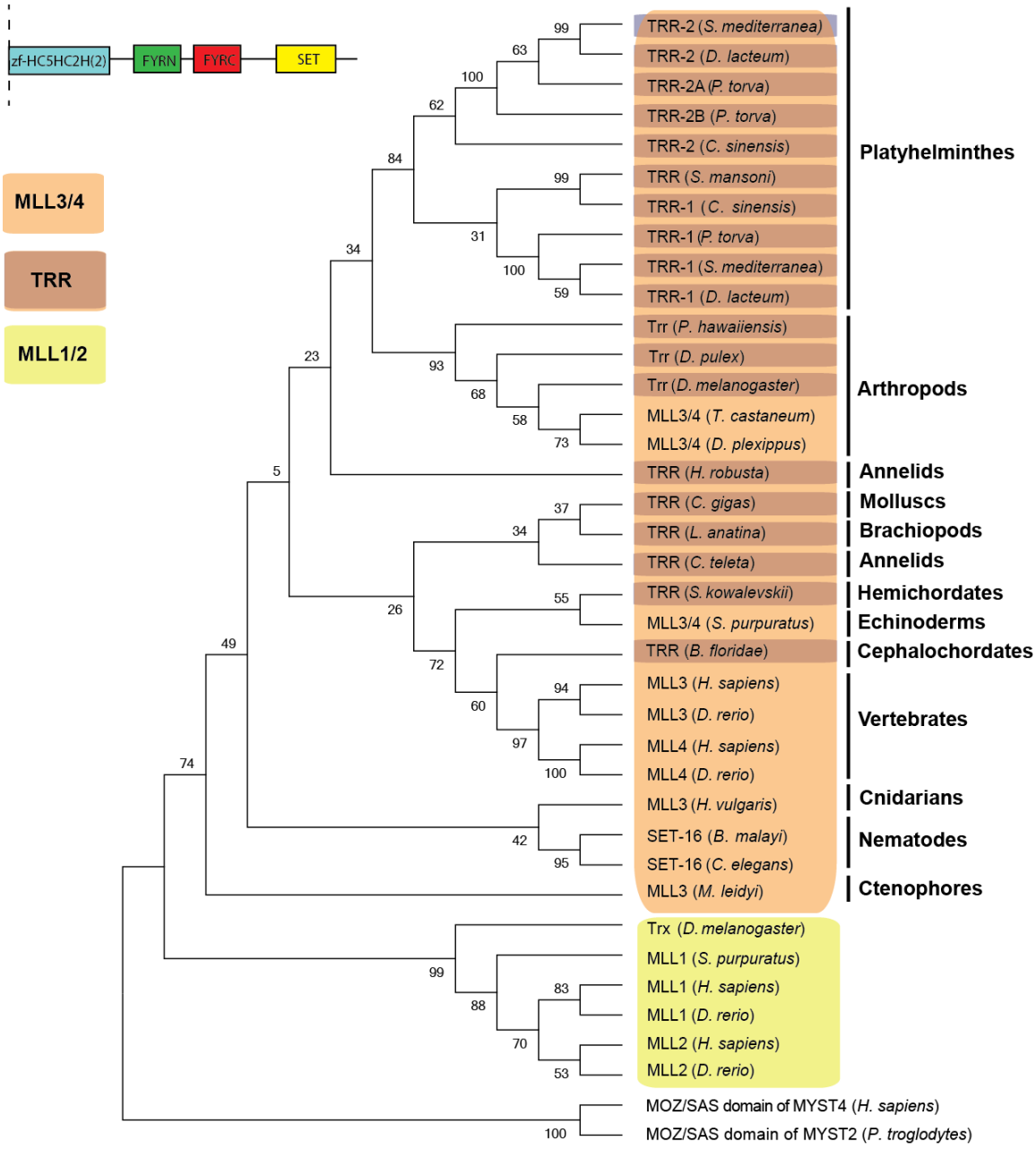

zf-HC5HC2H(2)
FYRN
FYRC
SET
MLL3/4
TRR
MLL1/2
TRR-2 (S. mediterranea)
TRR-2 (D. lacteum)
TRR-2A (P. torva)
TRR-2B (P. torva)
TRR-2 (C. sinensis)
TRR (S. mansoni)
TRR-1 (C. sinensis)
TRR-1 (P. torva)
TRR-1 (S. mediterranea)
TRR-1 (D. lacteum)
Trr (P. hawaiiensis)
Trr (D. pulex)
Trr (D. melanogaster)
MLL3/4 (T. castaneum)
MLL3/4 (D. plexippus)
TRR (H. robusta)
TRR (C. gigas)
TRR (L. anatina)
TRR (C. teleta)
TRR (S. kowalevskii)
MLL3/4 (S. purpuratus)
TRR (B. floridae)
MLL3 (H. sapiens)
MLL3 (D. rerio)
MLL4 (H. sapiens)
MLL4 (D. rerio)
MLL3 (H. vulgaris)
SET-16 (B. malayi)
SET-16 (C. elegans)
MLL3 (M. leidyi)
Trx (D. melanogaster)
MLL1 (S. purpuratus)
MLL1 (H. sapiens)
MLL1 (D. rerio)
MLL2 (H. sapiens)
MLL2 (D. rerio)
MOZ/SAS domain of MYST4 (H. sapiens)
MOZ/SAS domain of MYST2 (P. troglodytes)
Platyhelminthes
Arthropods
Annelids
Molluscs
Brachiopods
Annelids
Hemichordates
Echinoderms
Cephalochordates
Vertebrates
Cnidarians
Nematodes
Ctenophores

Figure 2.9. Phylogenetic tree with MLL3/4 and Trr proteins. Phylogenetic tree built with the C-terminus sequences (starting from the beginning of the PHD-like zinc-binding domain and finishing at the end of the protein) of MLL1, MLL2, MLL3, MLL4 and Trr homologues. The MOZ/SAS domain regions of two MYST acetyltransferase regions were used for rooting the tree. The sequences were aligned using MAFFT (Katoh and Standley, 2013) with the E-I-INS-I option chosen (recommended for sequences with multiple conserved domains and long gaps). The evolutionary history was inferred using the Maximum Likelihood method based on the Poisson correction model (Zuckerkandl and Pauling, 1965). The bootstrap consensus tree inferred from 500 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 38 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 165 positions in the final dataset. Evolutionary analyses were conducted in *MEGA 5.2.2* (Tamura et al., 2011). The orange colour shows the MLL3/MLL4/Trr group (Trr proteins are highlighted), while the yellow shows the MLL1/MLL2 group.

2.4.5. Domain conservation in SMED-LPT, SMED-TRR-1 and SMED-TRR-2

This chapter also aimed to examine whether the domains, recognised by Pfam in SMED-LPT, SMED-TRR-1 and SMED-TRR-2 proteins, were indeed conserved enough to have methyltrasferase (SET domain) and (methylated) histone-binding function (PHD-finger). To answer this question, the protein sequence of the first and second PHD-fingers of SMED-LPT were extracted and compared to the first and second PHD-finger of MLL3/MLL4 homologues in other species. The domain sequences were aligned using MAFFT software (Katoh and Standley, 2013). The alignment allowed us to conclude that the first and the second PHD-fingers of SMED-LPT protein had the characteristic for PHD-fingers C4HC3 pattern (**Figure 2.10A**). Therefore, SMED-LPT

probably had chromatin-binding properties. The two important features of a PHD-finger that is capable of binding H3K4me2/3 are a conserved tryptophan (W) between cysteine 4 (C4) and histidine (H) and aromatic acid residues forming the aromatic cage binding H3K4me2/3 (Peña et al., 2006; Shi et al., 2006; Li et al., 2006; Wysocka et al., 2006). The domain alignment in **Figure 2.10A** showed a W conservation in both the first and the second PHD-fingers. However, its location was different – before the last pair of cysteines. Furthermore, the sequences were not rich in aromatic amino acids, with the exception of the tyrosine (Y) and phenylalanine (F) found in close proximity to the conserved histidine (H). The only feature that tied SMED-LPT to being able to read H3K4me3 was the methionine (M) between the second and third conserved cysteines. A methionine at this position had been linked to H3K4me2/3 recognition in PHD-fingers (Sanchez and Zhou, 2011). On the other hand, the aspartate (D) found between the second pair of cysteines (in both PHD-fingers) was indicative of H3K4me0/1-binding capabilities (Org et al., 2008). Moreover, the second PHD-finger had a leucine (L) between the second and third conserved cysteines. This residue at this position had been associated with H3K4me0 recognition (Sanchez and Zhou, 2011).

Similar analysis was performed with the SET domain in SMED-TRR-1 and SMED-TRR-2 proteins. A high level of amino acid conservation was found between the SET domains of SMED-TRR-1 and SMED-TRR-2 and the SET domains of MLL3/4 homologues with established histone methyltransferase activity, such as SET-16 in *C. elegans* (Fisher et al., 2010) and Trr in *D. melanogaster* (Sedkov et al., 2003) (**Figure 2.10B**). This analysis indicated that the SET domain in SMED-TRR-1 and SMED-TRR-2 had all the important residues and it had, most likely, retained its methyltransferase

activity. One noteworthy observation was the presence of tyrosine residue (Y) at the end of the C-terminus of the domain. In G9a H3K9 methyltransferase, a tyrosine at this position would indicate that the enzyme performs mostly mono- and di-methylation and has very limited trimethylation abilities (Collins et al., 2005b). Given recent reports that Trr in *Drosophila* performs mostly H3K4 monomethylation (Herz et al., 2012), it seems plausible that this is the case in *S. mediterranea* as well.

The conclusions drawn regarding SMED-LPT's chromatin-binding and SMED-TRR-1 and SMED-TRR-2's methyltransferase activity fit with a previously described hypothesis (Mohan et al., 2011) where the first step in the activity of the trithorax-related complex is directing the methyltransferase activity via LPT's chromatin binding, followed by deposition of methyl groups on H3K4 by Trr.

Next, the PHD-like zinc-binding domains (zf-HC5HC2H and zf-HC5HC2H2) immediately 5' to the FYRN domain were also analysed. The pattern HC5HC2H was conserved in SMED-TRR-1 protein. In SMED-TRR-2, however, the final C was replaced by a G (**Figure 2.10C**). The structure of this domain was investigated because previous reports (Liu et al., 2014) indicated that the zf-HC5HC2H domain (known also as ePHD2) in the protein PHF6 had the capability of binding dsDNA, but not histones. This finding could mean that SMED-TRR-1 protein could, under some circumstances, have binding abilities in the absence of LPT, due to the its zf-HC5HC2H(2) domain.

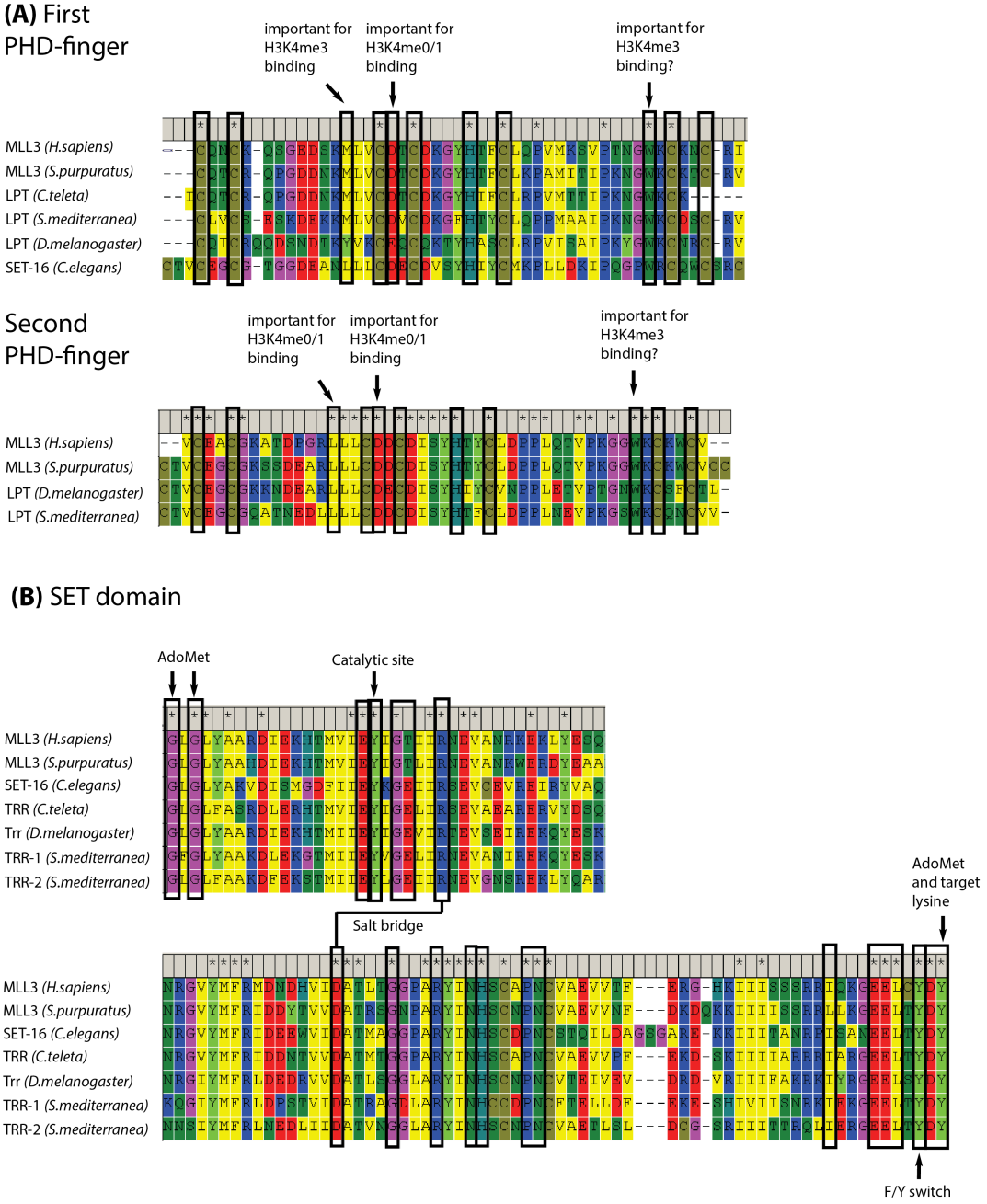

(A) First PHD-finger
important for H3K4me3 binding
important for H3K4me0/1 binding
important for H3K4me3 binding?
MLL3 (H.sapiens)
MLL3 (S.purpuratus)
LPT (C.teleta)
LPT (S.mediterranea)
LPT (D.melanogaster)
SET-16 (C.elegans)
Second PHD-finger
important for H3K4me0/1 binding
important for H3K4me0/1 binding
important for H3K4me3 binding?
MLL3 (H.sapiens)
MLL3 (S.purpuratus)
LPT (D.melanogaster)
LPT (S.mediterranea)
(B) SET domain
AdoMet
Catalytic site
MLL3 (H.sapiens)
MLL3 (S.purpuratus)
SET-16 (C.elegans)
TRR (C.teleta)
Trr (D.melanogaster)
TRR-1 (S.mediterranea)
TRR-2 (S.mediterranea)
Salt bridge
AdoMet and target lysine
F/Y switch
(C) zf-HC5HC2H/zf-HC5HC2H2 (PHD-like zinc-binding domain)
MLL3 (H.sapiens)
MLL3 (S.purpuratus)
SET-16 (C.elegans)
TRR (C.teleta)
Trr (D.melanogaster)
TRR-1 (S.mediterranea)
TRR-2 (S.mediterranea)

Figure 2.10. PHD-finger, SET domain and PHD-like zinc-binding domain conservation in trithorax-related proteins. Different amino acids are in different colours. Alignments of (**A**) the first and second PHD-fingers, (**B**) the SET domain and (**C**) the FYRN-domain-proximal PHD-like zinc-binding domain (zf-HC5HC2H or zf-HC5HC2H2) of the MLL3/MLL4 homologues across different species. Black rectangles show important structural amino acids based on Pfam (Finn et al., 2014), NCBI CDD (Marchler-Bauer et al., 2015) and literature for the respective protein domain. Asterisks indicate complete conservation of amino acids. The alignments were performed with MAFFT software and visualised with *MEGA 5.2.2* (Tamura et al., 2011).

The final analysis in this chapter addressed conservation of nuclear receptor binding motifs (NR boxes). It was reported that the hormone-dependent function of LPT-Trr in *Drosophila* was mediated via NR boxes of the LXXLL type (three leucines with two other amino acids inbetween) (Chauhan et al., 2012; Sedkov et al., 2003). Three NR boxes were found in *Drosophila* LPT and four in Trr (Chauhan et al., 2012; Sedkov et al., 2003). Sequence analysis of *Drosophila* LPT revealed that all three NR boxes were located at the C-terminus, outside of the PHD-fingers (Pfam analysis). In *Drosophila* Trr, three NR boxes were situated at the N-terminus, ouside of Pfam-recognisable protein domains, while the fourth NR box was found within the PHD-like zinc-binding domain (Pfam analysis). Planarian LPT and TRR-2 did not have any NR boxes, while TRR-1 had one (LRHLL), located between the FYRC and the SET domain (**Figure 2.11**). This finding possibly indicates that only TRR-1 or LPT-TRR-1 complex in planarians has hormone-related function with TRR-2 probably having a different role. Alternatively, due to its unusual positioning, SMED-TRR-1's NR box might be simply vestigial and not able to perform hormone receptor binding functions.

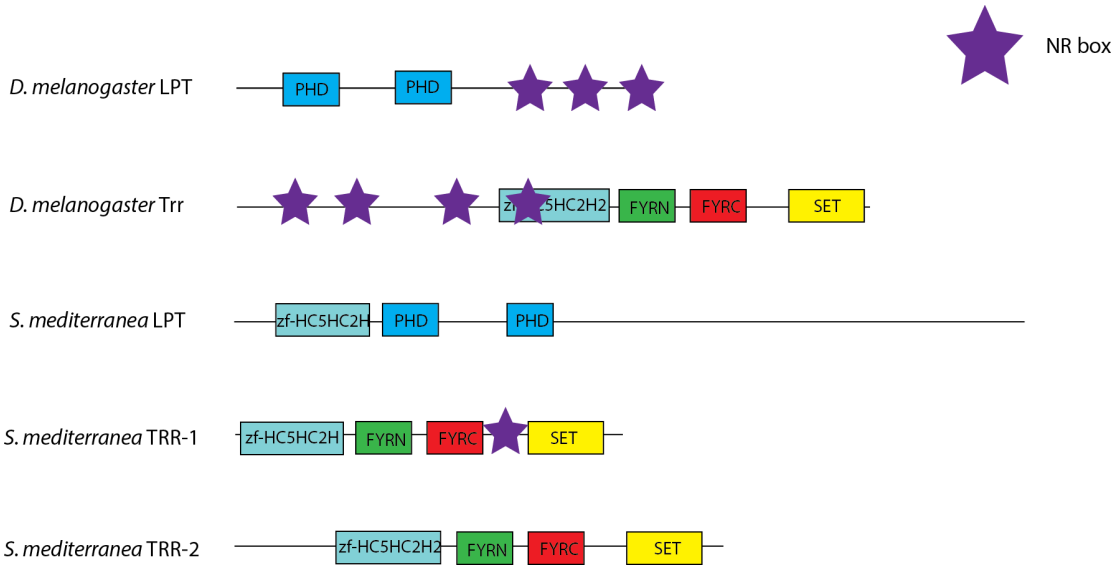


Figure 2.11. NR boxes in *D. melanogaster* and *S. mediterranea* trithorax-related proteins. Purple stars mark the position of NR boxes in *D. melanogaster* and *S. mediterranea*'s LPT and Trr proteins. PHD-fingers are in blue, PHD-like zinc-binding domain is in light blue, FYRC domain - in red, FYRN domain - in green and SET domain – in yellow.

2.5. CONCLUSION

S. mediterranea contains three genes that can be considered to be partial orthologues of the deuterostome *Mll3/4* family of genes – *Smed-LPT* is homologous to the N-terminus of the gene family, while *Smed-trr-1* and *Smed-trr-2* are homologous to the C-terminus. The presented experiments confirmed via 3' RACE that *Smed-LPT* is indeed a separate gene. Based on phylogeny, we concluded that the split between *LPT* and *trr* in *D. melanogaster* probably happened separately to that in *S. mediterranea*. This was the preferred interpretation, since the split was present in some Ecdysozoa arthropod species and in all Lophotrochozoa, but more basal species did not have the split. Additionally, it was found that Platyhelminthes are the only phyla that have a duplication of the *trr* gene. Furthermore, examination of domain architecture conservation suggests that PHD-fingers in SMED-LPT are almost certainly binding H3K4me0/1 and PHD-finger 1 is also potentially capable of binding H3K4me2/3. Analysis of the SET domains of SMED-TRR-1 and SMED-TRR-2 also confirmed that both are well conserved and probably possess full methyltransferase activity. Evidence suggests that the two proteins are involved in H3K4 mono-, rather than trimethylation. These findings confirm the speculation that SMED-LPT binds to chromatin and thus focuses the methyl group deposition of the SET-domain containing TRR(s). Furthermore, a speculation is presented that the PHD-like zinc-binding domain, invariably present at the N-terminus of the *trr* genes, might, under some circumstances, have limited DNA-binding activity. Finally, only SMED-TRR-1 possesses nuclear receptor-binding motifs, suggesting that this protein (out of the three planarian trithorax-related proteins) is the most likely candidate for regulating hormone-dependent processes.

CHAPTER 3. TRITHORAX-RELATED GENES ARE EXPRESSED IN PLANARIAN NEOBLASTS AND NEOBLAST PROGENY

3.1. ABSTRACT

Previous studies on the expression of planarian trithorax-related genes (Hubert et al., 2013) revealed that *Smed-trr-1* and *Smed-trr-2* were expressed in radiosensitive tissues, as well as CNS and gut. No data had been reported on *Smed-LPT* expression. The results presented in this chapter provide evidence that *Smed-LPT*, *Smed-trr-1* and *Smed-trr2* are expressed in over 90% of planarian stem cells, as well as in post-mitotic stem cell progeny. Trithorax-related genes are also up-regulated in the region of the regenerating brain and pharynx structures. Finally, some evidence that the three trithorax-related genes are co-expressed in the brain region is presented, confirming the hypothesis that LPT and Trr proteins interact in planarians, as they do in other animal models.

3.2. INTRODUCTION

3.2.1. MLL3/4/TRR TRANSCRIPT AND PROTEIN EXPRESSION DURING DEVELOPMENT, ADULTHOOD AND REGENERATION

Studies investigating LPT-Trr/MLL3/4 transcript and protein expression could serve to formulate an informative map of LPT-Trr/MLL3/4's involvement in embryonic development and adult tissues.

3.2.1.1. Zebrafish

One of the organisms with the most well studied development is zebrafish. Sun et al (2008) took advantage of this fact and used it as a model system to investigate expression of SET domain-containing genes during development. In zebrafish the switch between maternally produced gene products and the start of zygotic transcription in the newly conceived organism happens around 3-4 hours post-fertilisation (hpf) (Pelegri (2003); Haberle et al. (2014)). Since Sun et al. (2008) were interested in focusing on this switch period, they decided to perform *in situ* hybridisation of embryos at 0.74, 2 and 4 hpf. Among many other SET domain genes, they identified two copies of *mll3* and two copies of *mll4.* Their experiments showed that neither of the *mll3* genes was strongly expressed in the timeframe 0.75 – 4 hours post-fertilisation. *Mll3a,* however, showed low expression in the animal pole of the embryo in the region where the blastoderm was forming. *Mll4a* and *mll4b* showed a much more prominent expression in the observed developmental stages. A somewhat opposite pattern was seen for the two genes – *mll4a* was more prominent in the animal pole at 4 hpf and weaker at 0.75 and 2 hpf, while the reverse was true for *mll4b.* These temporally opposing expression patterns of *mll4a* and *mll4b*, most likely, indicate changes due to the maternal to zygotic transition. More specifically, maternally transcribed *mll4a* seems important for zygote formation, cleavage and early blastula stage, while zygote-transcribed *mll4b* has a role in the late stages of blastula (and possibly later throughout development).

Sun et al.'s (2008) study allows us to conclude that *mll4a* and *mll4b* (and potentially *mll3a*) are expressed in the blastoderm and the cells at the animal pole of the embryo preceding the formation of the blastoderm.

3.2.1.2. Mammals

In mouse development, the earliest tested time for *Mll3* transcript detection was day 8 post-coitum (dpc) (Brun et al., 2006). The authors observed *Mll3* expression throughout the embryo, with the exception of the heart and the extraembryonic amniotic membrane, and with stronger expression in the forebrain, specifically where optic placodes were to be formed. By 9.5 dpc the expression was more pronounced in the telencephalon and in the forming somites. By 11.5 dpc the expression of *Mll3* was still widely spread, with noticeable presence in the eyes. The authors commented on the eye expression going down by 13.5-14.5 dpc, which was accompanied by increasingly prominent expression in the digits. Interestingly, the retinoic acid receptor alpha (*RARalpha*) was also expressed in mouse digits in a similar pattern at the same developmental stages (Dolle et al., 1990). *Mll3* and RARalpha proteins are both associated with the COMPASS complex (Dolle et al., 1990; Goo et al., 2003), which might indicate that *Mll3*'s expression in the digits at the same time as *RARalpha* is related to retinoic acid signaling (Brun et al., 2006). To solidify this conclusion, studies showed that knockdown of *Mll3* and *Mll4* resulted in a dramatic decrease in the amount of retinoic acid-dependent H3K4 trimethylation on promoters (Lee et al., 2006a). *Mll3* and *Mll4* had also been associated with estrogen-mediated control of developmentally crucial *HOXB9* and *HOXC10* genes (Ansari et al. (2011); Ansari et al. (2012)). Tying *Mll3* to sex hormone-mediated processes makes the observation that *Mll3* was expressed in the mouse developing gonads less surprising (Brun et al., 2006). Between 12.5 and 14 dpc *Mll3* was only expressed in a subset of cells in the female and also in the male gonads. Furthermore, expression of *Mll3* was observed via RT-PCR, where some adult tissues

were also included in the analysis. *Mll3* was not detected in the adult heart and liver (Brun et al., 2006).

It would have been informative to see Brun et al. (2006) do *in situ* hybridisation experiments at developmental stages earlier than 8 dpc. Having that data, would have allowed us to make a comparison between *Mll3* expression in zebrafish blastoderm and mouse blastocyst.

In summary, the *Mll3* expression in the developing mouse is wide-spread with prominent presence in the developing brain, eyes, somites, digits and gonads. In some structures, like the eyes and the brain, *Mll3* transcript seems to only be needed (or needed more) during early organ formation. *Mll3*'s presence is encountered only in mesoderm and ecdoderm-derived organs in mouse (Brun et al., 2006).

Expression analysis of *MLL3* in human tissues revealed that the transcript was not present in the foetal liver and adult skeletal muscle, lowly expressed in adult heart, brain, lung, liver and kidney, and highly expressed in adult pancreas, peripheral blood, testis and placenta (Ruault et al., 2002). These findings indicate that *MLL3* has a role in tissues derived from all three germ layers in humans.

3.2.1.3. Drosophila

Studies in *Drosophila* provide insight into Trr protein localisation. An antibody against Trr recognised a product in nuclear extract from embryos, leading Sedkov et al. (2003) to conclude that Trr was a nuclear protein. This conclusion was supported by detecting Trr protein in the nuclei of salivary glands via immunohistochemistry. Trr

protein was also shown to co-localise with EcR-B1 (the predominant ecdysone nuclear receptor in salivary glands) in salivary gland polytene chromosomes. Furthermore, Trr showed very similar staining pattern to that of EcR-A (predominant ecdysone nuclear receptor in eye discs) and Usp (retinoid X receptor homologue) in eye imaginal discs. It is known that the steroid hormone ecdysone (induces molting) binds to EcR, which forms a heterodimer with Usp, an occurrence making both proteins indispensible for ecdysone function (Oro et al. (1990); Koelle et al. (1991)). Sedkov et al. (2003) further demonstrated that Trr bound to EcR when ecdysone was present, and trimethylated H3K4 on ecdysone-inducible promoters (discussed more in depth in **Chapter 4**). It was later demonstrated that transcription of *LPT* was increased during the early pupal stages, a time that was also marked by increased amounts of ecdysone and the beginning of metamorphosis (Chauhan et al., 2012).

Another study in *Drosophila* showed that LPT protein co-immunoprecipitated with the same complex components as did Trr (Mohan et al., 2011). This was interpreted as evidence that LPT and Trr are both part of the same complex and that Trr's methyltransferase function is performed via LPT's chromatin binding (Mohan et al., 2011). A year later, Chauhan et al. (2012) demonstrated that LPT was capable of binding Trr.

The presented so far transcript and protein expression data for MLL3/4/Trr suggest that these transcripts/proteins are expressed in a variety of differentiating tissues as well as pluripotent stem cell niches during early development. Based on the described studies, the proteins also seem to be involved in nuclear receptor-mediated cell differentiation, as indicated by *Mll3*'s association with retinoic acid

signaling and Trr's involvement in ecdysone-induced processes ultimately leading to *Drosophila* metamorphosis.

3.2.1.4. Smed-trr-1 and *Smed-trr-2* transcript expression in planarians

Hubert et al. (2013) discovered that *Smed-trr-1* and *Smed-trr-2* were expressed throughout the planarian body, with emphasis on the brain and gut. The expression of these two genes was found to be radiosensitive as demonstrated by *in situ* hybridisation and qPCR. This indicates that, at least some of *Smed-trr-1* and *Smed-trr-2*'s expression is in stem cells and/or in differentiating neoblast progeny cells.

The fact that *Smed-trr-1* and *Smed-trr-2* were found to be expressed in the regenerating brain in planarians nicely complements observations in mouse embryonic development (Brun et al., 2006). *Smed-trr-1* and *Smed-trr2* were also strongly expressed in adult brains, which is in somewhat of a contradiction with findings from human and mouse adult brains where expression was low or lacking altogether (Brun et al. (2006); Ruault et al. (2002)). This incongruence could be reconciled by the fact that every tissue in planarians, including the brain, undergoes constant cell turnover, hence, the planarian brain is under a constant state of 'development'.

Despite the fact that the expression of COMPASS-like transcripts investigated by Hubert et al. (2013) was found to be radiosensitive, this is not an unequivocal proof that they are expressed in neoblasts. Three days after lethal irradiation was chosen as a timepoint to verify whether gene expression had changed. However, it was demonstrated previously that by 3 days after lethal irradiation, it is not only the stem

cell-specific gene expression that is decreased, but also the expression in early stem cell progeny cells (Eisenhoffer et al., 2008). Hence, the claim that *Smed-trr-1* and *Smed-trr-2* are expressed in neoblasts is not very well substantiated. Furthermore, gut expression in planarian *in situ* hybridisation procedures is a common artefact of non-specific binding (personal observations). In Hubert et al. (2013) there is no image of a positive control to demonstrate that the *in situ* hybridisation procedure used is not resulting in detectable gut expression regardless of the investigated gene. Therefore, there are two contentious results in Hubert et al. (2013) – COMPASS-like genes' expression in stem cells versus stem cell early progeny and gut expression as an experimental artefact potentially masking the true extent of radiosensitivity.

3.2.2. AIM OF CHAPTER

This chapter aims to present an expression pattern for *Smed-LPT* and resolve whether trithorax-related genes' radiosensitive expression is due to their expression in stem cells, stem cell progeny or both. This problem is addressed via lethal irradiation, *Smed-H2B*(RNAi)(as a more specific way of depleting stem cells) and trithorax-related genes co-localisation with *Smed-H2B* experiments. *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2's* expression during regeneration is also considered. Finally, co-localisation experiments are used to establish whether trithorax-related genes are expressed in the same cells.

3.3. MATERIALS AND METHODS

3.3.1. REAGENTS

Reagents used for experiments in this chapter are listed in **Appendix A**.

3.3.2. PRIMERS USED

Primers used for experiments in this chapter are listed in **Appendix B**.

3.3.3. GENE CLONES USED

Gene clones used for experiments in this chapter are listed in **Appendix G**.

3.3.4. PRIMER DESIGN FOR RNA PROBES AND DSRNA

As for the RACE PCR primer design described in the previous chapter, Primer3Plus (Untergasser et al., 2007) was used. The following parameters were set: primer length – 22 bp – 26 bp; minimum Tm was set to 55°C, optimal – 57°C and maximum – 60°C; minimum GC content in a primer was 40%. If possible, the CG clamp value (the number of consecutive Gs and Cs at the 3' end of the primer) was set to 2. 3' complementarity and potential hairpin formation was avoided. All amplicons were in the range 300-800 bp.

If the amplified product was not to be cloned in a plasmid, 5' universal T7 linker (GGCCGCGG) was added to the start of the forward primer and 3' universal T7 linker

(GCCCCGGC) was added to the start of the reverse primer. These linkers are part of the 5' universal T7 primer and the 3' universal T7 primer.

3.3.5. PCR

For routine PCR gene amplification Sigma Aldrich's REDTaq® ReadyMix™ PCR Reaction Mix was used as per manufacturer's instructions. The standardised settings for PCR amplification were as follows:

Initial denaturation – 94°C for 5 minutes; 35-40 cycles (94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1-1.5 minutes); Final extension - 72°C for 10 minutes.

3.3.6. CLONING OF GENES

TA cloning with two preferred vector systems, pGEM®-T Easy (Promega) and pCR II (Invitrogen), was done. In each case, the manufacturer's instructions were followed. The competent cells used were DH5α™ (Invitrogen).

3.3.7. RNA PROBE SYNTHESIS

Total RNA extraction and cDNA synthesis were performed as described in the previous results chapter's methods section. Where possible, the primers for the region of interest for synthesis of RNA probe were designed so that conserved domains can be avoided. This way cross-detection of other genes with similar domains was less likely.

3.3.7.1. RNA probes from PCR

The gene of interest was amplified via RedTaq ReadyMix PCR with primers that have T7 linkers. The product was gel-extracted and used for a 'secondary' PCR where, for the synthesis of a template for anti-sense transcription, the gene-specific forward primer was used together with 3' T7 universal primer. This way the RNA polymerase binding site was added to the template in a position suitable for later RNA anti-sense transcription. *In vitro* RNA transcription with T7 RNA Polymerase was performed. RNA probes were either Digoxigenin(DIG)-labeled or Fluorescein(FITC)-labeled. Probes were precipitated with Ethanol and lithium chloride. After centrifugation, the pellet was resuspended in a solution containing equal volumes of formamide and TE pH 7.5 (+0.001% Tween-20). Probes were quantified via Nanodrop.

3.3.7.2. RNA probe from plasmid

Products of interest were cloned into either pGEM®-T Easy (Promega) or pCR II (Invitrogen) vectors. In order to generate template for synthesis of anti-sense RNA probe, amplification (using plasmid as a template) with gene-specific forward primer and either M13 forward or M13 reverse primer, depending on the orientation of the insert in the plasmid, was performed. If the template was amplified with gene-specific forward and M13 forward primers, T7 RNA polymerase was used to set up transcription (the M13 forward site of both pGEM®-T Easy and pCR II vectors has a binding site for T7 polymerase). Alternatively, if the template was amplified with gene-specific forward and M13 reverse primers, SP6 RNA polymerase was used to set up transcription (the M13 reverse site of both pGEM®-T Easy and pCR II vectors

has a binding site for SP6 polymerase). In either case, RNA was DNase-treated and precipitated with Ethanol and lithium chloride. After centrifugation, the pellet was resuspended in a solution containing equal volumes of formamide and TE pH 7.5 (+0.001% Tween-20). Probes were quantified via nanodrop.

3.3.8. dsRNA SYNTHESIS

The fragment of interest was cloned into a vector. RedTaq ReadyMix PCR reaction, using the cloned fragment as a template, was set up with M13 forward and M13 reverse primers. This resulted in the amplification of the fragment of interest with the site for T7 RNA polymerase binding at one end (M13 forward) and the site for SP6 RNA polymerase binding at the other end (M13 reverse). The purified PCR product was used to set up 2 RNA transcription reactions – one with T7 and another one with SP6. RNA was DNase-treated and precipitated with Ethanol and sodium acetate. After centrifugation, the pellet was resuspended in PCR-grade water and quantified with a Nanodrop and 1% agarose gel.

3.3.9. RNA INTERFERENCE

The dsRNA was diluted to 2 μg/μl. The dsRNA was administered via injections on a Nanoject II apparatus (Drummond Scientific) with 3.5'' Drummond Scientific (Harvard Apparatus) glass capillaries pulled into fine needles on a Flaming/Brown Micropipette Puller (Patterson Scientific). The needle was was filled with mineral oil and dsRNA. Worms were placed on a piece of black card in a petri dish of flattened

ice (to reduce movement). They were turned with their ventral side pointing upward. The needle was inserted in the middle of the anterior gut branch and, depending on the size of the worm, 96.6 to 128.8 nl were injected in each worm on each injection day.

The parameters described in Solana et al. (2012) were followed for *Smed-H2B*(RNAi). The dsRNA was prepared from a pCR II plasmid containing a *Smed-H2B* fragment. Injections were done on three consecutive days and animals were processed five days after the last injection.

For other gene knockdowns mentioned throughout this thesis, the same set up was used, the only difference being the injection schedules. For all knockdowns *GFP*(RNAi) was used as a control. GFP dsRNA was always injected under the same conditions as the experimental condition.

3.3.10. GAMMA RADIATION

The irradiation machine used was GSR D1 cabinet irradiator, supplied by RPS Service. The lethal irradiation dose used was 40 Gy.

3.3.11. IN SITU HYBRIDISATION PROTOCOLS

3.3.11.1. Carnoy-based *in situ* hybridisation

This method was preferred for colorimetric development and some single fluorescent *in situ* hybridisation procedures. For every experiment using this protocol

throughout the thesis, 10 animals per condition were processed, unless otherwise specified. Protocol was followed as described in Jaber-Hijazi et al. (2013).

3.3.11.2. King and Newmark (2013) *in situ* hybridisation protocol

This protocol was preferred for most single fluorescent and all double fluorescent *in situ* hybridisation procedures. An extremely detailed protocol with notes on reagent brands, solution preparation and different steps can be found in the Supplemental data (http://www.biomedcentral.com/content/supplementary/1471-213x-13-8-s1.pdf) of the Newmark planarian group's article (King and Newmark, 2013).

In order to detect *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* transcripts in a fluorescent *in situ* hybridisation (less sensitive that colorimetric procedure), a pooled probe detection method was used. This means that two probes (for each of the *Smed-LPT* and *Smed-trr-1* transcripts) or three probes (for *Smed-trr-2*) detecting different region from each of the three transcripts were synthesised in order to lower the detection threshold for each gene. This technique had been previously described (van Wolfswinkel et al., 2014).

For every experiment using this protocol throughout this thesis, 10 animals per condition were processed, unless otherwise specified.

3.3.12. Preparation of stock fluor-tyramide solutions

Three different fluor-conjugated NHS esters (TAMRA, DyLight 633 and DyLight 488) were used for fluorescent gene detection. The stocks were prepared as described in King and Newmark (2013).

3.3.13. Imaging

All colorimetric *in situ* hybridisation images were taken on Zeiss Discovery V8 (Carl Zeiss) microscope with a Canon EOS 600D or Canon EOS 1200D camera. All fluorescent images were taken on either Inverted Olympus FV1000 or FV1200 Confocal systems. In most cases, the images were taken as a z-stack and flattened afterwards on Fiji (Schindelin et al., 2012). Other times, only representative slices from the stacks were taken for figures.

3.4. Results and Discussion

3.4.1. Trithorax-related genes are expressed in CNS, gut and stem cell progeny

Sense and anti-sense RNA probes were synthesised and tested in an *in situ* hybridisation procedure (**Figure 3.1**). The fact that there was staining in the anti-sense probe condition, but not in the sense probe, clearly demonstrated that the whole-body expression pattern of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* was specific.



Figure 3.1. Expression patterns of trithorax-related genes in planarians. *In situ* hybridisation on wildtype animals with anti-sense and sense RNA probes for *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2*.

In order to assess whether *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* are expressed in stem cells and stem cell progeny, an experiment was performed where half of the animals were subjected to a lethal dose (40 Gy) of gamma irradiation and the other half were not. It is known that stem cell gene expression is completely abolished 2 days following a lethal irradiation procedure and that some stem cell progeny gene expression is also affected at this stage (Eisenhoffer et al., 2008). The expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* was found to be radiosensitive (**Figure 3.2A**). This finding agrees with Hubert et al. (2013). Following irradiation, *Smed-LPT*'s expression was only detectable in the cephalic ganglia, the gut region and, faintly, in the pharynx. This result means that the decreased detection of this gene is due to its partial expression in stem cells and/or early stem cell progeny. *Smed-trr-1* was the gene that showed the greatest radiosensitivity out of the trithorax-related genes under investigation. Following irradiation, only CNS and pharyngeal expression was detected. This contradicts partially Hubert at al.'s (2013) findings that *Smed-trr-1* is expressed in the gut, as well as neoblasts and CNS. In Hubert et al.'s (2013) work, no positive control was shown for the optimal performance of the *in situ* hybridisation procedure. In the present experiment, both a gene known to be expressed in the gut (*Smed-porcupine-1*) and a gene known to be expressed in stem cells, but not the gut (*Smed-H2B*) showed their respective expected expression patterns. Hence, the current *in situ* hybridisation procedure is working well, which suggests that the gut expression detected for *Smed-trr-1* in Hubert et al. (2013) was an experimental artifact.

Smed-trr-2, on the other hand, was the least radiosensitive gene, with substantial expression remaining in the gut, pharynx and CNS following radiation. Two

independent experiments confirmed this. The gene *Smed-porcupine-1* expressed in the gut tissue (Gurley et al., 2008) was used as a negative control in its role of radiation insensitive gene, while *Smed-H2B* (Guo et al., 2006) was used as a positive control.

Expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* was also studied during regeneration. For this purpose, ten worms per condition were cut and processed on the third day of regeneration, while another ten worms per condition were cut, lethally irradiated (40 Gy) 24 hours afterwards and fixed on the third day of regeneration (and second day after lethal irradiation). This experiment aimed to better illustrate the genes' expression in terminally differentiated (radio-insensitive) cells during regeneration. Trithorax-related genes were expressed in the regenerating brain and the pharynx forming in the remodeling tissue (**Figure 3.2B**). Hubert et al. (2013) also observed *Smed-trr-1* and *Smed-trr-2* being expressed in the regenerating brain, while there was no mention of the regenerating pharynx. Interestingly, the brain and pharynx expression in the case of all three genes was more obvious before, rather than after irradiation. This finding indicates that *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* are expressed in CNS and pharynx progenitor cells (radiosensitive), as well as terminally differentiated cells belonging to those structures. These findings are in line with Brun et al.'s (2006) observations that *mll3* is expressed in the developing brain of mouse embryos. The presented results also confirm Hubert et al.'s (2013) findings of *Smed-trr-1* and *Smed-trr-2's* expression in the regenerating planarian brain.

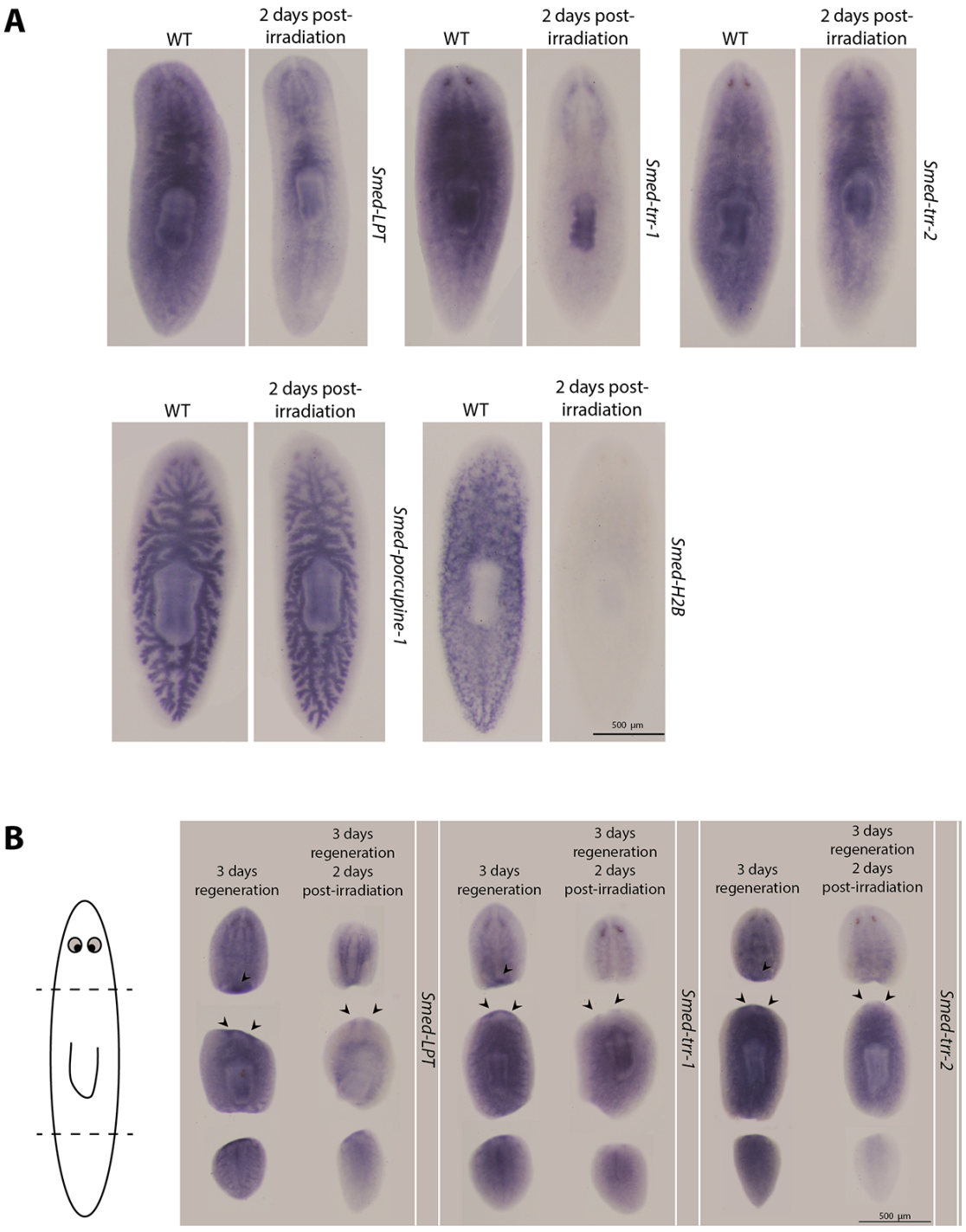


Figure 3.2. Expression of trithorax-related genes is radiosensitive. Expression pattern of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* following lethal irradiation. (**A**) Homeostatic wildtype (WT) animals are compared to homeostatic animals 2 days after 40 Gy gamma radiation exposure. *Smed-porcupine-1* is a negative control and *Smed-H2B* - a positive control for the procedure. The image is representative of 2 separate experiments using 10 worms per condition each. (**B**) Expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* is compared between three-day wildtype regenerates and three-day regenerates two days after 40 Gy gamma radiation exposure. Arrows indicate regenerating brain and pharynx. Ten worms per condition were used.

3.4.2. STEM CELLS EXPRESS TRITHORAX-RELATED GENES

The alternative, more specific way of testing whether a gene is expressed in planarian stem cells is via *Smed-H2B* knockdown (Solana et al., 2012). Ten worms per condition underwent control *GFP* knockdown and another ten worms per condition – *Smed-H2B* knockdown. The expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* was assessed on the fifth day following the last dsRNA injection, as described in Solana et al. (2012). *Smedwi-2* (Reddien et al., 2005b) and *Smed-porcupine-1* were used as a positive and negative control for the knockdown efficiency. In the case of *Smedwi-2*, expression only remained in the CNS after *Smed-H2B* knockdown, while *Smed-porcupine-1* was unchanged. As can be seen in **Figure 3.3A**, the effect of *Smed-H2B*(RNAi) on the expression of the trithorax-related genes was much weaker than that of irradiation. This finding confirms that the result we observed in **Figure 3.2A** was partially due to irradiation's effect on stem cell progeny, as well as stem cells. There was, however, still a decrease in *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2*'s expression following *Smed-H2B*(RNAi) treatment. As seen with irradiation, the effect was strongest for *Smed-trr-1*, weaker for *Smed-LPT* and the weakest for *Smed-trr-2*. In *Smed-trr-2*'s case, the reduction in expression following *Smed-H2B*(RNAi) was mostly visible in the posterior part of the animal. The observations stated here were reproducible over the course of two separate experiments. These findings mostly agree with predictions made on the basis of a compiled (by Dr. Damian Kao) dataset of three transcriptomic studies on FACS-sorted planarian cells (Önal et al., (2012); Labbé et al. (2012); Resch et al. (2012)). The predictions for the expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* in the three distinguishable by FACS cell populations are illustrated in **Figure 3.3B**. As expected, *Smed-trr-1* had the highest percentage of

its expression in stem cells and early stem cell progeny (x1+x2=67%) out of the three genes of interest. Next was *Smed-trr-2* with x1+x2 = 62%. *Smed-trr-2* had a higher percentage of expression in x2 compared to *Smed-trr1*. In the light of the post-irradiation and post-*Smed-H2B*(RNAi) expression studies, this could indicate that *Smed-trr-1*'s expression in x2 is mostly in G1 stem cells, rather than early stem cell progeny (supported by the fact that *Smed-trr-1*'s expression was drastically decreased after both neoblast-ablating treatments). *Smed-trr-2*'s expression in x2, on the other hand, could be due to mostly neoblast progeny expression in that fraction (supported by the fact that *Smed-trr-2*'s expression was influenced much more following irradiation than *Smed-H2B* knockdown). What these speculations suggest, is that *Smed-trr-1*'s neoblast expression is closer to 67% (x1+x2 estimate), while Smed-trr-2's neoblast expression is closer to 38% (x1 alone estimate).

The FACS-transcriptome predictions also fit the results from the irradiation and *Smed-H2B*(RNAi) studies for *Smed-LPT*, indicating that, at the very least, 28% (x1 estimate) of *Smed-LPT*'s expression is in stem cells.

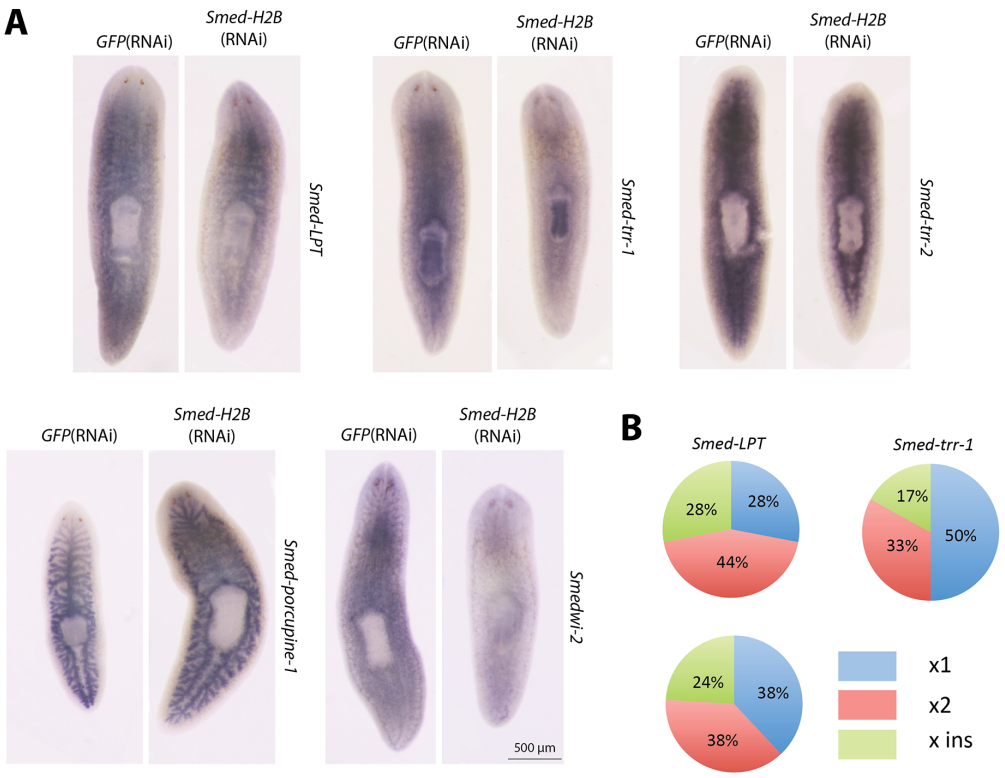


Figure 3.3. Trithorax-related genes are down-regulated following *Smed-H2B*(RNAi) and are expressed in x1, x2 and x ins FACS-sorted populations. Stem cell-specific expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2*. (**A**) Expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* in *GFP*(RNAi) and *Smed-H2B*(RNAi) animals five days after the last dsRNA injection. *Smed-porcupine-1* and *Smedwi-2* are negative and positive controls respectively. Images are representative of 2 experiments using 10 worms per condition each. (**B**) Pie charts showing estimated percentage of total RNA expression of *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* in x1, x2 and x ins FACS populations (combined data from (Resch et al. (2012); Önal et al. (2012); Labbé et al. (2012)).

In order to confirm that trithorax-related genes are indeed expressed in stem cells, a double *in situ* hybridisation was performed to investigate whether *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* co-localised with *Smed-H2B* (**Figure 3.4**). Results showed that 91% of all *Smed-H2B*-positive cells counted were *Smed-LPT*-positive, 97% were *Smed-trr-1*-positive and 92% were *Smed-trr-2*-positive. Ten animals were processed per condition, six worms were chosen at random from each condition and cells in the imaged area were counted.

The expression of trithorax-related genes in planarian stem cells nicely complements previous findings of zebrafish *mll3a*, *mll4a* and *mll4b* expression in the embryonic blastoderm (Sun et al., 2008).

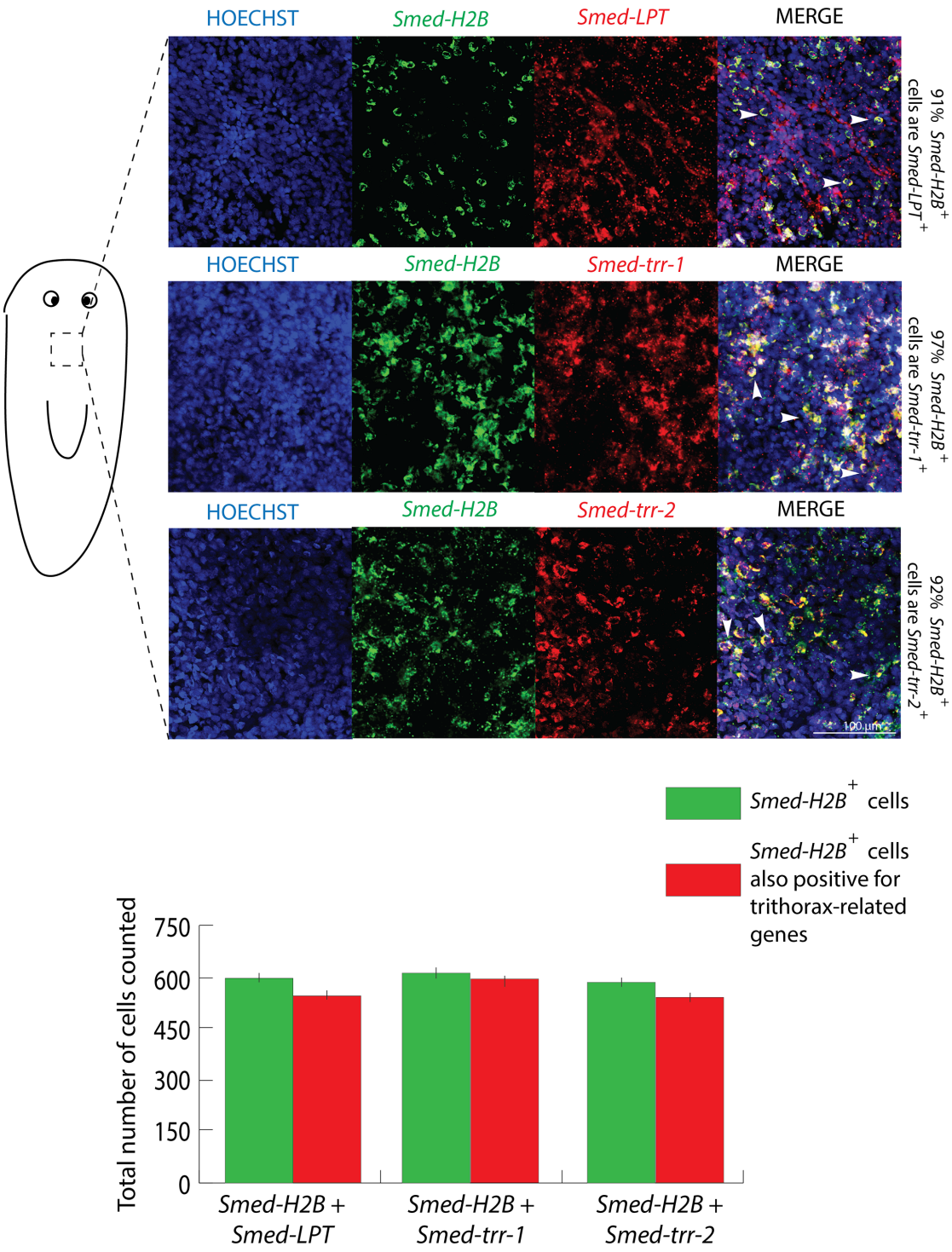


Figure 3.4. Most neoblasts express trithorax-related genes. Co-localisation between genes from the trithorax-related family in *S. mediterranea* (for which pooled probes, detecting different regions of the genes, were used) and *Smed-H2B*. Pre-pharyngeal region of the animal was imaged. Ten worms per condition were used. The bar graph shows the raw numbers of counted cells, based on which percentages were calculated. Error bars represent SEM. 92% of all *Smed-H2B*-positive cells are *Smed-LPT*-positive, 97% are *Smed-trr-1*-positive and 92% are *Smed-trr-2*-positive. White arrows show examples of double positive cells.

To summarise the results presented so far in this chapter, not only are trithorax-related genes expressed in stem cells and stem cell progeny, but also almost every stem cell expresses a trithorax-related gene.

It would be exciting to achieve better resolution regarding trithorax-related genes' expression in stem cells, i.e. whether they are expressed in the sigma population of neoblasts (cNeoblasts) and/or in committed neoblasts like the zeta population (van Wolfswinkel et al., 2014) (sigma and zeta populations of neoblasts are discussed in detail in the **Chapter 1**). Such experiments would inform claims of planarian trithorax-related genes being involved in pluripotent stem cell maintenance and/or commitment decisions. Due to technical limitations, at the time of writing this chapter, co-localisation experiments between sigma/zeta genes and trithorax-related genes were not possible.

3.4.3. TRITHORAX-RELATED GENES ARE EXPRESSED IN THE SAME CELLS

We have established that *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* have differences in their expression patterns. The final experiment in this chapter aimed to elucidate whether trithorax-related genes are, at least partially, expressed in the same cells. Examples of cells that were found to co-express *Smed-LPT* and *Smed-trr-1*, *Smed-LPT* and *Smed-trr-2*, *Smed-trr-1* and *Smed-trr-2* are presented in **Figure 3.5**. These findings agree with Mohan et al.'s (2011) discovery that LPT protein co-immunoprecipitates with Trr in *Drosophila* and, hence, support their hypothesis that LPT and Trr proteins work together. Therefore, it seems likely that in planarians, *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* are also capable of working together.

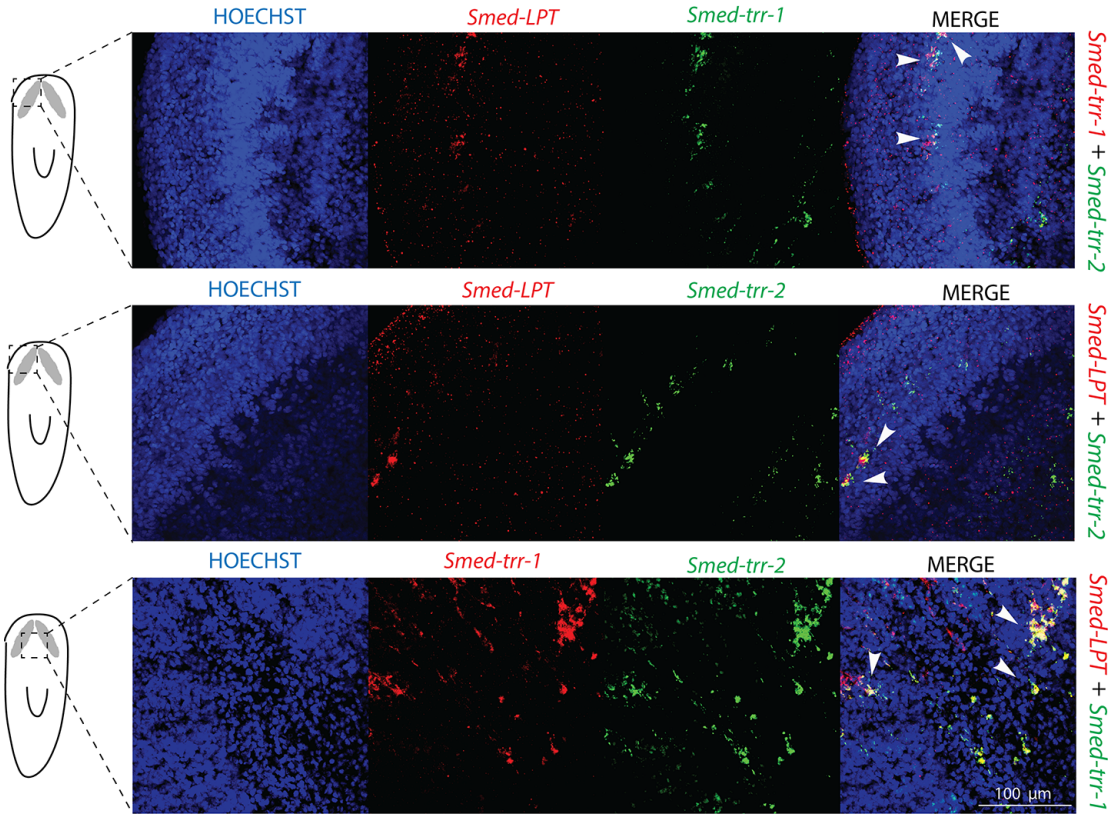


Figure 3.5. Trithorax-related genes co-localise. Co-localisation between the genes from the trithorax-related family in *S. mediterranea*. *Smed-LPT* + *Smed-trr-1*, *Smed-LPT* + *Smed-trr-2* and *Smed-trr-1* + *Smed-trr-2* co-localisation experiments are shown. Ten worms per condition were used. The schematics indicate which parts of the animals' head regions are shown (the cephalic lobes are in gray). White arrows indicate examples of double positive cells.

3.5. CONCLUSION

Smed-LPT, *Smed-trr-1* and *Smed-trr-2* are expressed in neoblasts, neoblast progeny cells and differentiated body structures like the brain, the gut (except *Smed-trr-1*) and the pharynx. The genes' expression in these structures is also observed during planarian regeneration. Co-localisation studies demonstrated that almost all planarian stem cells are positive for trithorax-related genes. Establishing whether trithorax-related genes are expressed in cNeoblasts and/or committed neoblasts was not possible. Trithorax-related genes co-localise with each other, suggesting that they might work together in the same cells.

CHAPTER 4. KNOCKDOWN OF TRITHORAX-RELATED GENES PERTURBS REGENERATION AND HOMEOSTASIS AND INDUCES DEFECTS IN STEM CELL DIFFERENTIATION

4.1. ABSTRACT

Planarians have to constantly produce new differentiated cells as part of homeostatic cell turnover. This process is significantly more demanding during regeneration. Perturbing the function of trithorax-related genes negatively impacts the production and maintenance of differentiated cells in the brain, gut, pharynx and epidermis, but not protonephridia and cilia. These defects are not due to stem cell maintenance failure. The defects are most pronounced in *Smed-LPT*(RNAi) animals and less obvious in *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi), potentially due to functional redundancy between *Smed-trr1* and *Smed-trr2*. SMED-LPT-TRR function could be related to nuclear receptor co-activation, but there is no solid evidence for this. The fact that not all tissues are affected following knockdown leads to the conclusion that trithorax-related genes may regulate differentiation of specific lineages.

4.2. INTRODUCTION

4.2.1. LPT-TRR FUNCTION IN DROSOPHILA

Studies of *Drosophila* LPT function are still in their infancy. The existence of LPT was only identified in 2011 (by Mohan et al.). However, significant advances have already been made. Chauhan et al. (2012) found that homozygous *LPT* mutants (where *LPT* promoter and parts of the *LPT* open reading frame are deleted) usually died before the second or early third instar larva stages. The authors hypothesised that a requirement for *LPT* much earlier in development could had been masked by the fact that the mother delivered substantial amounts of *LPT* mRNA during embryogenesis (Chauhan et al., 2012). Consistent with this hypothesis, Sedkov et al. (1999) demonstrated that lack of zygotic *trr*, combined with 50% reduction in maternal *trr* resulted in embryonic lethality.

Upon studying of the dead *LPT* mutant larvae, Chauhan et al. (2012) noticed that L2 and L3 mouth hooks suffered severe malformations, indicative of a defect in cuticle shedding, a process dependent on the steroid hormone ecdysone. The few mutant larvae that did survive, developed patches of necrotic tissue (indicative of hematopoietic problems) and failed to form a cuticle when they attempted entering the pupal stage (Chauhan et al., 2012; Dearolf, 1998).

Targeted RNAi-based depletion of *LPT* in *Drosophila* wing imaginal discs gave rise to shortened veins in the adult wings, while the same manipulation in the developing eye imaginal disc led to the development of smaller eyes and loss of antennal and head structures (Chauhan et al., 2012). The authors demonstrated that the function of LPT is dose-sensitive during development since mild overexpression of LPT led to lengthened veins in the wing, changes in cuticle pigment, eye outgrowths, as well as ectopic sex combs (Chauhan et al., 2012).

Interestingly, overexpression of *LPT* combined with RNAi-mediated depletion of *trr* or *utx* (H3K27 demethylase in the same complex) resulted in almost complete suppression of the *LPT* overexpression phenotype (Chauhan et al., 2012). This finding indicates that *LPT*'s functions in *Drosophila* are dependent on the methyltransferase Trr and the demethylase UTX. Furthermore, a combined depletion of *LPT* and *trr* led to a stronger phenotype than depleting either of them alone, possibly indicating that they cooperate on the same gene targets (Chauhan et al., 2012).

It had previously been demonstrated that Trr binds to the ecdysone receptor EcR (Farnesoid-X-receptor/liver-X-receptor homologue) and its heterodimer partner USP (Retinoid-X-receptor homologue) in the presence of ecdysone (Sedkov et al., 2003). Chauhan et al. (2012) showed that LPT binds to Trr, EcR and USP. LPT was also found to activate transcription from an Ecdysone Receptor response element (ECRE) as well as transcription of the early ecdysone-response genes E74A and E74B. This led to the conclusion that both LPT and Trr serve as nuclear receptor co-activators in ecdysone-dependent processes in *Drosophila* metamorphosis (Chauhan et al., 2012; Sedkov et al., 2003). These effects were connected to LPT-Trr regulating H3K4 trimethylation on specific promoters (Chauhan et al., 2012; Sedkov et al., 2003). These *Drosophila* findings were among the first indications that LPT-Trr/MLL3/4 COMPASS-like complex might have a function in organismal development via controlling hormone-dependent processes.

In order to explore further the function of LPT-Trr complex, Kanda et al. (2013) focused on the *Drosophila* eye. Mutations in *trr* led to affected cells showing growth advantage compared to wildtype cells in the eye. The eyes with a *trr* mutation also

showed defects in photoreceptor specification. The phenotypes observed in *trr and utx* mutants were very similar. The authors demonstrated that, like *utx* mutants, *trr* mutants showed increased amount of intracellular domain of Notch, leading to the conclusion that both *utx* and *trr* have a Notch suppressing function. Interestingly, unlike studies on LPT-Trr, so far, Kanda et al. (2013) noticed that the mutant phenotype was associated with a decrease in H3K4 monomethylation in the mutant cells, rather than to decreased trimethylation. Kanda et al. (2013) steered the study of LPT-Trr/MLL3/4 in a different direction, indicating that the COMPASS-like complex is probably involved in more histone modifications that just H3K4me3. These functions might not necessarily be hormone-related. Roles of LPT-Trr/MLL3/4 in histone modifications will be discussed in detail in **Chapter 5**.

Exploring the molecular pathways regulated by LPT and Trr, Sedkov et al. (2003) demonstrated that mutating *trr* in the eye imaginal disc led to perturbations in the Dpp (Decapentaplegic) signaling pathway, downstream of Hh (Hedgehog) signaling. This was reflected in delayed morphogenetic furrow progression and a 'rough eye' phenotype, indicative of photoreceptor differentiation defects (Sedkov et al., 2003), also seen in the Kanda et al. (2013) study. Chauhan et al. (2013), on the other hand, demonstrated that the LPT regulation of Dpp in the *Drosophila* wing was not Hh-dependent. Paradoxically, Kanda et al. (2013) reported an increase in pMAD (phospho – Mothers against Dpp; downstream of Dpp) staining in *trr* mutant larval eye imaginal discs, while Chauhan et al. (2013) reported a decrease of pMAD in *LPT* mutant third instar wing imaginal disc. These findings indicate one of three things: (1) there are irreconcilable experimental methodology differences between the cited

studies; (2) LPT and Trr differ in the ways they regulate Dpp signaling; (3) the LPT-Trr complex regulates Dpp differently across different tissues.

4.2.2. SET-16 FUNCTION IN C. ELEGANS

In *C. elegans*, a null mutation of the *MLL3/LPT/trr* homologue (*set-16*) was found to be lethal (Andersen and Horvitz, 2007). An RNAi-based study implicated *set-16* in RAS signaling attenuation (Fisher et al., 2010). The authors showed that *set-16* was expressed in the worm's vulval precursor cells with higher expression in the non-vulval descendents than in the vulval cells (Fisher et al., 2010). Interestingly, the worm's *set1* homologue (*set-2*) showed the reverse expression in the vulval precursor cells (Fisher et al., 2010). These results, together with the observation that UTX-1 and SET-16 both suppressed RAS signaling (crucial for vulval fate determination), led Fisher et al. (2010) to conclude that these two enzymes are involved in the prevention of cells mistakenly adopting vulval fate.

SET-16 and UTX-1's involvement in vulval cell specification was tied to H3K4 trimethylation and H3K27 demethylation activity respectively (Fisher et al., 2010). One of the consequences of this activity was the up-regulation of AJM-1 (Apical junction molecule-1), a repressor of RAS signaling (Fisher et al., 2010). RAS signaling hyperactivation had been linked to cancer (Mahgoub et al., 1998).

SET-16 and UTX-1's effect on H3K4 monomethylation was not tested in *C. elegans*.

4.2.3. MLL3/4 FUNCTION IN MAMMALS

MLL3 and MLL4 have been found in a retinoic acid receptor context in HeLa cells, consistent with a nuclear receptor coactivator function (Goo et al., 2003). Suppression of either *Mll3* or *Mll4* via RNAi in mice led to reduced H3K4 trimethylation as well as reduced expression of a retinoic acid receptor target genes (Lee et al., 2006a). These findings agree with LPT-Trr's nuclear receptor coactivator functions in *Drosophila* (Chauhan et al., 2012; Sedkov et al., 2003). Eissenberg and Shilatifard (2010) suggested that the existence of two LPT-Trr homologous enzymes in mammals (MLL3 and MLL4) might reflect an evolutionary diversification necessary because of the increased number of nuclear receptors in mammals compared to *Drosophila*.

A mutant heterozygous for a partial deletion of the SET domain of MLL3 had no detectable phenotype, while a homozygote was partially embryonic lethal in mice (Lee et al., 2006a). The surviving homozygous mice showed stunted growth and delayed eye opening (Lee et al., 2006a). Furthermore, embryonic fibroblast cells isolated from E12.5 *Mll3* or *Mll4* knockout mouse embryos showed a delayed proliferation response compared to cells from control animals (Lee et al., 2006a). Further defects of *Mll3* knockout mice included decreased white fat mass, impaired accumulation of lipids in the liver and brown adipose tissue upon high-fat diet, increased bile acid levels, hyperactive behaviour, higher energy expenditure, improvement in glucose tolerance (Kim et al., 2009, 2011; Lee et al., 2008a, 2008b). Lastly, female *Mll3* homozygous knockouts were either infertile or hypofertile, while males were hypofertile (Lee et al., 2006a).

Unlike *Mll3* homozygote mutants, *Mll4* homozygote mutation was completely embryonic lethal (Kim et al., 2015). Heterozygotes, on the other hand, much like surviving *Mll3* homozygote knockdown mice, had stunted growth, low body fat, improved glucose tolerance, increased bile acid production. A mutant heterozygous for *Mll3* and *Mll4* mutations showed an even stronger phenotype, demonstrating that MLL3 and MLL4 had a lot of common metabolic targets. RNA-seq analysis on *Mll3* and *Mll4* knockout mice revealed that the two enzymes' targets were predominantly in the liver. The down-regulated genes were mainly involved in ribosomal biogenesis and oxidative phosphorylation (primary energy-consuming cellular processes), which would explain why *Mll3* knockout mice showed increased expenditure of energy. Analysis of the up-regulated genes revealed genes involved in bile acid production and circadian clock regulation. It was demonstrated that retinoid-related orphan receptors (RORs) are the transcription factors responsible for MLL3/4's observed circadian clock function in the liver. MLL3 and MLL4 were reported to suppress bile acid production as a key part of their circadian rhythm control (Kim et al., 2015). Valekunja et al. (2013) also reported MLL3's involvement in circadian rhythm processes in the liver.

Another important finding regarding the function of MLL3/4 is that 50% of surviving homozygous *Mll3* knockout mice started exhibiting hydronephrotic kidneys, as well as hyperproliferation and tumour formation in the innermost layer of ureter cells within 4 months after birth (Lee et al., 2009). Despite the fact that terminally differentiated urothelial cells are not expected to proliferate, those in the tumours of *Mll3* knockouts showed an increase in PCNA staining (indicative of cell proliferation),

consistent with a broken tumour suppressor mechanism. Furthermore, a homozygous *Mll3* knockout in a heterozygote *p53* knockout background led to 100% of all surviving animals developing the described tumours in a much shorter time. More findings illustrating the connection between MLL3 and p53 were increased amount of DNA damage and decreased levels of p21 (a p53 target) in *Mll3* knockout mice. ChIP (Chromatin Immunoprecipitation) experiments also revealed that MLL3 was recruited to p21-p53REs (response elements) in MEF cell lines, which correlated with H3K4 trimethylation in those regions. Finally, the complex containing MLL3 was shown to directly interact with p53 (Lee et al., 2009).

Interestingly, the MLL3 and MLL4 complexes depend in their nuclear receptor co-activator activity on a subunit called ASC-2 (Goo et al., 2003). ASC-2 contains two NR (nuclear receptor) boxes – the C-terminal NR box interacts with the liver X receptors, while the N-terminal box is more promiscuous and binds more nuclear receptors, including retinoic acid receptor (RAR) (Lee et al., 2001). Goo et al. (2003) demonstrated that immunoaffinity purification with antibody against ASC-2 resulted in pulling down MLL3 and MLL4 among other proteins. Not surprisingly, the phenotypic defects observed in *Mll3* and *Mll4* mutants were very similar to what happens in *asc-2* knockout mutants (Lee et al., 2006a). Common features included decreased cellular doubling rate, stunted growth and decreased fertility (Lee et al., 2006a). It has was also shown that using neutralising antibodies against ASC-2 led to lack of transactivation by RAR (retinoic acid receptor) and other nuclear receptors (Lee et al., 1999). Retinoid treatment led to the recruitment of ASC-2 and its binding partners (including MLL3 and MLL4) to the β-RARE promoter region and the subsequent H3K4 trimethylation of the region (Goo et al., 2003). Similar findings

were reported by Lee et al. (2006) who found that phenotypes of ASC-2, MLL3 or MLL4 depletion all resulted in decrease in H3K4 trimethylation of the *RAR*-β*2* target gene of RAR. It was also reported that MLL3 and MLL4 form their own respective complexes with ASC-2 (Lee et al., 2006a). All of these findings point to the interpretation that ASC-2 binds nuclear receptors and the co-activator function of the complex happens via the H3K4 trimethylation performed by MLL3/4 (Goo et al., 2003; Lee et al., 2006a).

4.2.4. SMED-TRR-1 AND SMED-TRR-2 IN PLANARIANS

Hubert et al. (2013) performed some functional analyses on the Trr homologues in planarians. Using RNAi by feeding, they demonstrated the formation of a smaller regenerative blastema and delayed eye regeneration in *Smed-trr-1* knockdown animals, while no phenotypic defects were observed in *Smed-trr-2* knockdown planarians. When the authors performed a double knockdown of *Smed-trr-1* and *Smed-trr-2*, they observed a stronger *Smed-trr-1* knockdown phenotype – little or no regenerative blastema was formed following amputation and most animals curled ventrally and died from tissue lysis. The increased severity of the phenotype points towards (partial) functional redundancy between *Smed-trr-1* and *Smed-trr-2*. The ventral curling phenotype was also observed in a planarian RNAi screen for genes involved in regeneration (Reddien et al., 2005a). Ventral curling, accompanied by head regression, is the effect of lethally irradiating planarians and forcing them to undergo regeneration (Reddien et al., 2005a). This observation led Reddien et al. (2005a) to conclude that this phenotypic feature might be indicative of failure of

neoblast function. Hubert et al. (2013) did not observe head regression in their *trr* double knockdown animals, so it is not entirely clear whether the defect was due to a stem cell maintenance problem.

Hubert et al. (2013) also observed that *Smed-ptip*(RNAi) and *Smed-utx*(RNAi) (unique accessory subunits of MLL3/4 complexes) resulted in phenotypes similar to that following *Smed-trr-1*(RNAi) – smaller blastema and fainter photoreceptors.

4.2.5. NUCLEAR RECEPTORS IN PLANARIANS

Hubert et al. (2013) failed to find an ASC-2 (also known as NCOA6) homologue in planarians. Interestingly, yeast, *Drosophila* and *C. elegans* do not have ASC-2 orthologue either (Goo et al., 2003). This could indicate that ASC-2 appeared during vertebrate/deterostome evolution. Since ASC-2 targets and directs MLL3/4's nuclear receptor co-activator function (Goo et al., 2003), its appearance in vertebrates could reflect a need for fine-tuned co-activation of the increased number of nuclear receptors, compared to invertebrates, where ASC-2 is missing (Goo et al., 2003). Despite the fact that LPT-Trr complex in *Drosophila* does not have ASC-2 homologue to act as a nuclear receptor-binding agent, both LPT and Trr have several nuclear receptor binding motifs themselves (Chauhan et al., 2012; Sedkov et al., 2003). As shown in the **Chapter 1**, planarian LPT and TRR-2, were not found to have any nuclear receptor binding motifs (NR boxes). SMED-TRR-1 has one NR box. These findings could mean either that the LPT-TRR protein complex in planarians has functions additional to nuclear-receptor binding, or that the nuclear receptor family

in planarians is not very evolved, hence SMED-TRR-1's NR box is sufficient for nuclear receptor recognition.

Very little is known about nuclear receptors in planarians. Providing an exhaustive list of different types of nuclear receptors in flatworms is beyond the scope of this work. There are, however, four planarian nuclear receptors described so far. The first one is *Smed-tlx-1*, belonging to Retinoid-X-receptor-like subfamily (Escriva et al., 2004). *Smed-tlx-1* is homologous to the *Drosophila tailless* and was identified by Raška et al. (2011). The authors found that *Smed-tlx-1* was expressed throughout the body, but with much higher expression in the head region (Raška et al., 2011). Upon RNAi by feeding, the animals exhibited slow movement and started regressing their head (Raška et al., 2011). When the RNAi treatment was discontinued, the worms started recovering their heads, but demonstrated under-developed eyes (Raška et al., 2011). A more detailed functional characterization of *Smed-tlx-1* is lacking.

Another nuclear receptor belonging to the Retinoid-X-receptor-like superfamily is *Smed-hnf-4* (*hepatocyte nuclear factor 4*). Not much is known about this gene, other than that, together with *Smed-gata4/5/6*, it is expressed in planarian gut progenitor cells and fully differentiated gut cells (Scimone et al., 2014).

A similarly understudied nuclear receptor is the planarian thyroid receptor 1. The gene was identified as part of a screen and the only known fact about it is that it belongs to the so-called second wave (W2) of wound response genes (see **Chapter 1**) peaking in their expression 3-12 hours after wounding (Wenemoser et al., 2012).

Finally, retinoic acid treatment was shown to severely compromise head regeneration in planarians and delay eye formation (Ermakova et al., 2009). The

study also identified a retinoid X receptor homologue in planarians and speculated that the retinoic acid treatment had its effect via this receptor (Ermakova et al., 2009). However, no experiments were performed to explicitly test this.

4.2.6. Aim of chapter

Results from experiments knocking down *Smed-LPT*, *Smed-trr-1* and *Smed-trr2* are presented in this chapter, where the focus is on stem cell differentiation defects. Both regeneration and homeostatic phenotypes are considered and evidence is presented that phenotypes are not related to stem cell maintenance problems. A plethora of molecular markers are used to assess the recovery of missing tissues and organs or their homeostatic maintenance. Finally, the findings are discussed in the light of previous studies described in this chapter's introduction.

4.3. Materials and Methods

4.3.1. Primers for in situ hybridisation markers

Primers used for experiments in this chapter are listed in **Appendix B**.

4.3.2. Gene clones used

Gene clones used for experiments in this chapter are listed in **Appendix G**.

4.3.3. ANTIBODIES

The antibodies used for the experiments in this chapter were:

- mouse anti-SYNORF-1 (also known as anti-synapsin or 3C11) purchased from Developmental Studies Hybridoma bank and used at 1:100 dilution
- mouse anti-SMEDWI-1 – provided by Prof. Jochen Rink; used 1:1000
- anti-acetylated tubulin – purchased from Developmental Studies Hybridoma bank and used at 1:200 dilution
- Alexa Fluor 488 goat anti-mouse IgG – purchased from Molecular Probes; used 1:400

4.3.4. IMMUNOHISTOCHEMISTRY

Immunohistochemistry was performed either on its own (anti-synapsin antibody) or after *in situ* hybridisation (anti-SMEDWI-1 antibody). Immunohistochemistry was performed as described in Cebrià and Newmark (2005).

4.3.5. CELL COUNTING

Samples, for which cell counting was performed, were imaged on a confocal microscope (as previously described in **Chapter 3**). The resultant z-stacks were collapsed and flattened using FIJI (Schindelin et al., 2012). If the staining was cytoplasmic, cells were counted manually on Photoshop CS6; if the staining was nuclear, FIJI's automatic cell counter was used. The total number of cells was then normalised to imaged area. This was done as follows:

-Area in microns was measured on FIJI.

-This value 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 – 4 microns for all imaged stacks).

4.3.6. RNAi OF TRITHORAX-RELATED GENES

Double stranded RNA was prepared as previously described (**Chapter 3**). The dsRNA for each gene was diluted to 2 µg/µl according to Nanodrop measurements. RNAi was achieved via injections, as previously described. The injection schedule followed was three consecutive days of injections, followed by three days of rest and final four consecutive days of injections. After the final injection, animals were either cut on the next day or left whole for homeostatic observation and/or experimentation.

4.3.7. ECTOPIC SEROTONIN TREATMENT

Planarians were treated with 100 µM final concentration of Serotonin hydrochloride (Sigma Aldrich) in planarian water, as per (Currie and Pearson, 2013). The animals were observed for up to one hour and videos were taken.

4.3.8. PHENOTYPE SCORING AND BRIGHTFIELD IMAGING

Regenerating animals were scored and imaged every day or every other day. Homeostatic animals were scored and imaged every week. Scoring entailed observing each animal under a microscope and making a note of its appearance. The characteristics of interest were blastema size (for regenerates), presence of pharynx, presence of eyes, position of eyes, locomotion, head regression (for homeostasis) and outgrowth and swelling formations.

Animals were not fed during RNAi and after RNAi treatment.

Still images and videos were taken on a Zeiss Discovery V8 (Carl Zeiss) microscope with a Canon EOS 600D or Canon EOS 1200D camera.

4.4. Results and Discussion

4.4.1. Knockdown of trithorax-related genes

4.4.1.1. Knockdown of trithorax-related genes is effective

Smed-LPT, *Smed-trr-1* and *Smed-trr-2* were knocked down via RNAi. The efficacy of each of the three knockdowns was assessed via *in situ* hybridisation. The *in situ* probes used hybridised to mRNA regions different than the region targeted by each knockdown. In this way, whether the expression of the whole transcript is affected can be ascertained. Results shown in **Figure 4.1** confirmed this expectation. The animals used for this experiment had undergone a typical course of RNAi, amputated the day after the last dsRNA administration and fixed for *in situ* hybridisation on the third day of regeneration. This timepoint was chosen arbitrarily. There is no reason to expect different results upon choosing a different timepoint.

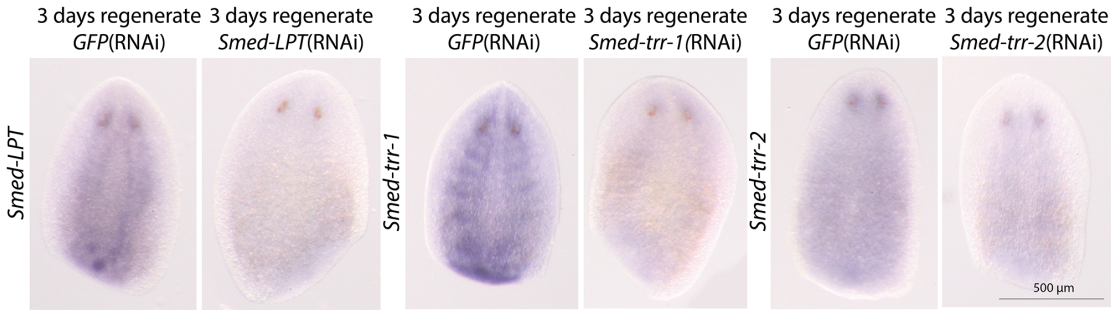


Figure 4.1. RNAi of trithorax-related genes is effective. Knockdown efficacy for *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* was checked via *in situ* hybridisation for a region non-overlapping with the region chosen for knockdown. The typical knockdown course was performed for each gene, after which amputation followed. The samples used were 3-day-regenerating head pieces. Ten regenerating pieces per condition were used.

4.4.1.2. Tissue regeneration and homeostasis are impaired following knockdown of trithorax-related genes

Regeneration and homeostasis phenotype assessment assays were used. In each case, twenty animals per condition were observed. In the regeneration assay, phenotypic defects due to knockdown started becoming obvious at later regeneration stages. Two regeneration timepoints (day 3 and day 10) are shown in **Figures 4.2A** and **Figure 4.2B** for comparison. During early regeneration, all control and experimental animals formed a blastema and looked similar. At regeneration day 5, however, the experimental knockdown animals started exhibiting delayed blastema growth and inability to replace missing structures. At day 10 of regeneration (**Figure 4.2B**) defects in experimental animals were more obvious. *Smed-trr-2*(RNAi) animals were almost identical to control animals, the only difference being that their eyes were slower to develop (3/20 tail pieces never formed eyes) and 10% of the middle pieces showed posterior bloating. Also, less than 20% of tail and middle pieces of *Smed-trr-2*(RNAi) animals exhibited slower mobility and occasional 'inchworming'. This observation differs from Hubert et al.'s (2013) conclusion that knockdown of *Smed-trr-2* results in no visible phenotype. This discrepancy could be easily explained by the fact that RNAi administered by injection (as in the present experiment) is more controlled and has a stronger effect than RNAi by feeding (the preferred method in Hubert et al. (2013)).

At day 10 of regeneration of *Smed-trr-1*(RNAi) animals, 14/20 head pieces, 14/20 middle pieces and 12/20 tail pieces had smaller blastema compared to controls. Eye regeneration was much slower than controls, with some worms never forming eyes

and others forming them in the post-blastema region. Over 60% of head, middle and tail pieces were moving slower than controls and/or inchworming.

The most pronounced phenotype at 10 days of regeneration was in *Smed-LPT*(RNAi) animals. Ten percent of the head pieces showed outgrowths/swellings and 50% had a dented posterior blastema. Only 3/20 head pieces seemed to have developed a pharynx. From the middle pieces, all animals had smaller blastemas (especially anterior) and 15/20 developed faint eyes in the post-blastema. Twenty-five percent of the middle pieces and 60% of the tail pieces did not develop eyes. One out of twenty tail pieces in the *Smed-LPT*(RNAi) condition had formed an outgrowth at the blastema region. Eighty percent of the animals in this experimental condition either did not move at all or inchwormed. Usually, the motility phenotype increased in severity with the progression of regeneration – from inchworming to complete immobility. A more detailed scoring of both regeneration and homeostasis phenotypes across the animals' lifespan is presented in **Appendix H**. The motility phenotype observed primarily in *Smed-LPT* and *Smed-trr-1* knockdown animals is discussed later in this chapter.

The observation that *Smed-LPT* and *Smed-trr-1* knockdown regenerating animals formed a normal blastema at first and later during regeneration it became smaller, might indicate that the initial tissue response required to make a blastema was intact. The problem could lie in the elaboration of this blastema into missing tissues later on. The results of an experiment challenging this interpretation are shown in **Figure 4.2C**. The regeneration of knockdown animals after two rounds of amputation (first cut - the day after final dsRNA injection; second cut – 24 hours later cut off

blastemas of middle piece; image 3 days after second cut) is shown (**Figure 4.2C**). Both anterior and posterior blastemas in *Smed-LPT*(RNAi) animals were underdeveloped when compared to control. Samples at three days regeneration after single amputation (**Figure 4.2A**), however, looked similar to control. The double amputation experiment caused an acceleration of the phenotype. This could indicate that the first stages in blastema development in the single cut experiment appeared unperturbed because there were enough relatively healthy stem cells ready to proliferate and migrate to the blastema upon amputation. In the double cut experiment, on the other hand, this source of relatively healthy cells was depleted (in response to the first cut) and the second cut response was performed by more severely affected by the knockdown cells. Hence, the phenotype became visible sooner in brightfield observation. These results argue against the interpretation that *Smed-LPT* is specifically involved in later stages of blastema commitment and the processes governing initial blastema assembly are intact in *Smed-LPT*(RNAi) conditions. What seems to be more likely is that the presence of relatively healthy cells at the onset of regeneration is masking the effect of *Smed-LPT*(RNAi). The double cut experiment was performed on *Smed-trr-1* and *Smed-trr-2* knockdown animals as well, but they responded no differently to controls.

Another way in which the *Smed-LPT*(RNAi) animals differred from *Smed-trr-1* and *Smed-trr-2*(RNAi) animals was in outgrowth formation. At 14 days of regeneration (single cut) 50% of *Smed-LPT* knockdown head pieces had formed outgrowths (**Figure 4.2D**). The head pieces were the most predisposed to outgrowths (**Figure 4.2E**). The outgrowth side of the *Smed-LPT* knockdown phenotype is addressed in more detail in the **Chapter 5**.

Smed-LPT(RNAi) regenerates had the shortest lifespan out of all knockdown conditions, followed by *Smed-trr-1*(RNAi), with *Smed-trr-2*(RNAi) animals surviving for up to two months following amputation (**Figure 4.2F**). In all experimental conditions, the regenerating animals with the shortest lifespan were the tail pieces (**Figure 4.2F**). This could be explained by the fact that the tail piece had to recover more vital tissues (brain, gut, pharynx) that the head and middle pieces. The challenges associated with this process were aggravated by the knockdown of trithorax-related genes.

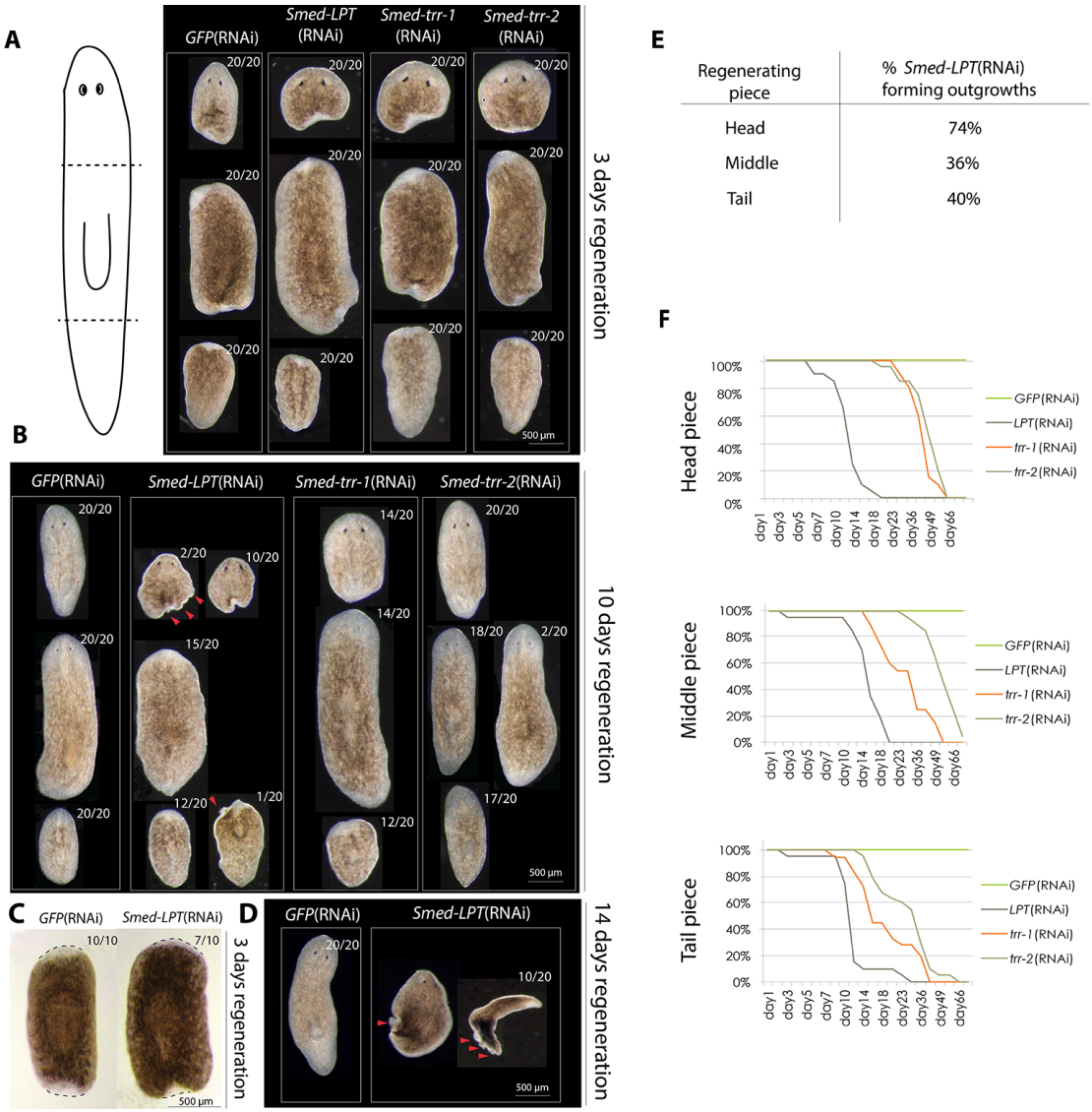


Figure 4.2. Knockdown of trithorax-related genes affects regeneration. (**A**) A schematic showing the regions where the amputations were performed. Head, middle and tail pieces are shown at day 3 of regeneration after *Smed-LPT*(RNAi), *Smed-trr-1*(RNAi), *Smed-trr-2*(RNAi) and *GFP*(RNAi). White numbers next to each piece indicate proportion of animals that exhibit the same phenotype at this timepoint. Twenty worms per condition were used. (**B**) Ten days of regeneration following knockdown of trithorax-related genes. Red arrows indicate outgrowth formation. Twenty worms per condition were used. (**C**) Following the last dsRNA administration, animals were cut into three pieces, 24 hours later the newly formed blastemas were cut off and animals were imaged on the third day after the second cut. *Smed-LPT*(RNAi) animals form smaller blastemas compared to control. Dotted line shows blastema boundaries in middle pieces. Ten worms per condition were used. (**D**) At day 14 of regeneration 10/20 *Smed-LPT*(RNAi) head pieces show outgrowth formation (indicated by red arrows). Twenty worms per condition were used. (**E**) Percentage of heads, middles and tails of *Smed-LPT*(RNAi) animals that form outgrowths throughout their lifespan. (**F**) Survival curves of heads, middles and tails following trithorax-related genes knockdown. Twenty worms per condition were used.

Brightfield observations of homeostatic knockdown animals were also performed. By week 2 of homeostatic tissue turnover, only 2/19 *Smed-LPT*(RNAi) animals looked like controls (**Figure 4.3A**). Four out of nineteen animals showed head regression (yellow arrow) with 1/19 showing outgrowth tissue. It was previously described that head regression and ventral curling are characteristic of failure in neoblast maintenance (Reddien et al., 2005a). However, only 21% of the *Smed-LPT*(RNAi) animals underwent head regression and none of them exhibited ventral curling. Therefore, it does not seem likely that the phenotype was due to neoblast maintenance defects (to be addressed in more detail in the next section of this chapter). Around half of *Smed-LPT*(RNAi) animals had also lost their pharynx by 2 weeks of homeostasis.

Smed-trr-1(RNAi) and *Smed-trr-2*(RNAi) animals, on the other hand, looked like controls at this stage. Furthermore, no noticeable homeostatic defects were present at 10 weeks following *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) (**Figure 4.3B**). *Smed-LPT*(RNAi) animals stopped moving and died within 3 weeks after the last dsRNA injection (**Figure 3.3C**). *Smed-trr-1* and *Smed-trr-2* knockdown homeostatic animals could survive up to twelve weeks after the last dsRNA administration (**Figure 4.3C**). Immediately before death, *Smed-trr-1* and *Smed-trr-2* knockdown worms started experiencing inability to move. Once this phenomenon occurred, the animals lysed on the next day, with no additional pathology visible in brightfield. No head regression or ventral curling were observed, suggesting that death did not ensue as a result of stem cell maintenance problems.

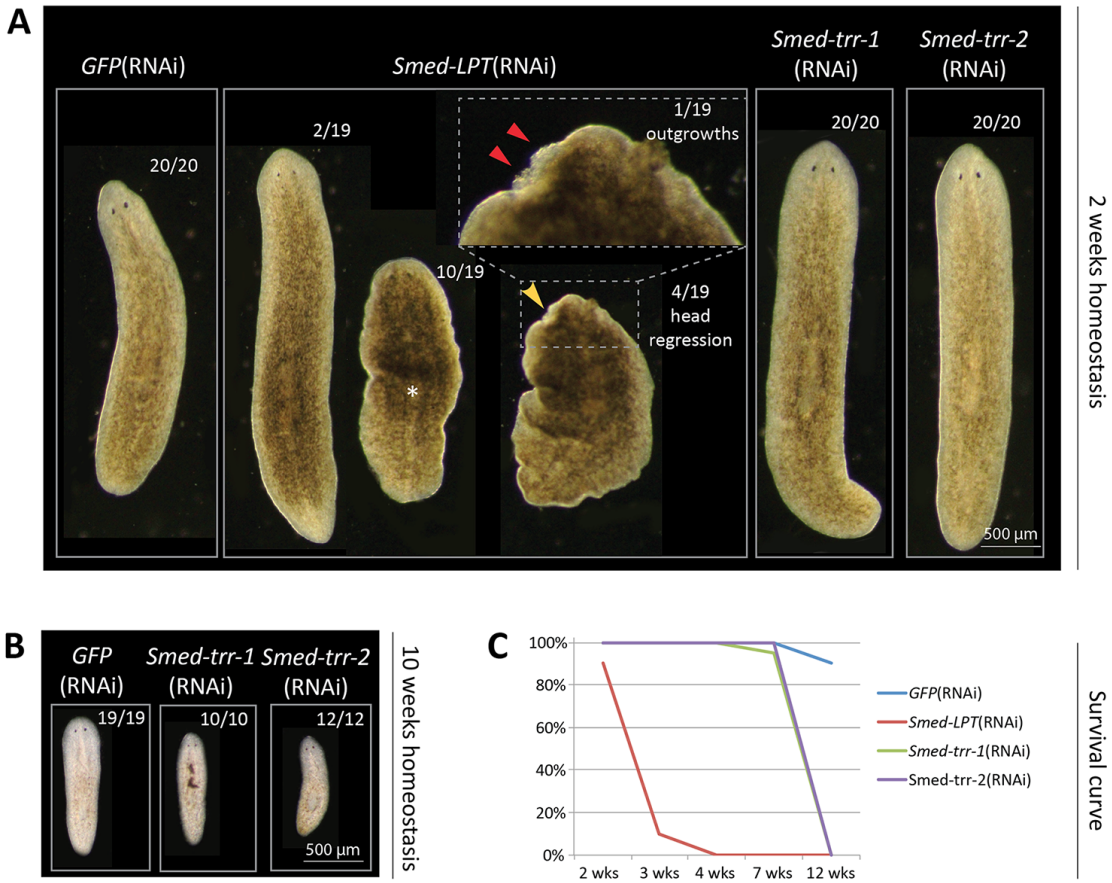


Figure 4.3. Knockdown of trithorax-related genes affects tissue homeostasis. (**A**) Illustration of pathology of two weeks of homeostatic cell turnover, following the last dsRNA administration. White numbers at the top right corner of each sample correspond to amount of animals in this condition showing similar features at this timepoint. Yellow arrow points toward head regression, while red arrows point to outgrowth formations. The asterisk indicates pharyngeal loss. Nineteen to twenty worms per condition were used. (**B**) Phenotype of *Smed-trr-1* and *Smed-trr-2* knockdown animals compared to control at 10 weeks of homeostasis. Ten to nineteen worms per condition were used. (**C**) Survival curves of the homeostatic animals following trithorax-related genes knockdown.

Our brightfield observations showed that *Smed-LPT* knockdown had a much stronger effect than either of the two *trr* knockdown. This finding was not surprising when considered in the light of Hubert et al.'s (2013) results where a double knockdown of *Smed-trr1* and *Smed-trr-2* produced a much stronger effect than either of the two single knockdowns. Such an observation is indicative of functional redundancy between the two *trr* genes. Similar double knockdown experiments (via injection) are described later on in this thesis (**Chapter 6**) and will not be discussed in the current chapter.

Our brightfield observations seem to contradict the expectation that the LPT-Trr complex could regulate Dpp signalling, based on experiments in *Drosophila* (Chauhan et al., 2013; Kanda et al., 2013; Sedkov et al., 2003). Planarian studies found that Dpp signalling was involved in midline and dorso-lateral specification during planarian regeneration and homeostasis (Molina et al., 2007; Reddien et al., 2007). None of the phenotypes described in Molina et al. (2007) and Reddien et al. (2007) showed any commonalities with the phenotypes described in this chapter. Hence, a functional relationship between LPT-Trr and Dpp in planarians was deemed unlikely.

4.4.2. Defects following trithorax-related genes knockdown are not due to failure in neoblast maintenance or an overall decrease in early neoblast progeny production

One possible explanation of the defects observed in **Figures 4.2** and **Figure 4.3** is that the stem cell pool was unable to maintain itself, as a consequence of which

homeostatic renewal and regeneration of missing tissues was compromised. Despite the fact that ventral curling was not observed in any of the samples, this possibility was examined further. Trithorax-related genes knockdown animals were stained with the neoblast marker *Smed-H2B.* Anti-SMEDWI-1 antibody was also used in these experiments. SMEDWI-1 protein is known to be expressed in stem cells as well as in their committed early progeny cells, regardless of the lineage (Guo et al., 2006; Wenemoser and Reddien, 2010). Hence, a double-labeling experiment with *Smed-H2B* would separate stem cells (double positive cells) from their immediate committed progeny (only positive for SMEDWI-1). Results showed that during late regeneration (**Figure 4.4A**) and homeostasis (**Figure 4.4B**), there was no significant difference between knockdown animals and controls in the observed number of stem cells. This finding contradicts the hypothesis that defects following trithorax-related genes knockdown are due to failure in neoblast maintenance. A colorimetric *in situ* hybridisation with *Smed-H2B* confirmed these results.

There were no significant changes in overall early stem cell progeny production as shown by SMEDWI-1^{+}/*Smed-H2B* $^{-}$ cell numbers. It should be noted that these cells represent all early stem cell progeny from across all lineages. Therefore, the lack of difference could mean one of two things: either the differentiation defects following knockdown of trithorax-related genes affect later stages of stem cell progeny commitment (when cells are no longer positive for SMEDWI-1 protein) or early stem cell progeny is affected differently across lineages. Experiments described in this chapter help explore some specific defects in progeny production. The behaviour of neoblast progeny and fully differentiated cells belonging to different lineages under trithorax-related genes knockdown conditions is assessed.

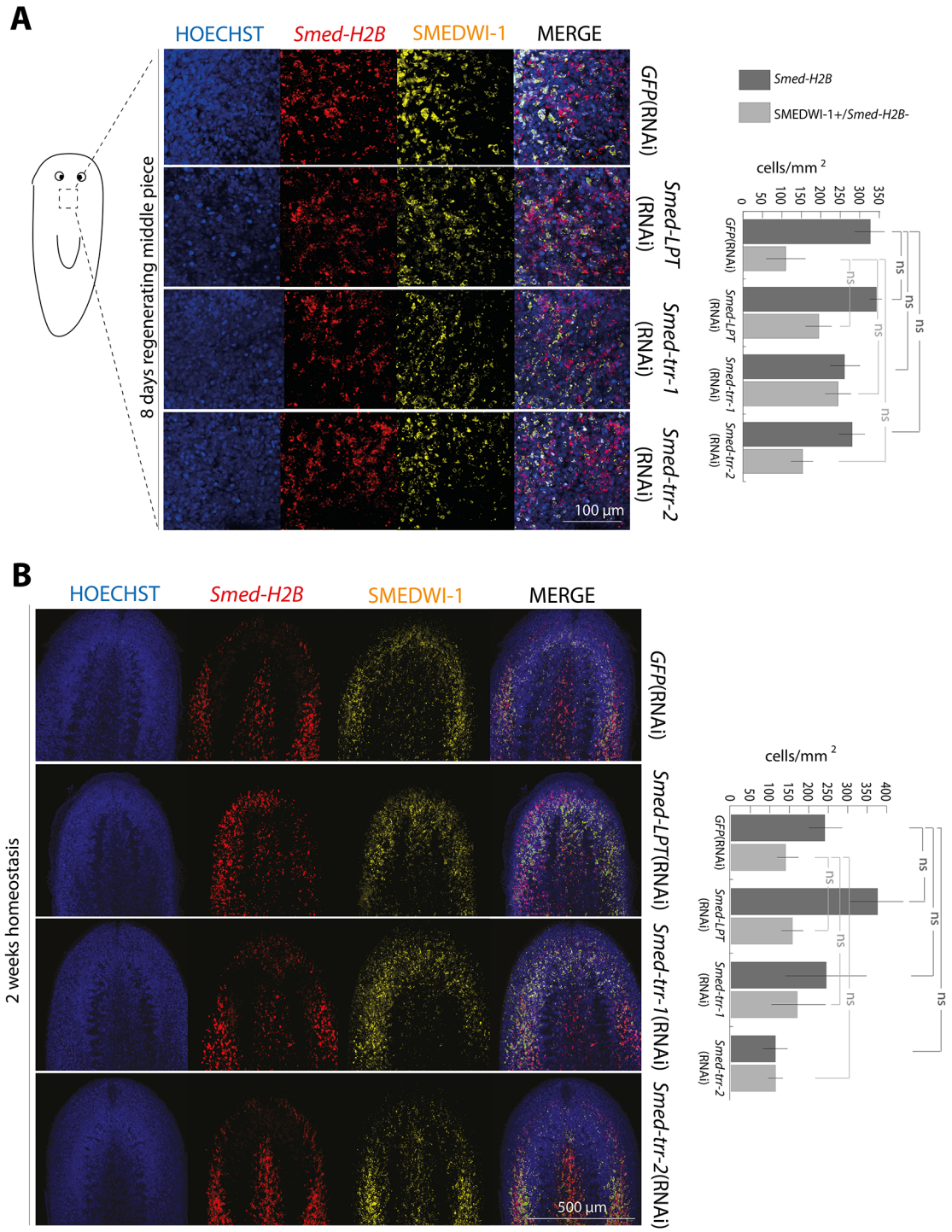


Figure 4.4. Stem cell maintenance and overall early neoblast progeny production is unaffected following trithorax-related genes knockdown. (**A**) 8-day regenerating middle pieces were stained with *Smed-H2B* and anti-SMEDWI-1 antibody. A schematic is shown illustrating the pre-pharyngeal region used for these images. Ten worms per condition were used. The differences in mean cell counts were graphed and assessed via 2-tailed ttest assuming unequal variance, where 'ns' stands for 'not significant'. (**B**) The anterior part of 2 weeks homeostasis knockdown animals from a double staining experiment with *Smed-H2B* and anti-SMEDWI-1 antibody is shown. Ten worms per condition were used. The differences in mean cell counts were graphed and assessed via 2-tailed ttest assuming unequal variance, where 'ns' stands for 'not significant'. Error bars in both graphs represent SEM.

4.4.3. TRITHORAX-RELATED GENES KNOCKDOWN IMPAIRS ANTERIOR STRUCTURES REGENERATION

One of the obvious effects of knocking down trithorax-related genes was difficulty in forming eyes during anterior regeneration. This issue was investigated using *Smed-SP6-9* (Lapan and Reddien, 2011), a marker for optic cups and stem cell progeny cells in the process of differentiating to become optic cup cells. As early as three days of regeneration, the formation of the characteristic trail pattern of optic cup precursor cells (shown in the cartoon) was visible in *Smed-trr-2*(RNAi) and control animals (**Figure 4.5**). The *Smed-SP6-9*-positive cells in *Smed-trr-1*(RNAi) animals were in somewhat of a disarray and those in *Smed-LPT* knockdown animals were absent. At day 8 of regeneration accumulation of *Smed-SP6-9*-positive cells was visible in the region where the optic cups should be in both controls and *Smed-trr-2*(RNAi) animals. There was also a visible trail pattern of these cells going from the posterior of the animal towards the optic eye cup. In *Smed-trr-1* knockdown animals there was an accumulation at the region of the optic cups, but a proper trail of *Smed-SP6-9*-positive cells was not obvious. This was the case in 60% of the animals. Ninety percent of *Smed-LPT* knockdown worms showed almost no *Smed-SP6-9*-positive cells.

These findings indicate that at early stages of regeneration, when developing eyes in the blastema were yet to become visible to the naked eye (day 3 regeneration), the arrangement of optic cup precursor cells was already perturbed in *Smed-trr-1*(RNAi). This was reflected by delayed formation of weaker eyes later on in those animals. In *Smed-LPT*(RNAi) animals no *Smed-SP6-9*-positive cells were being made, hence the

eyes observed at later regeneration stages were either absent, faint or with a visible photoreceptor, but not an eye cup.

Difficulties with eye formation following *LPT/Trr/Mll3/4* perturbation were reported in a number of studies in *Drosophila* (Kanda et al., 2013; Sedkov et al., 2003), mouse (Lee et al., 2006a) and planarians (Hubert et al., 2013). It is possible that these eye defects in planarians were a consequence of a conserved function of *Smed-LPT* and *Smed-trr-1* (and possibly *Smed-trr-2*) as nuclear receptor co-activators of the SMED-TLX-1 receptor. This hypothesis is supported by previous studies implicating this nuclear receptor in planarian eye regeneration (Raška et al., 2011).

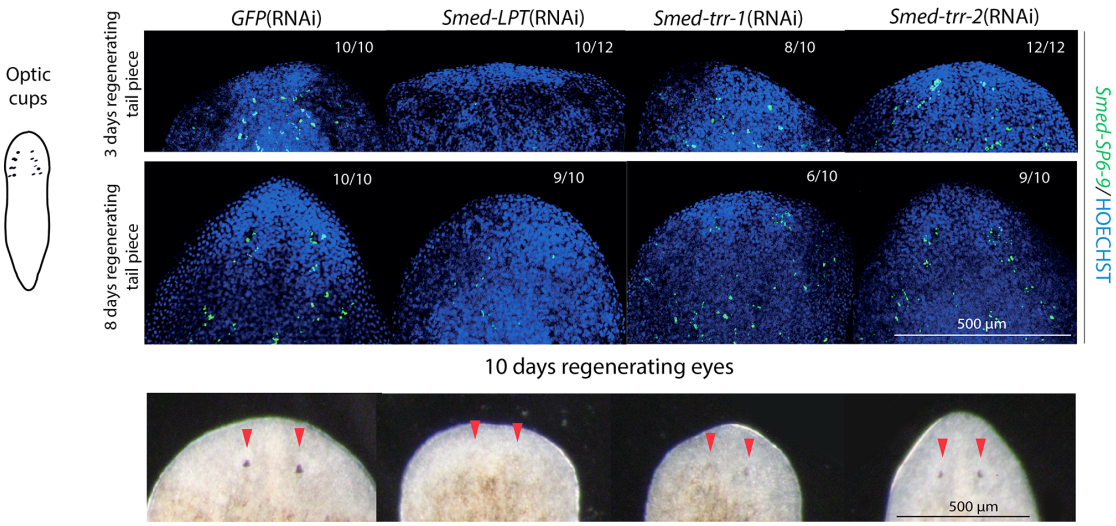


Figure 4.5. Trithorax-related genes knockdown affects optic cup regeneration. Optic cup formation in trithorax-related genes knockdown tail pieces at 3 and 8 days of regeneration. Schematic shows the expected pattern of *Smed-SP6-9*-positive cells. White numbers at top right corner of each image indicate proportion of animals exhibiting the same characteristics per sample. Brightfield images of typical 10-day regenerates from each condition are shown for reference. Red arrows point towards the eyes. Ten to twelve worms were used per condition.

Next, the production of different neuronal subtypes under trithorax-related knockdown conditions was examined. A schematic of a simplified CNS with two cephalic lobes, two ventral nerve cords (VNCs) and some enervation perpendicular to the VNCs is shown in **Figure 4.6**. The first investigated neuronal subtype was GABA-ergic neurons. The *Smed-GAD* (glutamic acid decardoxylase; converts glutamate to GABA) marker (März et al., 2013) was used to label these cells. At day 8 of regeneration there were significantly less GABAergic neurons produced in a regenerating tail piece that had undergone *Smed-LPT* or *Smed-trr-2* knockdown (**Figure 4.6**). The brains in *Smed-trr-1* knockdown pieces were not significantly different from control when it came to total number of GABAergic neurons. No earlier regeneration timepoints are shown in the figure, since GABAergic neurons appear to form relatively late in brain regeneration. Results were reproducible in a colorimetric *in situ* hybridisation experiment.

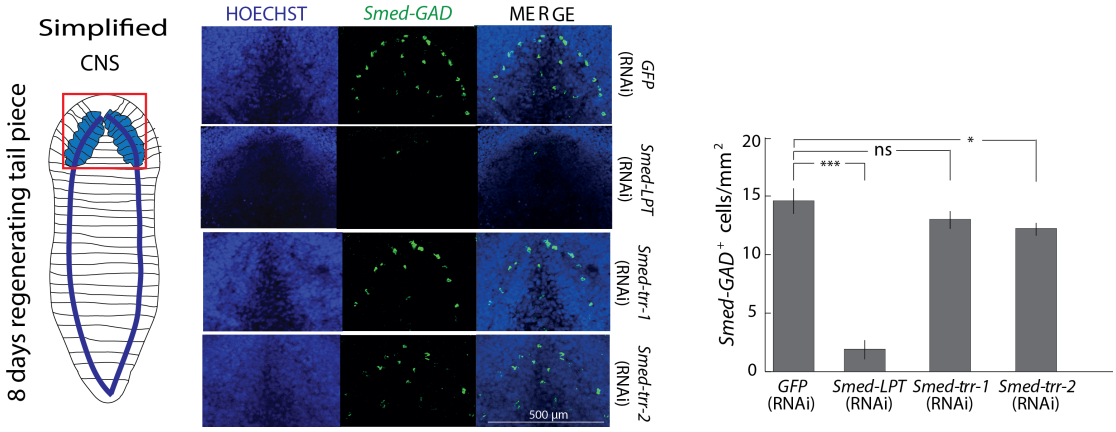


Figure 4.6. GABAergic neuron production is decreased following trithorax-related genes knockdown. A simplified sketch of the planarian CNS is presented (2 cephalic lobes, 2 VNCs and horizontal enervation). The red rectangle shows imaged area in the samples. Eight-day regenerating tail pieces from knockdown worms were tested for *Smed-GAD*-positive neurons. Images are representative of 2 experiments using ten worms per condition each. The bar graph shows the mean number of GABAergic neurons normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with *** for $p<0.001$, with * for $p<0.05$ and 'ns' for 'not significant'. Error bars represent SEM.

The next neuronal subtype under investigation was dopaminergic neurons as labeled by *Smed-TH* (tyrosine hydroxylase; converts L-tyrosine to L-DOPA), a marker published in Fraguas et al. (2011). Regenerating tail pieces were used in this analysis and two regeneration timepoints were tested – day 5 and day 8 (**Figure 7A** and **7B**). At day 5 *Smed-LPT* and *Smed-trr-1* knockdown regenerates showed significantly smaller amount of recovered dopaminergic neurons. At day 8, *Smed-LPT*(RNAi) animals still had significantly less *Smed-TH*-positive cells, while *Smed-trr-1*(RNAi) planarians had caught up with controls, unlike *Smed-trr-2*(RNAi) planarians which now had significantly less dopaminergic neurons compared to control. Results for the day 8 regeneration timepoint were also reproduced in a colorimetric *in situ* hybridisation experiment.

In summary, the production of dopaminergic neurons following knockdown of a trithorax-related gene was negatively affected.

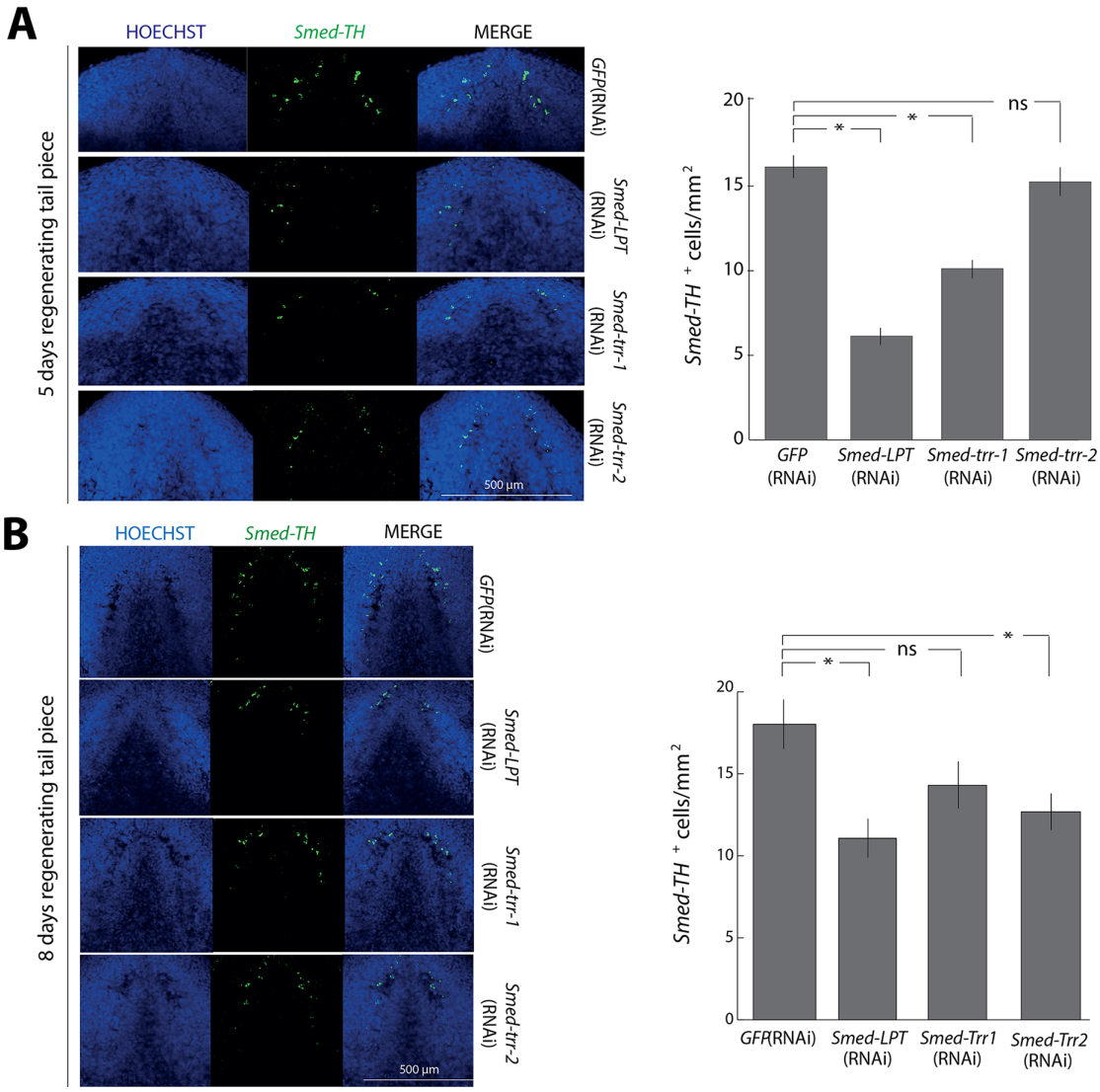


Figure 4.7. Dopaminergic neuron production is decreased following trithorax-related genes knockdown. (**A**) Five-day regenerating tail pieces from knockdown worms were tested for *Smed-TH*-positive neurons. Ten worms per condition were used. The bar graph shows mean numbers of dopaminergic neurons normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05. (**B**) Eight-day regenerating tail pieces from knockdown worms were tested for *Smed-TH*-positive neurons. Images are representative of 2 experiments using ten worms per condition each. The bar graph shows mean numbers of dopaminergic neurons normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05 and 'ns' for 'not significant'. Error bars represent SEM.

The production of acetylcholinergic neurons was also investigated under trithorax-related genes knockdown conditions. Early (3 days) and late (8 days) regeneration timepoints were used. Acetylcholinergic neurons were detected via *in situ* hybridisation with the marker *Smed-chat* (choline acetyl-transferase homolog) used in Wagner et al. (2011). It is known that *Smed-chat* labels differentiating neurons of different lineages and fully differentiated acetylcholinergic neurons (Wagner et al., 2011). To make use of this property, SMEDWI-1 antibody was employed in combination with *Smed-chat* RNA probe. Cells that are positive for both *Smed-chat* and SMEDWI-1 should be neuronal progenitor cells, while cells that are only positive for *Smed-chat* should represent the terminally differentiated acetylcholinergic neurons. There were no significant differences between knockdown conditions at day three of regeneration (**Figure 4.8A**). At day 8 of regeneration, *Smed-LPT*(RNAi) animals had significantly less terminally differentiated acetylcholinergic neurons compared to controls (**Figure 4.8B**). There were no observed differences in neuronal progenitor cell populations across conditions. The results for day 8 regeneration were replicated in a colorimetric *in situ* procedure with *Smed-chat*.

The current findings could mean that the initial recovery of acetylcholinergic neurons is intact in *Smed-LPT*(RNAi) animals, but the neurons' maintenance and continued production at later stages of regeneration suffers. Also, early neuronal progenitors seemed to be unaffected after trithorax-related genes knockdown, meaning that the defects in neuronal production probably happen at later precursor differentiation stages (cells no longer positive for SMEDWI-1).

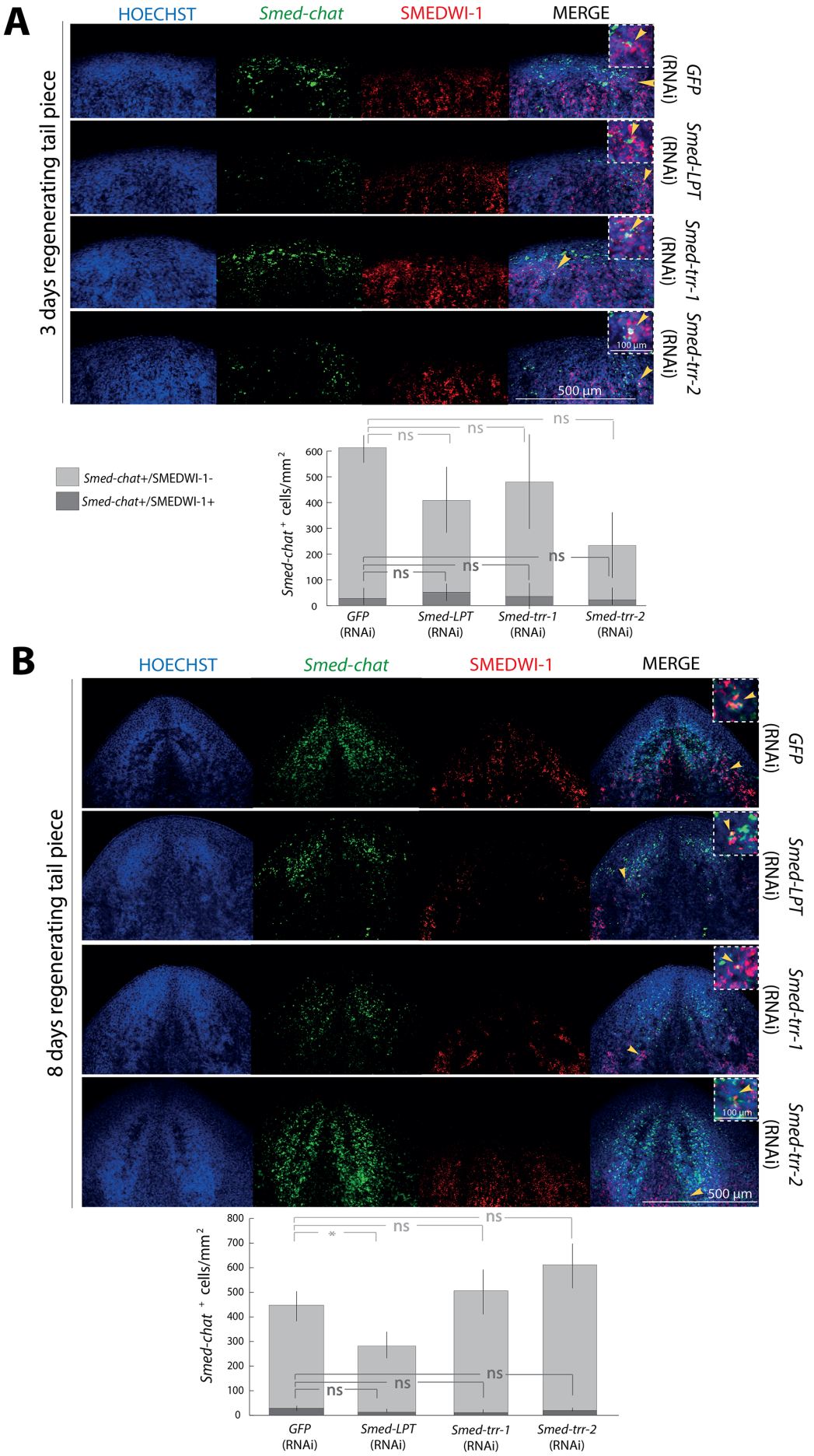

A
HOECHST
Smed-chat
SMEDWI-1
MERGE
3 days regenerating tail piece
GFP (RNAi)
Smed-LPT (RNAi)
Smed-trr-1 (RNAi)
Smed-trr-2 (RNAi)
100 µm
500 µm
Smed-chat+/SMEDWI-1-
Smed-chat+/SMEDWI-1+
Smed-chat+ cells/mm2
ns
B
8 days regenerating tail piece
*

Figure 4.8. Acetylcholinergic neuron production is decreased following trithorax-related genes knockdown. (**A**) Three-day regenerating tail pieces from knockdown worms were tested for *Smed-chat*-positive and SMEDWI-1-positive cells. The double positive cells represent neuronal precursors, while those positive only for *Smed-chat* are terminally differentiated acetylcholinergic neurons. Yellow arrows show double-positive cells. The top right corner of every image has a zoomed in area with a double-positive cell. Ten worms per condition were used. The bar graph shows mean numbers of acetylcholinergic neurons normalised per imaged area in light gray. The dark gray colour represents mean neuronal precursor numbers normalised per imaged area. 2-tailed ttests assuming unequal variance were used. Significance is noted with * for $p<0.05$ and 'ns' for 'not significant'. (**B**) The same analysis as in (**A**), but on day 8 regenerates. Images are representative of 2 experiments using ten worms per condition each. Error bars represent SEM.

The final neuronal subtype under investigation was serotonergic neurons. The study was performed via *in situ* hybridisation experiments with *Smed-TPH* (tryptophan hydroxylase; involved in serotonin production) used in Fraguas et al. (2011). Three- and eight-day regenerating tail pieces were used for this analysis. The presence of serotonergic neurons was significantly decreased in *Smed-LPT*(RNAi) animals in both regeneration timepoints (**Figure 4.9A** and **Figure 4.9B**). A colorimetric *in situ* hybridisation with *Smed-TPH* at day 8 regeneration tail pieces confirmed the reduction of signal in *Smed-LPT*(RNAi) animals.

These findings probably indicate that following *Smed-LPT* knockdown the initial serotonergic neuron-producing response is compromised and this defect gets amplified further at later stages of regeneration.

Smed-TPH was reported to have an important role in melanin production in the planarian photoreceptor pigment cups (Lambrus et al., 2015). In **Figure 4.9B** *Smed-*

TPH is seen to mark the pigment cups of developed photoreceptors. The most notable abnormality in photoreceptors is seen in *Smed-LPT*(RNAi) animals where the two photoreceptors seem to be underdeveloped compared to controls. This explains why when *Smed-LPT*(RNAi) animals form eyes they are usually faint.

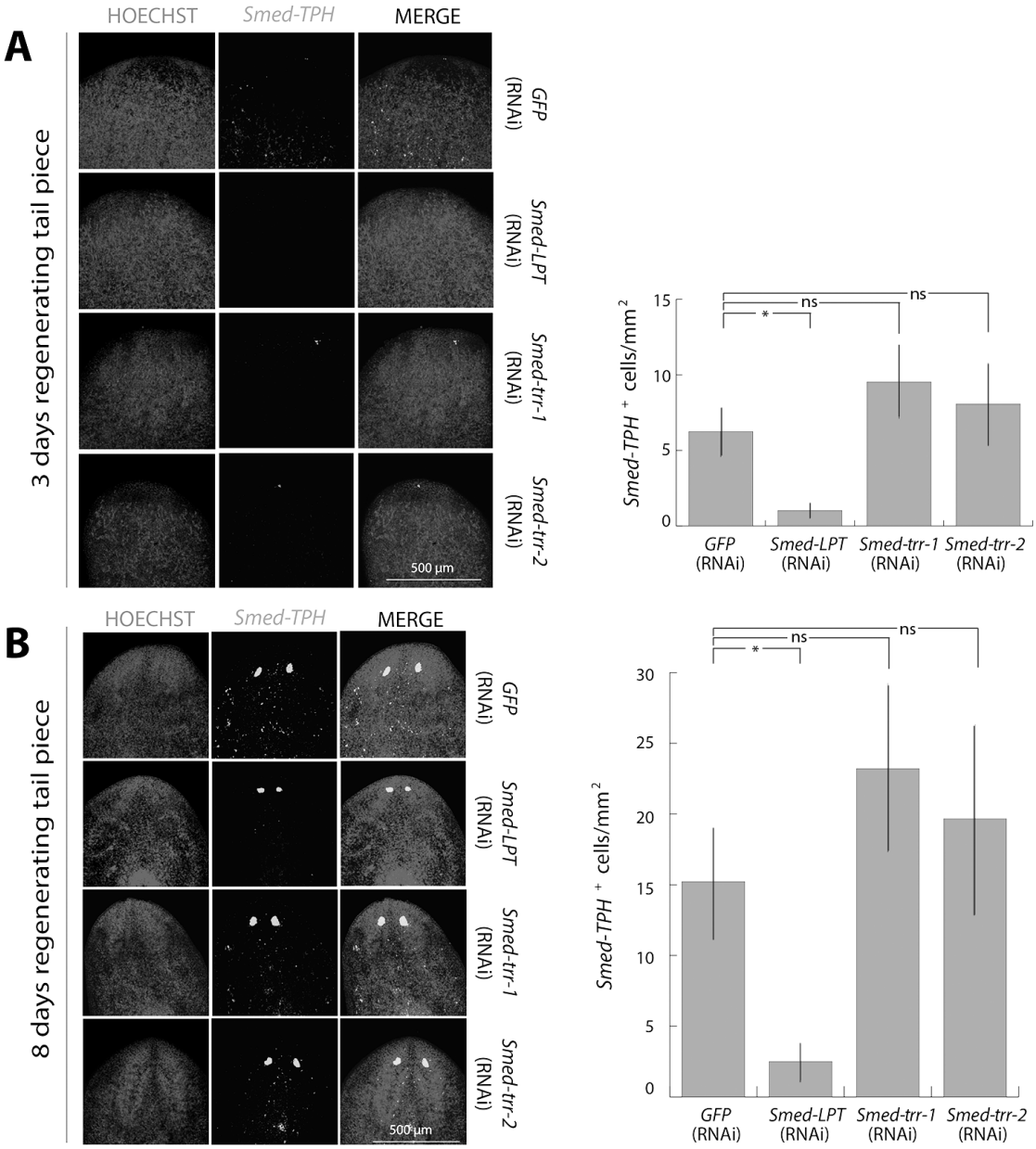


Figure 4.9. Serotonergic neuron production is decreased following trithorax-related genes knockdown. (**A**) Three-day regenerating tail pieces from knockdown worms were tested for *Smed-TPH*-positive neurons. Ten worms per condition were used. The bar graph shows mean numbers of serotonergic neurons normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05 and 'ns' for 'not significant'. (**B**) Eight-day regenerating tail pieces from knockdown worms were tested for *Smed-TPH*-positive neurons. Images are representative of two experiments with 10 worms per condition each. The bar graph shows number of serotonergic neurons normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05 and 'ns' for 'not significant'. Error bars represent SEM.

The serotonin neuron deficiency following *Smed-LPT*(RNAi) was further investigated. *Smed-TPH* labels terminally differentiated serotonergic neurons, while *Smed-pitx* (*Pituitary homeobox transcription factor*) is expressed in terminally differentiated serotonergic neurons, as well as their progenitor cells (Currie and Pearson, 2013; März et al., 2013). An experiment studying the production of *Smed-pitx*-positive cells during regeneration showed that there was a significant increase in the amount of *Smed-pitx*-positive cells in the anterior blastema of *Smed-LPT*(RNAi) regenerates compared to control. This was true for both early (3 days) and late (8 days) regeneration timepoints (**Figure 4.10A** and **Figure 4.10B**). No significant difference from control was observed for *Smed-trr-1* and *Smed-trr-2* knockdown animals.

Smed-pitx was suggested to act as terminal selector for planarian serotonergic neurons (Currie and Pearson, 2013; März et al., 2013). More specifically, it regulates the differentiation of a late stem cell progeny cell into a fully differentiated serotonergic neuron, as well as the establishment and maintenance of its serotonergic function (März et al., 2013). The present findings could indicate that knockdown of *Smed-LPT* results in intact production of *Smed-pitx*-positive serotonergic neuron precursor cells, which are unable to undergo terminal differentiation, hence the decrease in *Smed-TPH*-positive cells (**Figure 4.9**). As a result of this differentiation failure, *Smed-pitx*-positive cells accumulate. The results also demonstrate that *Smed-pitx* expression is not sufficient for progressing to serotonergic neuronal terminal differentiation.

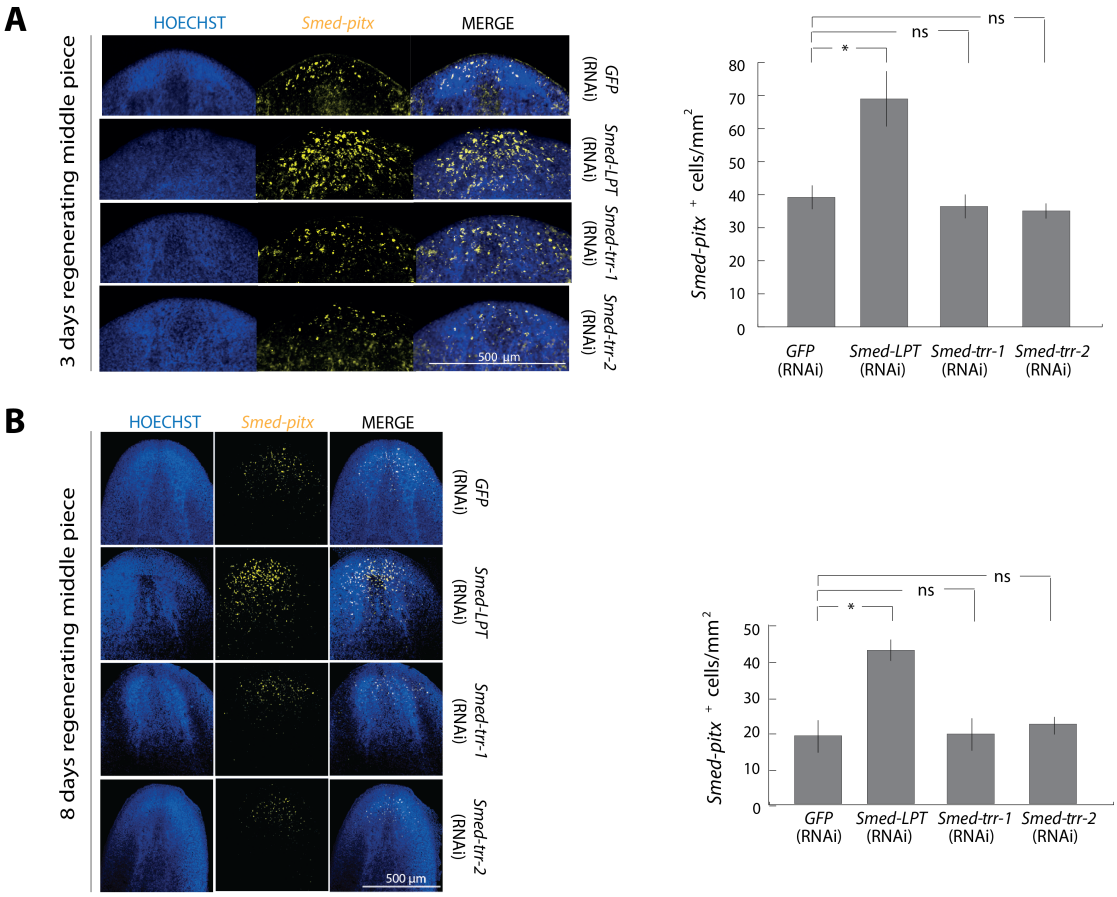


Figure 4.10. *Smed-pitx*-positive cell production is increased following trithorax-related genes knockdown. (**A**) Three-day regenerating tail pieces from knockdown worms were tested for *Smed-pitx*-positive cells. Ten worms per condition were used. The bar graph shows mean numbers of *Smed-pitx*-positive cells normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05 and 'ns' for 'not significant'. (**B**) Eight-day regenerating tail pieces from knockdown worms were tested for *Smed-pitx*-positive neurons. Ten worms per condition were used. The bar graph shows mean numbers of *Smed-pitx*-positive cells normalised per imaged area. A 2-tailed ttest assuming unequal variance was used. Significance is noted with * for p<0.05 and 'ns' for 'not significant'. Error bars represent SEM.

The overall size of regenerated brains in *Smed-LPT*(RNAi) was visibly smaller and more under-developed than in corresponding control animals (based on Hoechst staining in experiments presented so far). This was further confirmed by the defects seen in different neuronal subpopulations in this experimental condition. These findings could be, at least partially, explained by the speculation that lack of *Smed-LPT* results in failure of LPT-Trr complex to act as a co-activator to the planarian TLX-1 receptor. Previous findings have shown that *Smed-tlx-1* is highly expressed in the head and knockdown of *Smed-tlx-1* results in difficulties maintaining head structures (Raška et al., 2011). Furthermore, the importance of retinoic acid in head regeneration was demonstrated (Ermakova et al., 2009), suggesting that the planarian retinoid X receptor might also be affected by *Smed-LPT*(RNAi).

The much weaker phenotypic effects of *Smed-trr-1* and *Smed-trr-2* knockdown could be explained by functional redundancy, in agreement with Hubert et al. (2013).

4.4.4. Motility defects in trithorax-related genes knockdown animals are due to serotonin deficiency

Currie and Pearson (2013) demonstrated that planarian motility is dependent on serotonin signaling regulating the proper beating of the ventral cilia. If serotonin signaling was perturbed, planarians were only capable of performing muscle-based motion resulting in inchworming (Currie and Pearson, 2013). Hence, the inchworming locomotion that was observed in animals following trithorax-related genes knockdown could be due to two potential causes – lack of cilia or lack of serotonergic control of cilia beating (or both). The first hypothesis was tested with an

antibody against acetylated tubulin, labeling planarian ciliated cells. Both dorsal and ventral cilia recovery and homeostasis were unaffected following trithorax-related gene knockdown (**Figures 4.11A** and **Figure 4.11B**).

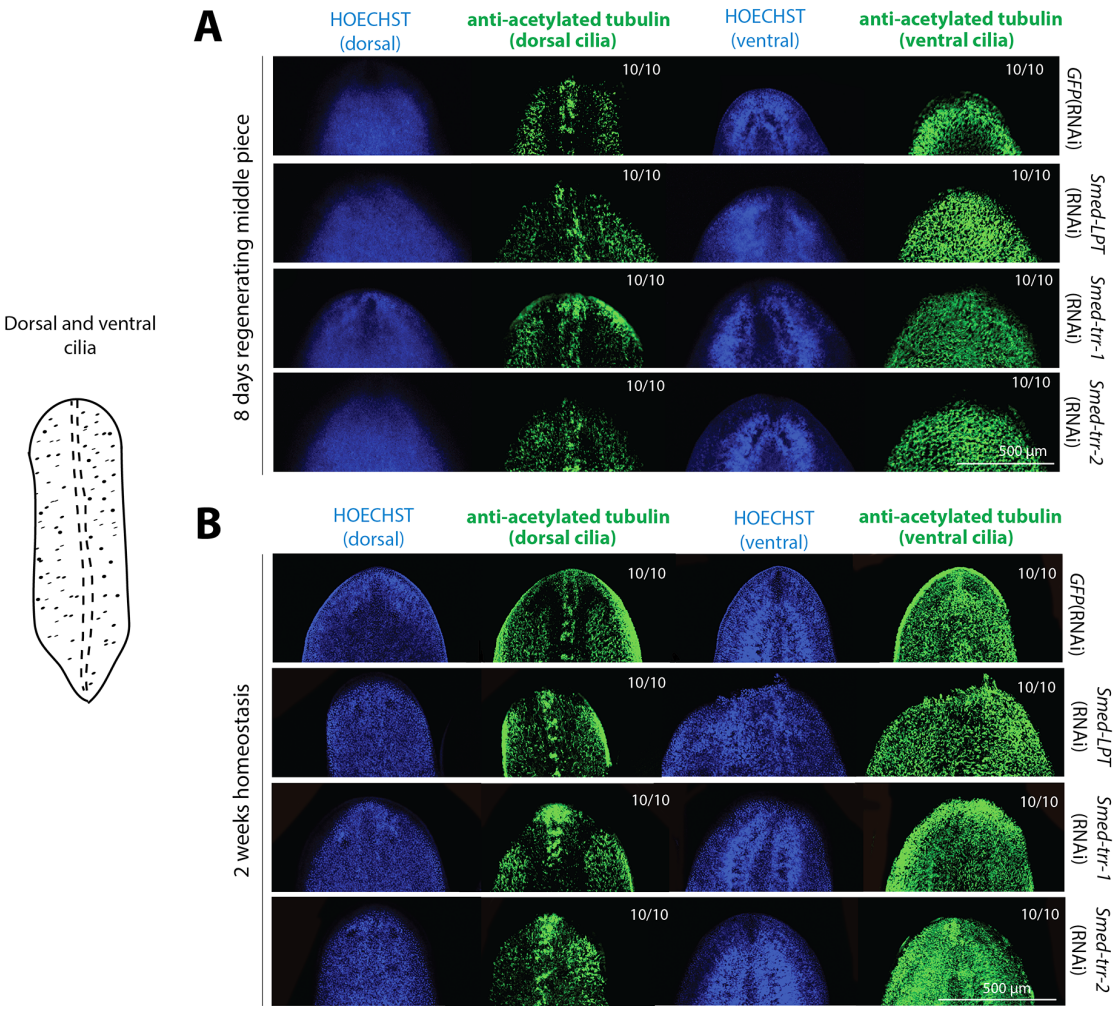
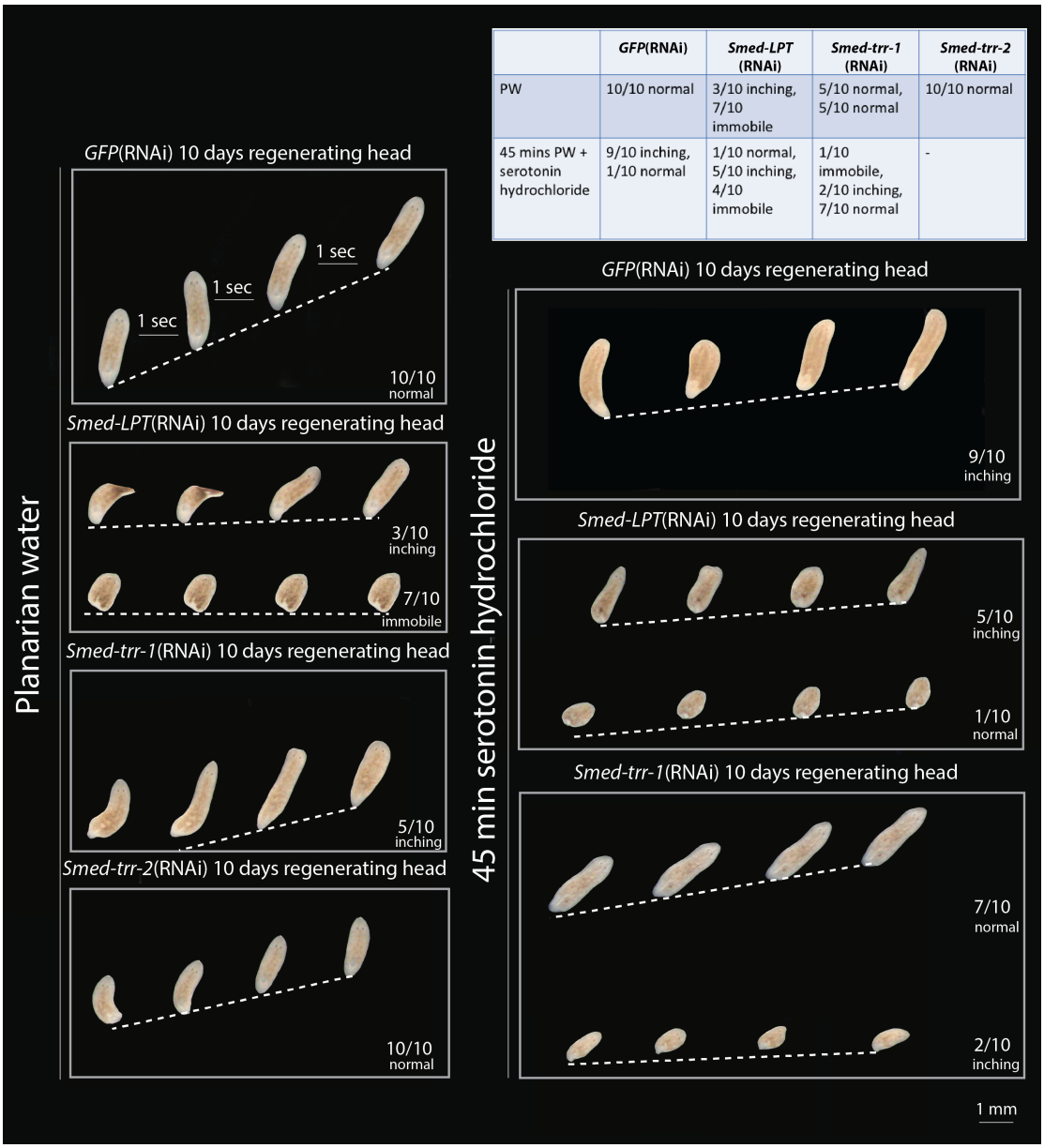


Figure 4.11. Cilia regeneration and homeostasis are unaffected following knockdown of trithorax-related genes. A schematic illustrating the positioning of ventral and dorsal cilia in the planarian body is shown. (**A**) 8-day regenerating middle pieces stained with anti-acetylated tubulin antibody in order to visualise dorsal and ventral cilia. Ten worms per conditions were used. (**B**) 2 weeks homeostatic animals stained with anti-acetylated tubulin antibody in order to visualise dorsal and ventral cilia. White numbers at the top right corner of each image indicate proportion of animals per sample that showed the same characteristics as the imaged animal. Ten worms per condition were used.

The next step was to test whether the inchworming locomotion was due to the fact that regenerating planarians that had undergone trithorax-related genes knockdown had fewer fully differentiated serotonergic neurons than controls (*Smed-LPT* knockdown animals in **Figure 4.9**). Having fewer serotonergic neurons should lead to decreased amount of serotonin production leading to unsynchronised cilia beating and inchworming. In order to test this hypothesis, *Smed-LPT* and *Smed-trr-1* knockdown animals were subjected to ectopic serotonin hydrochloride treatment, as described by Currie and Pearson (2013) in an attempt to reverse the inchworming phenotype. *Smed-trr-2* knockdown animals were not subjected to the treatment because the inchworming behaviour was not as prevalent in them and some batches of RNAi animals did not exhibit it at all. First, 10-day regenerating *Smed-LPT*(RNAi) and *Smed-trr-1*(RNAi) middle pieces underwent 45 minutes of 100 μM serotonin hydrochloride (dissolved in planarian water) treatment. No improvement was observed in the experimental conditions' locomotion (data not shown). It was hypothesised that the regenerating middle pieces already had a very limited amount of serotonin neurons so the ectopically introduced serotonin might not had been utilised very effectively. Therefore, it was decided that the treatment should be administered to regenerating head pieces, which still show inchworming behaviour, but should possess a much more functional and better preserved brain (and serotonergic neurons) than the middle pieces. The results from this experiment are shown in **Figure 4.12**. The blue table shows number of animals exhibiting a specific locomotor phenotype before and after serotonin hydrochloride treatment. Still shots (with one second interval between them) from movies before and after the treatment are presented. The angle of the dashed line indicates the locomotive

progress that the worm had achieved for three seconds (represented in 4 still shots) – the steeper the slope, the greater the progress. Since serotonin excess could also result in disturbance of cilia beating (Currie and Pearson, 2013), nine out of ten serotonin hydrochloride-treated *GFP*(RNAi) animals exhibited inchworming behaviour. The serotonin hydrochloride treatment mildly improved the locomotive defects of *Smed-LPT* and *Smed-trr-1* knockdown animals. Three out of ten *Smed-LPT*(RNAi) worms were inching and 7/10 were immobile before the treatment. After the treatment, only 4/10 were immobile, 5/10 were inching and one out of ten was moving normally. Similar improvements were seen for *Smed-trr-1* knockdown worms as well. Five out of ten worms were inching before treatment, which was reduced to 2/10 and 1/10 immobile after the treatment.

In summary, planarian inchworming behaviour following knockdown of trithorax-related genes was related to insufficient serotonin neurotransmission due to the lower amount of serotonergic cells being produced. This conclusion seems to best explain the defects in *Smed-LPT*(RNAi) animals. The locomotive problems in *Smed-trr-1* knockdown animals are slightly harder to reconcile with the fact that no significant decrease in serotonergic neuronal production was observed for them. This could potentially be explained by a couple of reasons. One possibility is that serotonergic neuronal production was in fact negatively affected by *Smed-trr-1* knockdown, but since the defect was more subtle than in *Smed-LPT*(RNAi), the chosen *Smed-TPH in situ* hybridisation assay was not sensitive enough to detect that effect. Alternatively, it is also possible that in *Smed-trr-1* (and *Smed-trr-2*) knockdown animals serotonergic neurons are being made, as indicated by *Smed-TPH* staining in **Figure 4.9**, but they are not fully functional.



	GFP(RNAi)	*Smed-LPT* (RNAi)	*Smed-trr-1* (RNAi)	*Smed-trr-2* (RNAi)
PW	10/10 normal	3/10 inching, 7/10 immobile	5/10 normal, 5/10 normal	10/10 normal
45 mins PW + serotonin hydrochloride	9/10 inching, 1/10 normal	1/10 normal, 5/10 inching, 4/10 immobile	1/10 immobile, 2/10 inching, 7/10 normal	-

Figure 4.12. Ectopic serotonin treatment improves mobility of trithorax-related knockdown animals. Locomotive defects in trithorax-related knockdown animals are shown. The blue table shows the number of regenerating head pieces per condition exhibiting the respective locomotive defect before and after 45-minute-long serotonin hydrochloride treatment. Still shots from movies are shown for worms in each condition. There was one second interval between the shots. The dashed line indicates the locomotive progress that the worm has achieved in 3 seconds (from first to last shot) – the steeper the line, the greater the progress. White numbers at the bottom right corners indicate number of animals from the respective condition showing the illustrated behaviour. Ten worms per condition were used.

4.4.5. Pharynx, gut and epidermis, but not protonephridia, are affected following knockdown of trithorax-related genes

The recovery after amputation and maintenance during homeostasis of tissues other than the brain and cilia were investigated. At day three of regeneration, *Smed-laminin* staining (Cebrià and Newmark, 2007) was clearly localised at the site of the newly forming pharynx in control animals (**Figure 4.13A**). Such localisation was much less obvious in *Smed-trr-2*(RNAi) animals and completely absent in *Smed-LPT*(RNAi) and *Smed-trr1*(RNAi) animals. At day 8 of regeneration, when the pharynx was fully formed in controls, 9/10 *Smed-trr-2*(RNAi) animals had pharynx comparable to controls, while 1/10 had an underdeveloped pharynx (**Figure 4.13B**). Seven out of ten *Smed-trr-1*(RNAi) animals had an underdeveloped pharynx and 3/10 had no pharynx. The pharyngeal defects were most severe in *Smed-LPT*(RNAi) animals, where 8/10 worms failed to form a pharynx by day 8 of regeneration, while the other 2/10 animals had underdeveloped pharynges. The day 8 regeneration *Smed-laminin* stain experiment was performed twice and the results were comparable.

These findings confirm previous brightfield observations that trithorax-related knockdown animals had difficulties regenerating a pharynx. The next logical step was to confirm the pharyngeal defect in homeostatic tissue maintenance conditions. For this purpose, 2 weeks homeostasis knockdown animals were stained with anti-synapsin antibody labeling the CNS and the pharynx (**Figure 4.13C**). In 4/10 *Smed-LPT*(RNAi) animals the pharynx was lost by the end of the second week following knockdown. No differences were observed for *Smed-trr-1* and *Smed-trr-2* knockdown animals.

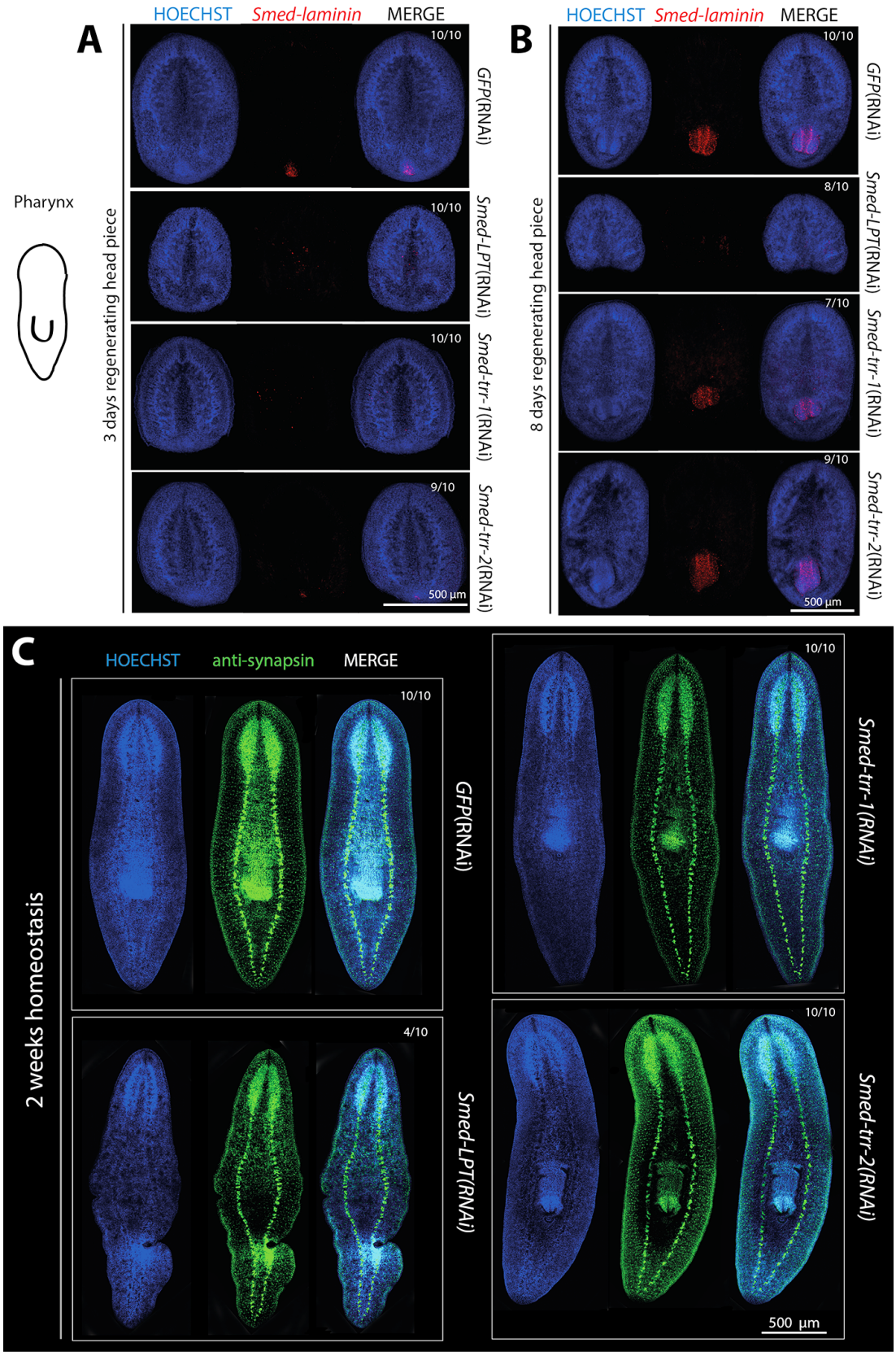

A
HOECHST
Smed-laminin
MERGE
Pharynx
3 days regenerating head piece
10/10
GFP(RNAi)
10/10
Smed-LPT(RNAi)
10/10
Smed-trr-1(RNAi)
9/10
Smed-trr-2(RNAi)
500 µm
B
HOECHST
Smed-laminin
MERGE
8 days regenerating head piece
10/10
GFP(RNAi)
8/10
Smed-LPT(RNAi)
7/10
Smed-trr-1(RNAi)
9/10
Smed-trr-2(RNAi)
500 µm
C
HOECHST
anti-synapsin
MERGE
2 weeks homeostasis
10/10
GFP(RNAi)
4/10
Smed-LPT(RNAi)
10/10
Smed-trr-1(RNAi)
10/10
Smed-trr-2(RNAi)
500 µm

Figure 4.13. Pharynx regeneration and homeostatic maintenance are compromised following trithorax-related genes knockdown. A schematic illustrating the pharyngeal position in a planarian is shown. (**A**) *Smed-laminin* staining of 3-day-regenerating head pieces. (**B**) *Smed-laminin* staining of 8-day-regenerating head pieces. Images are representative of two experiments. (**C**) Anti-synapsin antibody staining of two weeks homeostatic knockdown animals. White numbers at the top right corner of each merged image represent the proportion of animals exhibiting the same characteristics within each condition. Ten worms per condition were used.

The next observed structure was the planarian gut. Regenerating middle pieces were used for this assay. They were chosen because they allowed us to observe recovery of the anterior gut branch, as well as the two posterior branches. As early as 3 days after amputation, there were no discernible defects in any of the knockdown animals (**Figure 4.14A**). At day 8 of regeneration, however, only 3/10 *Smed-LPT*(RNAi) animals looked similar to control, the rest 7/10 suffered severe defects in gut recovery (**Figure 4.14B**). The branch definition was starting to become less clear and the gut arborisation appeared simplified. No noticeable differences from control were observed for *Smed-trr-1* and *Smed-trr-2* knockdown animals. This result was confirmed when a later regeneration timepoint (day 16) was tested (**Figure 4.14C**). The same results were achieved over 2 separate experiments on day 3 and day 8 regenerates (4 experiments altogether), each confirming the results shown in **Figure 4.14A** and **Figure 4.14B**. Ten worms per condition were used each time.

The observations made on gut maintenance during homeostasis knockdown conditions are in agreement with the findings from the regeneration studies. At 2 weeks of homeostasis 50% of the *Smed-LPT*(RNAi) animals showed degradation of

the existing gut (**Figure 4.14D**). Again, no noticeable differences from control were observed for *Smed-trr-1* and *Smed-trr-2* knockdown planarians.

One hypothesis that could explain the gut regeneration and maintenance defects in *Smed-LPT*(RNAi) animals is that the LPT-Trr complex in planarians is a co-activator of the nuclear receptor *Smed-hnf4.* It was previously shown that *Smed-hnf4* is expressed in gut progenitors as well as fully differentiated gut cells (Scimone et al., 2014). It was reported that ASC-2 (part of the mammalian MLL3/4 complexes) binds to HNF4α receptor in human cell lines (Surapureddi et al., 2008), meaning that the mammalian MLL3/4 complex serves as its co-activator. Unfortunately, no *Smed-hnf4* knockdown experiments had been reported in planarians, so there is no confirmation that malfunction of *Smed-hnf4* would lead to failure to regenerate and/or maintain the planarian gut.

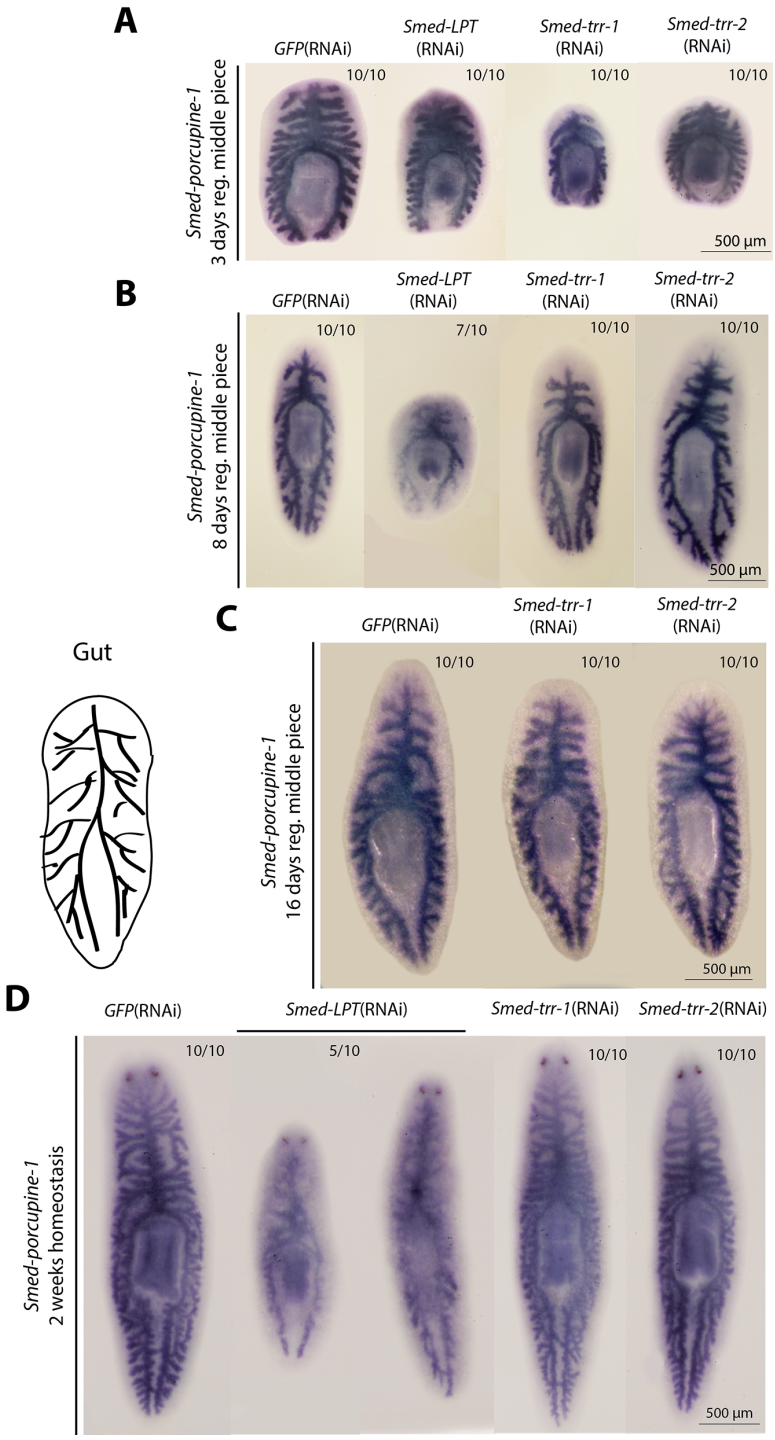


Figure 4.14. Gut regeneration and homeostatic maintenance are affected following knockdown of trithorax-related genes. A schematic illustrating the planarian gut structure is shown. (**A**) Day 3 regenerating middle pieces stained with *Smed-porcupine-1*. Ten worms per condition were used. (**B**) Day 8 regenerating middle pieces stained with *Smed-porcupine-1*. Ten worms per condition were used. (**C**) Day 16 regenerating middle pieces stained with *Smed-porcupine-1*. Ten worms per condition were used. (**D**) 2 weeks homeostasis knockdown animals stained with *Smed-porcupine-1*. Ten worms per condition were used. Black numbers at the top right corner of images indicate proportion of animals per condition with the same characteristics.

Protonephridia regeneration and maintenance was also studied under trithorax-related gene knockdown conditions. *Smed-CAVII-1* (carbonic anhydrase) (Rink et al., 2011; Scimone et al., 2011) labeling protonephridia tubule-associated cells was used in these experiments. There was no significant difference between knockdowns and controls (**Figure 4.15A**, **Figure 4.15B**, **Figure 4.15C**). These findings indicate that in planarians there appears to be no conservation of MLL3's regulatory functions of the excretory system, reported in Lee et al. (2009). This is somewhat surprising, since a small proportion of *Smed-trr-2* knockdown animals exhibited a slight posterior bloating, usually characteristic of problems with the excretory system (Reddien et al., 2005a). Rink et al. (2011) and Scimone et al. (2011) have published extensive studies on the organisation of flatworm protonephridia, focusing on tubules and 2 different tubule-associated cell types (Scimone et al., 2011). *Smed-CAVII-1* only labels one type of tubule-associated cells. We cannot exclude that other protonephridia structures are affected following knockdown of trithorax-related genes. Regardless of whether other protonephridia cells were affected, planarian cilia and *Smed-CAVII-1*-positive cells were not influenced by the knockdown of trithorax-related genes. These findings indicate that trithorax-related gene family does not act as a global controller of differentiation, but is rather fine-tuned to certain tissues and cell types.

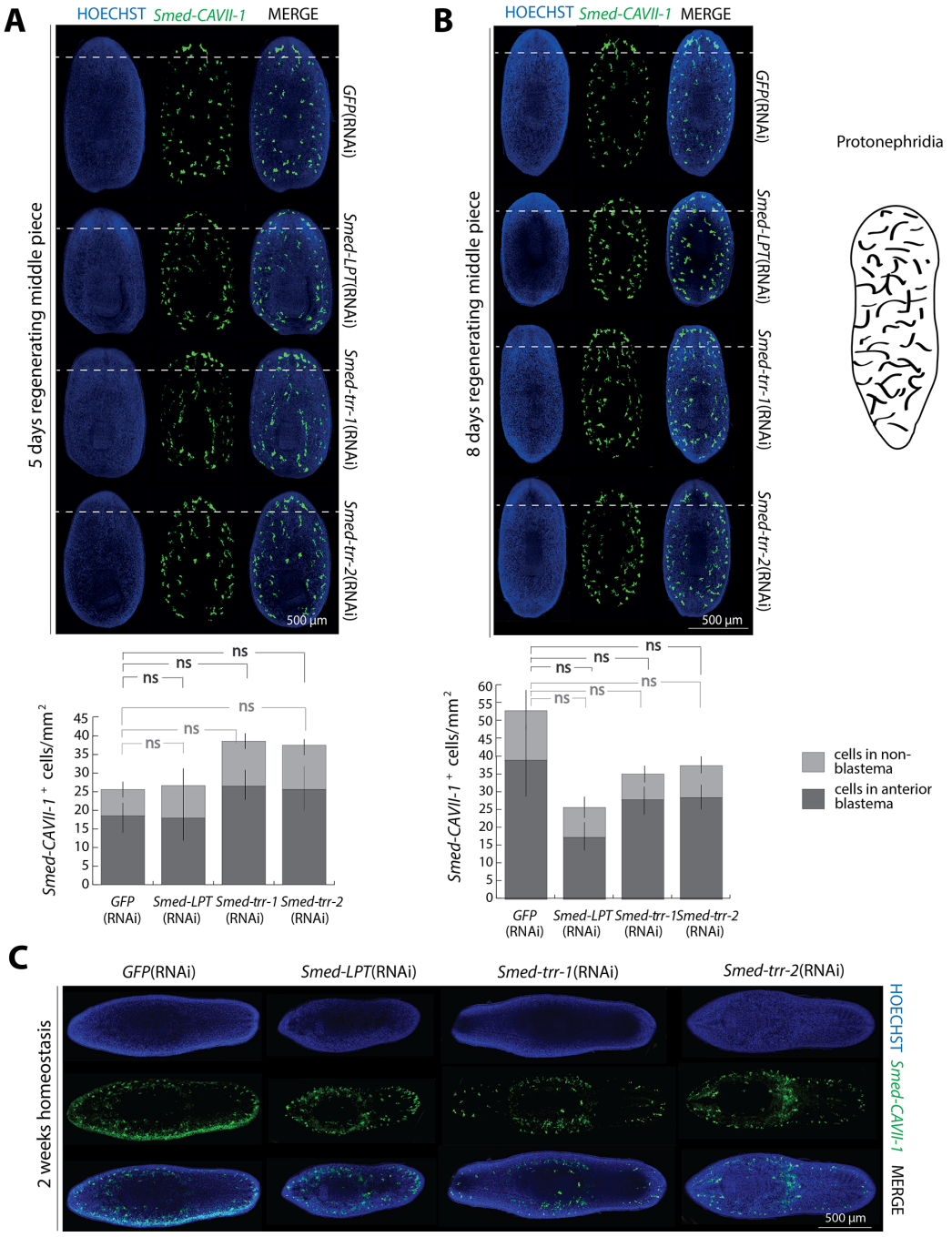


Figure 4.15. Protonephridia regeneration and homeostasis are unaffected following trithorax-related genes knockdown. A schematic illustrating the planarian protonephridia organisation is shown. (**A**) *Smed-CAVII-1 in situ* hybridisation on day 5 regenerating middle pieces. Ten worms per condition were used. White dashed line indicates where the anterior blastema ends. The bar graph presents the mean numbers of *Smed-CAVII-1*-positive cells normalised per area. Numbers are presented for the blastema region (dark gray) and non-blastema region (light gray). A 2-tailed ttest assuming unequal variance was performed where 'ns' stands for 'not significant'. (**B**) *Smed-CAVII-1 in situ* hybridisation performed on day 8 regenerating middle pieces. Ten worms per condition were used. A 2-tailed ttest assuming unequal variance was performed where 'ns' stands for 'not significant'. (**C**) *Smed-CAVII-1 in situ* hybridisation performed on 2 weeks homeostatic knockdown animals. Ten worms per condition were used.

The final experiment in this chapter sheds light on trithorax-related genes' involvement in the production of epidermal progenitor cells. The markers used in these experiments - *Smed-NB.21.11e* and *Smed-AGAT-1,* were first identified by Eisenhoffer et al. (2008). *Smed-NB.21.11e* is expressed in early stem cell progeny committed to the epidermal lineage, while *Smed-AGAT-1* is expressed in late epidermal stem cell progeny. At day 8 of regeneration both *Smed-NB.21.11e*-positive cells and *Smed-AGAT-1*-positive cells were significantly decreased in *Smed-LPT*(RNAi) regenerating middle pieces when compared to controls (**Figure 4.16A**). There was no significant difference in these cell numbers in *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) animals when compared to control. The day 8 regeneration experiment was performed twice. No significant difference was observed in the 2 weeks homeostasis condition (**Figure 4.16B**).

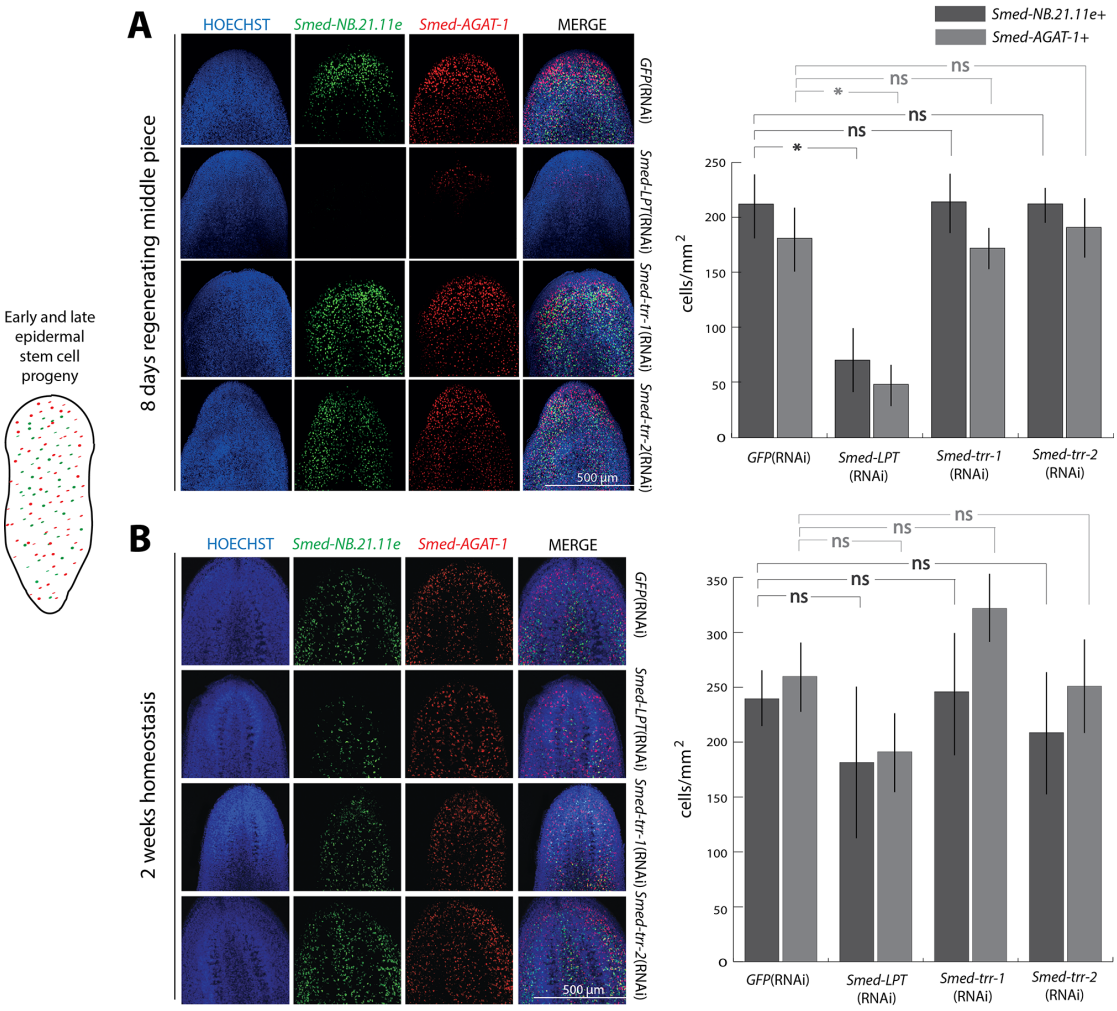


Figure 4.16. Epidermal progenitor cells regeneration and homeostasis are perturbed following trithorax-related genes knockdown. A schematic illustrating the organisation of *Smed-NB.21.11e*- and *Smed-AGAT-1*-positive cells is shown. (**A**) Day 8 regenerating middle pieces stained with *Smed-NB.21.11e* and *Smed-AGAT-1* markers. The bar graph illustrates the mean number of each of the two cell types normalised per imaged animal area – dark gray corresponds to *Smed-NB.21.11e*-positive cells, while light gray corresponds to *Smed-AGAT-1*-positive cells. Statistical comparisons were performed via 2-tailed ttests assuming unequal variance, where * signifies $p<0.05$ and 'ns' – 'not significant'. Images are representative of two separate experiments with 10 animals per condition each. (**B**) 2 weeks homeostasis knockdown animals stained with *Smed-NB.21.11e* and *Smed-AGAT-1* markers. The bar graph illustrates the mean number of each of the two cell types normalised per imaged animal area – dark gray corresponds to *Smed-NB.21.11e*-positive cells, while light gray corresponds to *Smed-AGAT-1*-positive cells. Statistical comparisons were performed via 2-tailed ttests assuming unequal variance where 'ns' denotes 'not significant'. Ten animals per condition were used.

4.5. CONCLUSION

The experiments presented in this chapter show that *Smed-LPT*(RNAi) has a much stronger effect than either of the two *trr* knockdowns. This trend is seen in both the brightfield observation experiments and the molecular marker staining experiments. A potential explanation for this is that *Smed-trr-1* and *Smed-trr-2* are (partially) redundant so when one of them is knocked down, the other is compensating. This hypothesis is in line with Hubert et al.'s (2013) findings. The experiments presented here also reveal that the differentiation phenotypes seen in *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* knockdown animals are not a consequence of stem cell maintenance. The experiments also show that trithorax-related genes are important for regenerative blastema formation, pharynx, gut, epidermis and brain tissue regeneration (**Figure 4.17**). They are not, however, involved in the production and maintenance of *Smed-CAVII-1*-positive protonephridia cells and dorsal and ventral cilia cells. These findings demonstrate that trithorax-related genes in planarians are not master-regulators of differentiation of all cell types, but only regulate cell differentiation in selected lineages.

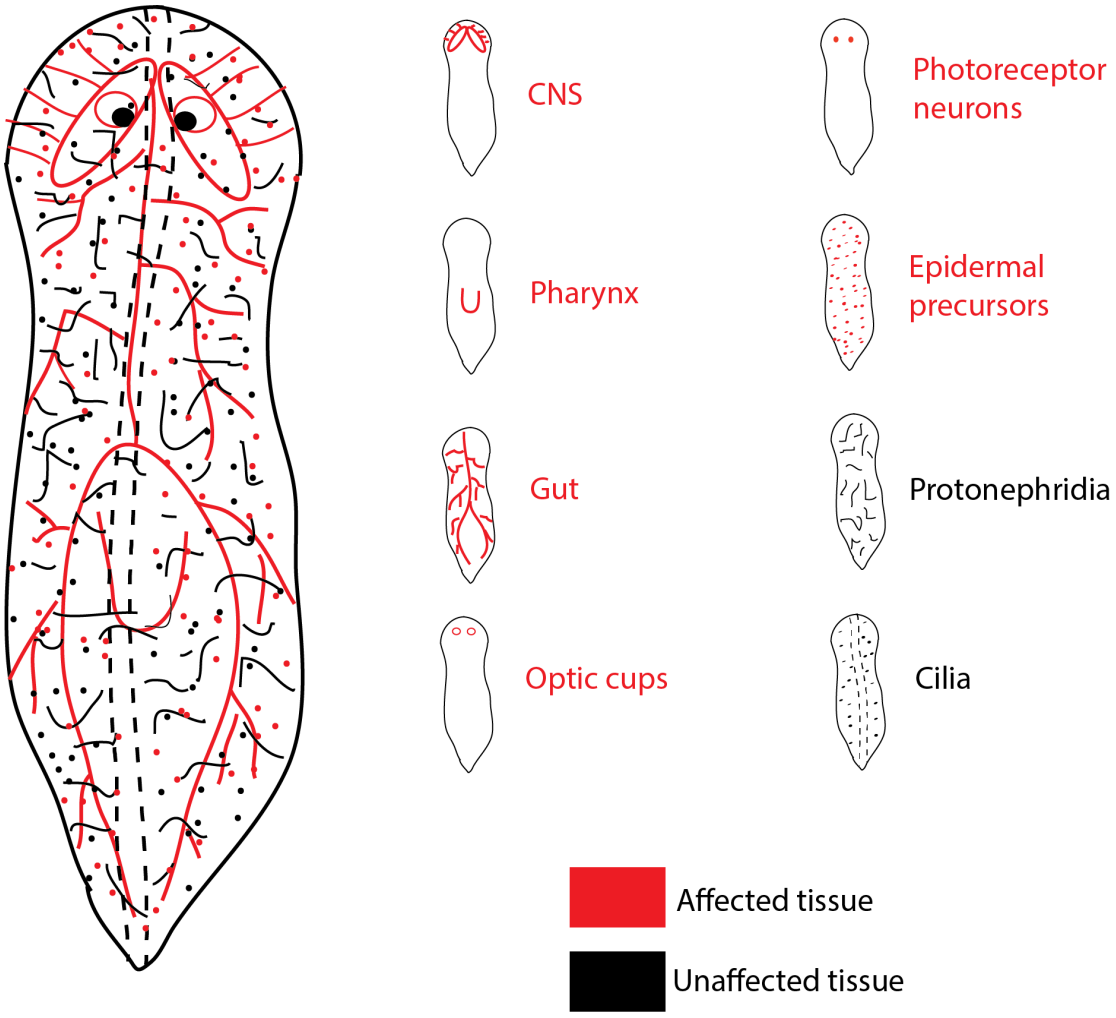


Figure 4.17. Summary of differentiation defects associated with trithorax-related genes knockdown. In red are tissues that are affected and in black are those unaffected by trithorax-related genes knockdown.

5. CANCEROUS FEATURES OF *SMED-LPT* KNOCKDOWN

5.1. ABSTRACT

Even though the histone methyltransferase *Mll3* is one of the most commonly mutated genes in human cancer, carrying both driver and passenger mutations, not many studies have focused on it. This chapter shows that knockdown of the gene *Smed-LPT* (corresponding to the N-terminus of *Mll3*) led to hyperproliferation of stem cells and the formation of tumour-like outgrowths in planarians. The outgrowths often contained stem cells and aberrant gene expression of *Smed-NB.21.11e* and *Smed-AGAT-1*. These defects were correlated with epidermal disorganisation and abnormalities in nuclear shape. *Smed-LPT* was also found to control stem cell migration. The data presented on *Smed-LPT* suggests that trithorax-related genes (and *Mll3*) may have a conserved role in maintaining the correct balance between stem cell proliferation and differentiation. This would explain *Mll3*'s persistent identification as a cancer driver gene.

5.2. INTRODUCTION

Human tumours evolve over years and are dependent on the gradual accumulation of mutations transforming them from benign to malignant structures (Vogelstein et al., 2013). A comprehensive attempt to synthesise the 'hallmarks of cancer' (Hanahan and Weinberg, 2011) presented the primary and auxiliary changes in tumour behaviour and capability. Hanahan and Weinberg (2011) also emphasised

the importance of two so-called ‘enabling characteristics’ of cancer – genomic instability (increased amount of mutations) and the inflammatory state of pre-malignant and malignant sites (tumour progression is often aided by inflammation). The hallmarks of cancer as defined by Hanahan and Weinberg (2011), alongside with two other characteristics, not mentioned in their article are illustrated below (**Figure 5.1**).

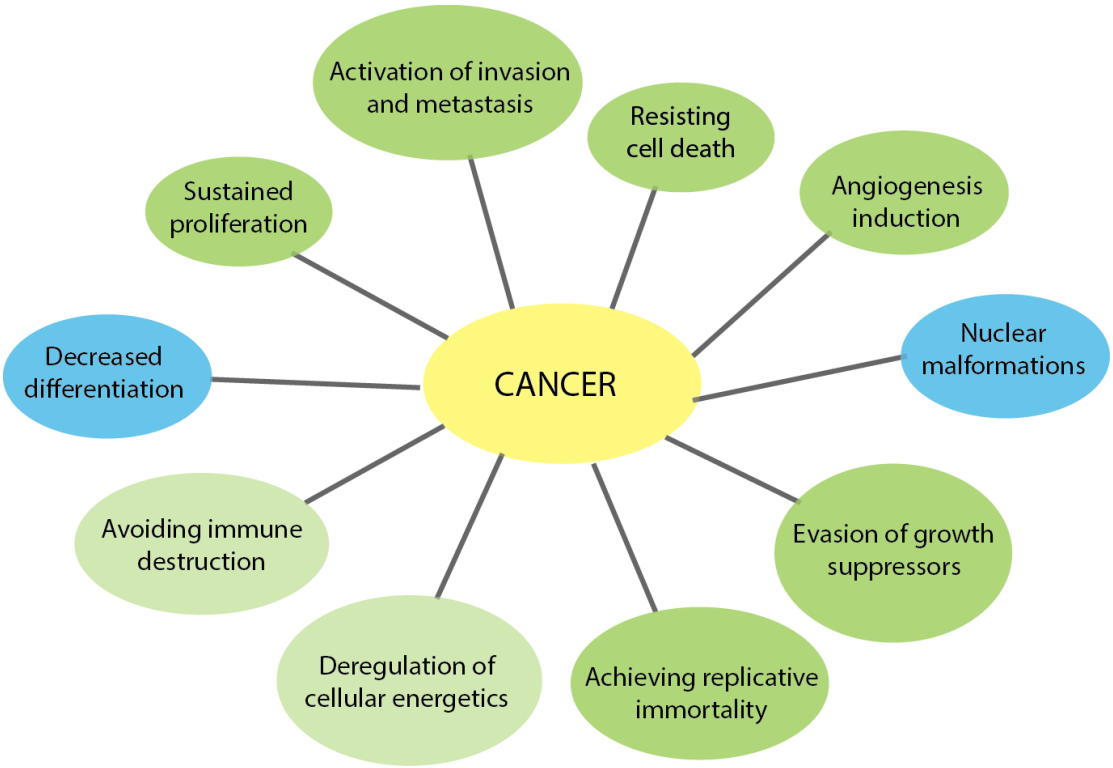


Figure 5.1. Hallmarks of cancer. In darker green are the hallmarks described by Hanahan and Weinberg (2011), while lighter green indicates 'emerging hallmarks' (Hanahan and Weinberg, 2011). Blue colour shows other cancer hallmarks (Zink et al., 2004; Killmann et al., 1963).

As studies from both the genetic and pre-genetic era have shown, planarians and mammals often have a remarkably similar response to carcinogen treatment (Best and Morita, 1982; Foster, 1963, 1969; Schaeffer, 1993) or perturbation of getenetic mechanisms implicated in cancer (Oviedo et al., 2008; González-Estévez et al., 2012). Planarians also have the attractive property of being an *in vivo* experimental system with rapid cell turnover, which results in much faster tumour formation compared to mammals. That being said, two of the cancer hallmarks presented in **Figure 5.1** cannot be studied in planarians. Flatworms do not have a circulatory system, so understanding cancer-related angiogentic response would be impossible. Furthermore, planarians would no be suitable for fully exploring the role of immunity in malignancy, since they only possess components of the innate and not the adaptive immune system (Abnave et al., 2014; Peiris et al., 2014).

Nevertheless, planarian neoblast abundance and functional robustness could prove valuable for understanding the role of 'stemmness' and cellular heterogeneity in cancer. One of the most influencial views on tumour malignancy, the Cancer Stem Cell (CSC) model, puts primary importance on the role of stem cells in disease progression.

5.2.1. MODELS OF CANCER

5.2.1.1. The problem of cancer cell heterogeneity

Cancer cells are undergoing constant mutational and regulational change, as a result of which they are extremely heterogenous. Vogelstein et al. (2013) described four main types of heterogeneity: (1) intratumoural – between cells in the primary

tumour; (2) intermetastatic – between different metastatic lesions, usually due to the fact that each lesion originates from a different clone in the primary tumour; (3) intrametastatic – between cells within one metastatic lesion, due to novel mutations as the lesion develops; (4) interpatient – same tumour type can be very different between patients.

One could easily imagine that the landscape of all of these levels of heterogeneity could be further complicated by epigenetic changes and different fate commitment of the tumour cells. Models trying to rationalise tumour heterogeneity are presented below.

5.2.1.2. Clonal evolution (stochastic) model

The clonal evolution or stochastic model of cancer development suggests that every cell within a tumour is equally likely to have been the cell of origin (Plaks et al., 2015). The model postulates that the cell heterogeneity within a tumour is due to the accumulation of somatic mutations, which then leads to divergence into genetically different subclones (Nowell, 1976). This process is aided by the increased genomic instability of tumour cells compared to normal cells (Nowell, 1976). One of the earliest examples of work supporting this model is the finding that at different stages of colorectal tumour development, genetic changes were found to correspond to different cell phenotypes (Vogelstein et al., 1988). Such findings fortified the idea that tumourigenesis has a lot in common with Darwinian evolutionary concepts (Nowell, 1976). This hypothesis poses that an advantageous mutation conferring growth advantage to a cell gets inherited in the daughters of this cell (Kreso and

Dick, 2014). With time, these daughter cells will form a clonal population and take over the site of origin (Kreso and Dick, 2014). When the next mutation providing a selective advantage hits a cell within that clone, it will give rise to its own subclone (Kreso and Dick, 2014). The co-habitation of different clones and subclones is thought to follow the 'survival of the fittest' law (Kreso and Dick, 2014).

5.2.1.3. Hierarchical/Cancer Stem Cell (CSC) model

The CSC model sees tumours as an organ dependent on stem cells, which have lost their ability to exert homeostatically informed control over their function (Rajendran and Dalerba, 2014) and put more weight on proliferation rather than differentiation (Killmann et al., 1963). This view explains the cell heterogeneity within a tumour as a result of epigenetic regulation associated with different stages of stem cell commitment (Reya et al., 2001). This leads to the formation of a crooked hierarchy-like structure vaguely reminiscent of healthy organs (Rajendran and Dalerba, 2014). Support for this model comes from the histological observations of cancerous tissue showing organisational similarities with the tissue of origin (Rajendran and Dalerba, 2014).

Considering that, in order to initiate a malignant transformation, cells should accumulate a number of subsequent genetic mutations (Fearon and Vogelstein, 1990), it seems that healthy stem cells are the perfect target for cancer initiation because of their long life and self-renewal capacity (Rajendran and Dalerba, 2014). Differentiated cells would be a suboptimal target for cancer initiation because of their shorter lifespan and their inability to propagate (Rajendran and Dalerba, 2014).

Evidence for the CSC model was found in acute myeloid leukaemia studies. Different types of cells FACS-sorted from human leukaemic clones were transplanted in immune-deficient mice (Bonnet and Dick, 1997; Lapidot et al., 1994). It was found that CD34+/CD38- (marker profile characteristic of hematopoietic stem cells) leukaemic cells had a marked advantage in initiating a tumour upon transplantation (Bonnet and Dick, 1997; Lapidot et al., 1994). Furthermore, the resultant tumours contained the cell types encountered in the original tumours before transplantation (Bonnet and Dick, 1997; Lapidot et al., 1994). Similar findings were reported for many solid tumours as well (Chu et al., 2009; Clarke, 2005; Collins et al., 2005a; Galli et al., 2004). In sum, CSC model is adopted as a model in cases where: (1) only a subpopulation of cells within a tumour are capable of initiating a tumour after transplantation in an immune-deficient mouse; (2) the tumours resultant from the xenograft transplantation experiment contain the same cell types as the original tumour; (3) The tumours after transplantation can be serially passaged many times (Rajendran and Dalerba, 2014).

The model also allows for the assumption that, as the disease progresses, mutations leading to differentiated cells acquiring self-renewal properties result in creating additional CSC pools within a tumour (Rajendran and Dalerba, 2014; Weissman, 2005; Chaffer and Weinberg, 2015). Support for this idea was granted in the context of chronic myeloid leukaemia (Jamieson et al., 2004). This is also a potential explanation for the findings that, upon transplantation, tumours originating from either CD133- or CD133+ melanoma cells, were able to recapitulate the heterogeneity of the original tumour (Quintana et al., 2008; Shackleton et al., 2009).

Interestingly, tumour types associated with higher amounts of CSCs were found to have greater likelihood of reoccurring and an overall worse prognosis (Eppert et al., 2011; Liu et al., 2007; Merlos-Suárez et al., 2011). The CSC model proposes that in order to eradicate cancer efficiently, therapy should target the sustaining core of the tumour – the CSCs (Rajendran and Dalerba, 2014). However, concerns have been raised that, given the plasticity of CSCs shown in experiments like Quintana et al (2008) and Shackleton et al. (2009), targeting CSCs specifically might be extremely difficult (Clevers, 2011).

Another main point of concern raised by Clevers (2011) is the interpretation of xenograft assays used to support the CSC model. Cells usually undergo harsh treatments before transplantation, after which they are transplanted in an environment very different from that of the original tumour (Clevers, 2011). Furthermore, there evidence exists that a multipotent cell can be converted into a pluripotent cell upon transplantation (Blanpain et al., 2007). All of these arguments point to the strong possibility that a lot of xenograft experiments investigating CSCs in different tumours could have been misinterpreted (Clevers, 2011). After correcting for the between-species barrier (human-mouse) in transplant experiments, different tumour-initiating efficiency for different tumour cells (evidence for CSCs) was still found for some mouse leukaemias (Yilmaz et al., 2006), breast cancer (Cho et al., 2008), squamous cell carcinomas (Malanchi et al., 2008), but not other cancers (Kelly et al., 2007; Williams et al., 2007). Kelly et al. (2007) found that as many as at least one in ten leukaemia cells transplanted from mouse into a mouse host resulted in *de novo* tumour formation. Such results, go against the hypothesis that only a selected

few CSCs are capable of tumour initiation and essentially renders the concept of CSC meaningless in this experimental context (Clevers, 2011).

Planarian studies have shown that neoblasts possess clonogenic properties and are able to completely repopulate a stem cell-deprived host (Wagner et al., 2011, 2012). Hence, investigating the mechanisms via which neoblasts are able to perform such feats, might be informative for understanding cancer clonogenic behaviour.

5.2.1.4. Combined view

The two models described above are not mutually exclusive. They portray two different sources of tumour cell heterogeneity, which probably act together (Gerlinger et al., 2012; Lagasse, 2008). Clevers (2011) suggested a model explaining how this might be happening (**Figure 5.2**). An important message from this model is that genetic heterogeneity, most likely, influences CSC frequency at different stages of tumourigenesis (Kreso and Dick, 2014).

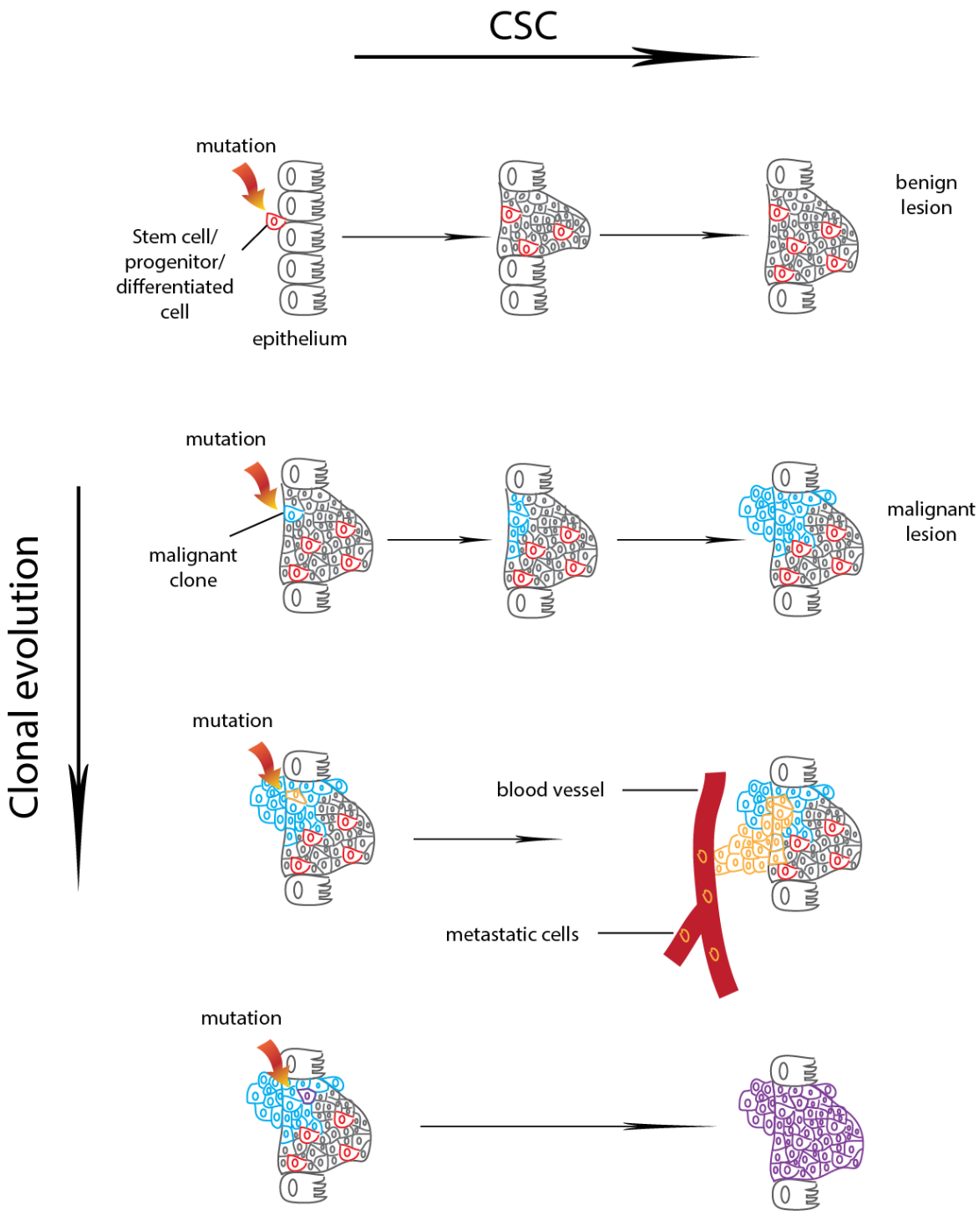


Figure 5.2. A hypothesised symbiosis between the clonal evolution and the CSC model of tumourigenesis. The first genetic mutation happens in a stem cell, progenitor or differentiated cell (red cell) and leads to a benign transformation. The second genetic mutation hits a cell in the benign lesion (blue cell) and this results in a malignant and invasive transformation. The third hit happens within the malignant subclone (orange cell) and leads to further malignant transformations with metastatic cells entering the blood vessels. The final mutation hit (purple cell) leads to the whole tumour being overtaken by cells that behave like cancer stem cell. At each stage of tumour progression there is a certain amount of cells behaving like CSCs (red, blue, orange or purple). The figure is recreated from Clevers (2011).

The combined view of tumour heterogeneity and progression can be further elaborated by adding the influence of the tumour microenvironment (TME) into the equation (Kreso and Dick, 2014). TME is comprised of the non-cancerous cells in the tumour, together with the adjacent to the tumour – healthy stromal cells (Hanahan and Coussens, 2012). A growing body of evidence supports the theory that cancer cells are able to go back and forth along the stem cell-differentiated cell continuum (Quail et al., 2012). Evidence suggests that TME could influence stem-like behaviour in cancer cells (Vermeulen et al., 2010). It was demonstrated that co-transplantation of stromal cells and tumour cells markedly improved the tumour-initiating properties of the transplant in the host (Medyouf et al., 2014).

In conclusion, tumour cell heterogeneity is a complex issue, dependent on genetic variation, epigenetic fluidity of cell fate and the cell interactions with the TME.

5.2.2. CANCER AND EPIGENETICS

Epigenetic events are just as important as genetic changes in cancer (Easwaran et al., 2014). There is growing evidence that at the onset of tumour formation, the boundaries between heterochromatin and euchromatin are perturbed (Baylin and Jones, 2011). The results of such disturbance could lead to altered replication dynamics and an overall changed transcriptional landscape (Easwaran et al., 2014). Most studies investigating epigenetic modifications in the context of cancer are concerned with DNA methylation, with histone modifications receiving significantly less attention.

5.2.2.1. DNA methylation

In cancerous environment, DNA methylation is often found to suffer widespread losses, as well as more spatially concentrated accumulations (Berman et al., 2012; Hon et al., 2012). The losses are associated with chromosome instability, perturbed genomic imprinting and loss of repression of transposable elements (Berdasco and Esteller, 2010). The accumulations happen on normally unmethylated CpG islands in gene promoters, leading to inappropriate gene silencing (Berman et al., 2012). Such Epi-driver mutations (genes with altered expression, but no genetic changes; Vogelstein et al., 2013) in a tumour suppressor gene could have the same effect as a Mut-driver mutation (genes with driver genetic mutations; Vogelstein et al., 2013) in a tumour suppressor since both mutations are heritable upon cell division (Easwaran et al., 2014). One example is Belinsky et al.'s (1998) study showing that the tumour suppressor CDKN2A (Cyclin-dependent kinase inhibitor 2A) promoter got progressively more methylated in response to exposure to a tobacco-related carcinogen. Importantly, ageing is a major cause for many genes' promoter hypermethylation, a finding that ties in well with the increase in cancer risk with age (Ahuja et al., 1998; Toyota et al., 1999).

Irregular DNA methylation is often seen in premalignant cells (Esteller et al., 2000; Suzuki et al., 2004). From the onset of tumour formation, the epigenome seems to lose stability and become chaotic, potentially contributing to the cell population heterogeneity (Hansen et al., 2011). In cancer, the wide spatially distributed losses of DNA methylation appear random and non-clonal, while the targeted gains tend to be clonal (Aryee et al., 2013). In 75% of the studied cases, hypermethylated promoters in cancer belong to genes identified by bivalent promoters in ESCs and adult

stem/progenitor cells (Easwaran et al., 2012). As discussed in detail in **Chapter 1**, bivalent promoters are 'poised' for transcription and associated with maintenance of the stem cell pool (Bernstein et al., 2006). Hypermethylating a bivalent promoter and thus repressing the gene downstream (usually a developmental gene), might contribute to the differentiation-deficient stem cell-like properties of cancer (Easwaran et al., 2012). In support of this hypothesis, it was shown that treatment with DNA demethylating compounds could restore cell differentiation capabilities (Borodovsky et al., 2013).

Interestingly, just like Mut-driver gene mutations, the pattern of CpG island hypermethylation in tumour suppressors differs between cancer types (Costello et al., 2000; Esteller et al., 2001).

Recently, a less well-studied epigenetic modification of cytosines-hydroxymethylation, was also connected to cancer (Jeschke et al., 2016). The enzymes performing hydroxymethylation of cytosines, TET proteins (ten eleven translocation), are associated with tumour suppression (Jeschke et al., 2016). Furthermore, genome-wide levels of hydroxymethylation in most cancers are severely reduced in promoters, gene bodies and intergenic regions (Jeschke et al., 2016).

5.2.2.2. Histone modifications

Some evidence exists that one important way in which CSCs differ from non-CSCs is related to histone modifications. In Acute Myeloid Leukaemia (AML), CSCs showed unchanged DNA methylation pattern, but altered H3K3me3 and H3K27me3

distribution when compared to non-CSCs (Yamazaki et al., 2013). In the case of Glioblastoma multiforme (GBM), loss of H3K27me3 (a repressive mark) on some promoters was associated with aberrant activation of many transcription factors in CSCs (Rheinbay et al., 2013). The authors pointed to one specific example – removed repression of ASCL1 resulted in increased Wnt signaling associated with CSC self-renewal (Rheinbay et al., 2013).

In line with the evidence of the repressive effect of DNA hypermethylated bivalent promoters in cancer (Easwaran et al., 2012), Roesch et al. (2010) reported that a subpopulation of melanoma slow-cycling cells was positive for the enzyme that demethylates the promoter transcriptional activation mark H3K4me3 - JARID1B. Knockdown of JARID1B in melanoma conditions resulted in fast increase of tumour growth followed by tumour exhaustion, indicating that JARID1B and H3K4me3 demethylation are crucial for sustained tumour growth (Roesch et al., 2010). Furthermore, the hypermethylation of tumour suppressor promoters was also associated with deacetylation of histones H3 and H4 and increase in H3K27me3, as well as H3K9 methylation, all indicative of gene repression (Jones and Baylin, 2007).

Tissue-specific expression of histone modifiers was observed in cancer, potentially reflecting the unique patterns of histone modifications in different tumour types (Özdağ et al., 2006). One example is CpG hypermethylation-mediated repression of nuclear receptor binding SET domain protein 1 (NSD1) leading to decrease in H3K36me3 and H3K20me3 (Berdasco et al., 2009). This pattern is characteristic of nervous system tumours, but not other types like leukaemia or breast cancer (Berdasco et al., 2009).

5.2.3. THE MLL FAMILY IN CANCER

An exciting advance in cancer research has managed to connect epigenetic alterations to gene mutations in the vast majority of tumours (Easwaran et al., 2014). This means that, very often, genes coding for chromatin modifying agents or DNA methylation machinery are found to contain Mut-driver mutations in cancer (Shen and Laird, 2013). A prominent example is the family of MLL histone methyltransferases.

5.2.3.1. MLL1

Chromosomal translocations of *Mll1* are among the most frequent driver mutations in acute myeloid and lymphoid leukaemias (Mohan et al., 2010a). It is known that leukaemias involving *Mll1* translocation have high recurrence rate and among the worst prognoses (Mohan et al., 2010a).

As discussed in **Chapter 1**, MLL1 is involved in HOX gene regulation via trimethylating H3K4 on selected gene promoters (Wang et al., 2009a). It is known that perturbed HOX gene expression could lead to tumour formation (Shah and Sukumar, 2010). Mice carrying *Mll1* translocation were shown to have HOX gene overexpression, correlated with overproliferation of haematopoietic progenitor cells and the onset of leukaemia (Shah and Sukumar, 2010).

Mll1 has at least 64 translocation partners (Meyer et al., 2009), potentially leading to variations in pathology (Mohan et al., 2010a). In the translocation process, the resultant chimeric gene contains the N-terminus of *Mll1* and the C-terminus of a

translocation partner, meaning that the resultant protein retains the ability to bind chromatin, but not to methylate H3K4 (Mohan et al., 2010a). The first *Mll1* translocation partner to be characterised was RNA polymerase II elongation factor (*ELL*), leading to the hypothesis that rate of transcriptional elongation might underlie the pathology of this translocation (Shilatifard et al., 1996).

Another common *Mll1* translocation partner is *DOT1* – a H3K79 methyltransferase (Mohan et al., 2010b). It is known that DOT1 is involved in Wnt signaling targets' regulation (Mohan et al., 2010b) and that activation of Wnt is associated with leukaemia initiation (Wang et al., 2010).

5.2.3.2. MLL3

Somatic missense and nonsense mutations in *Mll3* (across its whole open reading frame) are among the most common gene alterations in all types of cancer (Gui et al., 2011; Kandoth et al., 2013; Parsons et al., 2011). *Mll3*'s pan-cancer mutation rate (based on a study of 12 different cancers) was reported to be almost 7%, with high mutation rates in bladder urothelial carcinoma (25%), lung adenocarcinoma (18.4%) and lung scamous cell carcinoma (15.5%) (Kandoth et al., 2013). Some of the landmark studies elucidating MLL3's role in malignancy are outlined below.

As previously discussed in **Chapter 4**, half of all homozygous *Mll3* (SET domain) knockout mice showed proliferation of usually unproliferative terminally differentiated urothelial cells, as well as occasional ureter epithelial tumour formation (Lee et al., 2009). However, a hundred percent of the mice with a

homozygous *Mll3* knockout on a *p53* heterozygous knockout background developed tumours. They also died much quicker than the single *Mll3* knockout homozygotes. These results, alongside with the fact that MLL3 was shown to directly interact with p53 (Lee et al., 2009), led to the conclusion that, at least some, *Mll3* driver mutations in cancer could be working through an effect on p53.

MLL3 was also shown to have a function in acute myeloid leukaemia. AML sufferers often have monosomy 7 or large deletions of 7q (7/del(7q)), associated with severe resistance to conventional treatment (Grimwade et al., 2010). Also, such deletions have only been seen on one allele (Jerez et al., 2012). It was recently reported that 7q houses the haploinsifficient tumour suppressor *Mll3*, whose attenuation perturbed hematopoietic stem and progenitor cells (HSPCs) differentiation and drove leukaemia progression in mice (Chen et al., 2014). Interestingly, the homozygous *Mll3* knockout mice in Lee et al.'s (2009) study did not develop leukaemia, meaning that different dose-dependent perturbations of *Mll3* could result in different pathologies.

Researchers observed that, out of twenty-four 7/del(7q) human AML cases, 10 also showed mutations in the RAS pathway, an incidence much higher than in non-7/del(7q) AML cases (Chen et al., 2014). Similar findings were reported for *p53* mutations in 7/del(7q) AML (Chen et al., 2014). This led Chen et al. (2014) to conclude that MLL3's role in AML might be tied to those of RAS signaling and p53. Confirming this model, they demonstrated that transplanting HSPCs with RNAi for *p53*, *Mll3* and *Nf1* (*neurofibromin 1*; suppressor of RAS) resulted in anemia and

higher peripheral white blood cell count, indicative of leukaemia. Neither of the single RNAi conditions produced the same result (Chen et al., 2014). Using CRISPR/Cas9, Chen et al. (2014) demonstrated that when the same three genes were perturbed, only the cells heterozygous for the *Mll3* deletion led to leukaemogenesis. This finding, together with the fact that RNAi methodology provides incomplete gene suppression, indicate that *Mll3* is a haploinsufficient tumour suppressor (Chen et al., 2014). Sequencing results revealed that *Mll3* suppression in mice led to down-regulation of genes involved in hematopoietic differentiation, just like in AML in humans (Chen et al., 2014). The researchers also managed to tie this effect to decreases in H3K4me3 and increases in H3K27me3 in the downregulated genes (Chen et al., 2014). This connection fits with the expectation that MLL3 COMPASS-like complex is involved to some extent in trimethylation of H3K4 and demethylation of H3K27 via UTX (Herz et al., 2012; Hu et al., 2013; Sedkov et al., 2003).

Focusing on the histone methyltransferase properties of MLL3 in cancer, Weirich et al. (2015) looked at the methylation capabilities of three mutated MLL3 proteins. Three missense mutations in the SET domain region of MLL3 were chosen for investigation – S4757C (found in lung cancer), N4848S (endometrium and CNS cancers) and Y4884C (large intestine cancer) (Weirich et al., 2015). N4848S mutated protein was found to be incapable of histone methylation, Y4884C showed reduced methylation ability, while S4757C's methylation ability was comparable to wild type MLL3 (Weirich et al., 2015). Interestingly, Y4884C exhibited preference to perform H3K4 di- and tri-methylation, while wild type MLL3 was primarily involved in H3K4 monomethylation (Weirich et al., 2015). The fact that these mutations changed the

catalytic activity of MLL3 was interpreted as evidence for them being cancer drivers (Weirich et al., 2015). Results from this study mean that the MLL3 involvement in the carcinogenic landscape might not be due just to decrease of histone methylation activity, but also to shifts in histone methylation patterns.

According to Vogelstein et al. (2013), in order for a gene to be classified as an oncogene, more than 20% of its catalogued cancer mutations should be missense and occurring at the same positions. Tumour suppressors, on the other hand, tend to show mutations throughout their protein-coding sequence, in at least 20% of documented cases, resulting in inactivation (Vogelstein et al., 2013). According to these rules, searching the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Forbes et al., 2015) for mutations in *Mll3* across all recorded 111 cases, resulted in *Mll3* being classified as a tumour suppressor in 44% of cases and as oncogene – in 5% of cases (Vogelstein et al., 2013). Notable mentions of *Mll3* acting as a Mut-driver mutation in cancer include medulloblastoma (Parsons et al., 2011), gastric adenocarcinoma (Zang et al., 2012), pancreatic adenocarcinoma (Mann et al., 2012), pancreatic ductal adenocarcinoma and glioblastoma multiforme (Balakrishnan et al., 2007), breast cancer (Ellis et al., 2012), and diffuse large B-cell lymphoma (Pasqualucci et al., 2011).

5.2.4. CANCER IN PLANARIANS

Considering the fact that planarians are an organism extremely rich in proliferative stem cells, but never develop tumours in the wild, it stands to reason that their

cancer-protective mechanisms must be incredibly robust. Efforts have been made to elucidate those mechanisms.

The most obvious gene to focus on was *p53.* Researchers identified only one planarian gene belonging to the *p53/p63/p73* family and, based on its lack of the SAM domain (absent in p53, but present in p73 and p63 in other organisms), called it *Smed-p53* (Pearson and Sánchez Alvarado, 2010). However, upon BLAST homology search, it is not unequivocally clear that *p53* is the closest match for this planarian gene. *Smed-p53* is expressed mostly in the neoblast early progeny and, to a lesser extent, in stem cells and late stem cell progeny and differentiated cells. The authors observed that *Smed-p53*(RNAi) led to initial stem cell hyperproliferation and decreased stem cell progeny production, followed by stem cell depletion, ventral curling and lysis. Interestingly, when lower doses of *Smed-p53*(RNAi) were used, the hyperproliferative window increased in length sufficiently to allow the development of outgrowths. In the lower dose RNAi experiment, animals did not undergo ventral curling and lysis (features of stem cell depletion), but, due to pharynx malformations, died because of starvation. However, the authors did not show the state of stem cells closer to death in the lower dose RNAi condition. It could be the case, that stem cells were still decreased in number, but just not to the extent necessary to cause ventral curling and death (Pearson and Sánchez Alvarado, 2010). Overall, the observed crash in self-renewal of stem cells appears similar to the defects seen in *p63* knockout mice (Senoo et al., 2007). Furthermore, Suh et al. (2006) demonstrated that p63 in the mammalian female germline performs the DNA damage-sensing function performed by p53 in somatic cells. Knowing that planarian neoblasts share a lot of similarities with germline cells (Guo et al., 2006; Salvetti et al., 2005; Solana et

al., 2009), it seems possible that neoblasts might be under the control of a protein closer to p63, rather than p53. In conclusion, *Smed-p53* could not be labeled a canonical *p53*, since it exhibits functional similarities related to other *p53* family members. Its role in planarian neoblast function is dose-dependent.

Another conspicuous candidate for exploration was the tumour suppressor Retinoblastoma (Rb). Similar to *Smed-p53*, following RNAi, components of the Rb pathway – *Rb*, *EZF4-1* and *Dp*, led to initial overproliferation (not resulting in outgrowth formation), resolved in stem cell maintenance failure at later stages of the phenotype progression (Zhu and Pearson, 2013). Not surprisingly, stem cell depletion was correlated with decrease in early stem cell progeny production and increased cell death. The phylogenetic analyses showed that *Smed-Rb* was more closely related to vertebrate p107/p130 (Rb-like proteins; hypothesised to be the ancestral forms of Rb) than to Rb itself. As such, *Smed-Rb* does not exhibit typical tumour suppressor characteristics (Zhu and Pearson, 2013).

Researchers have also looked into the roles of the tumour suppressor PTEN (Phosphatase and tensin) in the context of planarian biology. Two (at least partially) functionally redundant PTEN orthologues (*Smed-PTEN-1* and *Smed-PTEN-2*) were found in planarians (Oviedo et al., 2008). Double knockdown of these orthologues resulted in decreased motility, head regression (in homeostatic worms), failure to form a blastema (in regenerating worms) and outgrowth formation. The authors reported disorganisation in musculature, cilia and photoreceptors and discontinuities in the basement membrane, as well as stem cell hyperproliferation and defects in differentiation (Oviedo et al., 2008). Like in *Pten* depletion experiments in mice

(Yilmaz et al., 2006), the effect of the double *PTEN* knockdown in planarians was partially reversed by treatment with rapamycin (Oviedo et al., 2008). *Smed-Akt* was also up-regulated in all three FACS-sorted populations in knockdown animals, indicating that the PTEN regulation of PI3K-Akt pathway, often perturbed in human cancer, is conserved in planarians. Interestingly, however, a triple knockdown (*Smed-PTEN-1*, *Smed-PTEN-2* and *Smed-Akt*) did not prevent overproliferation and tumour formation, meaning that planarian PTEN proteins regulate stem cell proliferation in a matter independent from *Smed-Akt* (Oviedo et al., 2008).

The last example of a planarian gene involved in tumourigenesis is *Smed-smg-1* (*suppressor with morphogenetic effect on genitalia*). The gene is a member of the phosphoinositide 3-kinase-related kinase family (PIKKs) with a function in nonsense-mediated decay (Grimson et al., 2004) and was studied in a flatworm context by González-Estévez et al. (2012). This was the first study to discover a tumour suppressor function for this gene, followed by a confirmation of this function in mice (Roberts et al., 2013) and human liver cancer (Han et al., 2014). *Smed-smg-1*(RNAi) led to hyperproliferation, accumulation of undifferentiated cells, as well as early and late stem cell progeny, accompanied with defects in differentiation (González-Estévez et al., 2012). Furthermore, shortly before death, *Smed-smg-1*(RNAi) animals developed invasive outgrowths containing stem cells (among other cells), invading the epidermis and the nerve plexus, epidermal hyperplasia, as well as disturbed organisation of the basement membrane. Double knockown experiments with *Smed-smg-1* and *Smed-tor* or *Smed-smg-1* and *Smed-raptor*, as well as rapamycin treatment of *Smed-smg-1*(RNAi) animals, revealed that *Smed-smg-1*'s molecular function is related to suppressing mTOR signaling (González-Estévez et al., 2012).

All of the cited planarian studies illustrate that flatworms could be a valuable addition to cancer research. Understanding the molecular regulations imposed on adult stem cells in an *in vivo* model organism like planarians could be informative for devising strategies to combat cancerous changes in highly proliferative niches in humans (i.e. gastro-intestinal tissues, hematopoietic system, etc.) (Oviedo et al., 2008). Support for this argument is lent by the fact that, often, there is conservation between molecular mechanisms governing cancerous transformations in humans and planarians, as demonstrated by González-Estévez et al. (2012) and Oviedo et al. (2008).

5.2.5. AIM OF CHAPTER

This chapter aims to characterise the outgrowths forming as a consequence of *Smed-LPT*(RNAi). *Smed-LPT*(RNAi)'s influence on proliferation, nuclear shape, and epidermis is discussed, as well as the cellular composition of the outgrowths. Lastly, defects in stem cell migration following *Smed-LPT* knockdown are presented.

5.3. MATERIALS AND METHODS

5.3.1. REAGENTS

The reagents used for experiments in this chapter are listed in **Appendix A**.

5.3.2. STAINING WITH ANTI-H3P ANTIBODY

Either a stand-alone immunohistochemistry protocol was used, or it was combined with an *in situ* hybridisation staining. Anti-H3P antibody was always used 1:1000 and samples were incubated over the course of 48 hours on a rocker at 4°C. Subsequent PBST washes were performed for 6-7 hours (every 30 mins) at RT, then ON on a rocker at 4°C.

5.3.2. MILD SMED-LPT KNOCKDOWN

For the shielded irradiation experiments reported in this chapter, a milder version of the *Smed-LPT* knockdown was used. This entailed 4 consecutive days of dsRNA (2 μg/μl) injections, followed by 48 hours of rest and then partial irradiation. Animals were amputated behind the eyes 24 hours later. Controls (*GFP* knockdown) were treated in the same way.

5.3.3. ANAESTHETISING AND PARTIALLY IRRADIATING PLANARIANS

The anaesthetising and partial irradiation procedures were inspired by Guedelhoefer and Alvarado (2012). The process was optimised in the Aboobaker laboratory by Ellen Aboukhatwa. Ellen Aboukhatwa, in collaboration with Prof. Mark Hill and Dr. James Thompson, designed a lead shielding device using a 6 mm deep and 1 mm wide lead strip for the purpose of X-ray attenuation. The animals were anaesthetised using 0.2% Chloretone in planarian water, following which they were ordered out of the water in a line on a petri dish. The petri dish was then put on a platform with the

custom-made lead shield so that the shield covers the pharyngeal region of the worms. The irradiation source was placed underneath the animals and delivered 30 Gy X-ray irradiation in the period of 1 minute and 18 seconds. The radiation source used was Gulmay MP1 controller using 225kV X-rays at 17.0mA.

The shielded irradiation procedure was performed by Dr. Nobuyoshi Kosaka and Dr. Prasad Abnave.

5.4. RESULTS AND DISCUSSION

This chapter is going to focus on *Smed-LPT*, since this is the only one of the planarian trithorax-related genes that consistently leads to outgrowth formation following knockdown. The reasons for this result are probably related to functional redundancy between *Smed-trr-1* and *Smed-trr-2,* as discussed in **Chapter 4**.

5.4.1. THE APPEARANCE OF OUTGROWTHS IN SMED-LPT(RNAI) ANIMALS IS PRECEDED AND ACCOMPANIED BY INCREASED STEM CELL PROLIFERATION

Smed-LPT(RNAi) led to overall increased proliferation during regeneration as measured by staining with an antibody against phosphorylated serine 10 on histone H3 (H3P). Twelve hours (6 hours after the first mitotic peak) and 2 days after amputation (second mitotic peak) *Smed-LPT*(RNAi) regenerating middle pieces showed a significantly elevated number of mitotic cells when compared to control (**Figure 5.3A**). Interestingly, there was also a significant increase at day 8 (192 hours) of regeneration. Day 8 of regeneration had been consistently noticed, between

different RNAi injection batches, to either be the last day before outgrowths started appearing or the first day of outgrowth formation. The proliferation curve experiment was performed twice. Representative examples of anti-H3P staining at the timepoints that showed a significant difference are shown (**Figure 5.3B**), with a specific emphasis on the proliferation pattern at day 8 of regeneration. Regions of accumulated proliferating cells were present, possibly marking future sites of outgrowth formation. A typical progression of the outgrowth phenotype seen in *Smed-LPT*(RNAi) animals (starting from day 8 regeneration) can be seen in **Figure 5.3C** with arrows indicating outgrowths. The proliferation dynamics after knockdown of the rest of the trithorax-related genes during both regeneration and homeostasis are shown in **Appendix I**.

These data showed that the appearance of outgrowths in *Smed-LPT*(RNAi) animals was preceded by increased proliferation. In mouse, MLL3 is required for p53 regulation and depletion of both of these genes simultaneously leads to a strong tumourigenic phenotype (Kim et al., 2011). Considering the fact that *Smed-p53*(RNAi) results in stem cell depletion (Pearson and Alvarado, 2010) and *Smed-LPT*(RNAi) never showed signs of neoblast depletion (**Chapter 4**), it is unlikely that *Smed-LPT*(RNAi) defects are related to *Smed-p53* repression in planarians. However, another option is that *Smed-LPT*(RNAi) leads to a mild/incomplete repression of *Smed-p53*, which induces tumour production (Pearson and Alvarado, 2010).

Chapter 4 already demonstrated that planarian trithorax-related genes are involved in lineage specification and, after knockdown, lead to stem cells not being able to recover missing (or maintain existing) structures via differentiation. The favouring of

proliferation over differentiation is a typical hallmark of cancer (Hanahan and Weinberg, 2011; Killmann et al., 1963; Vogelstein et al., 2013).

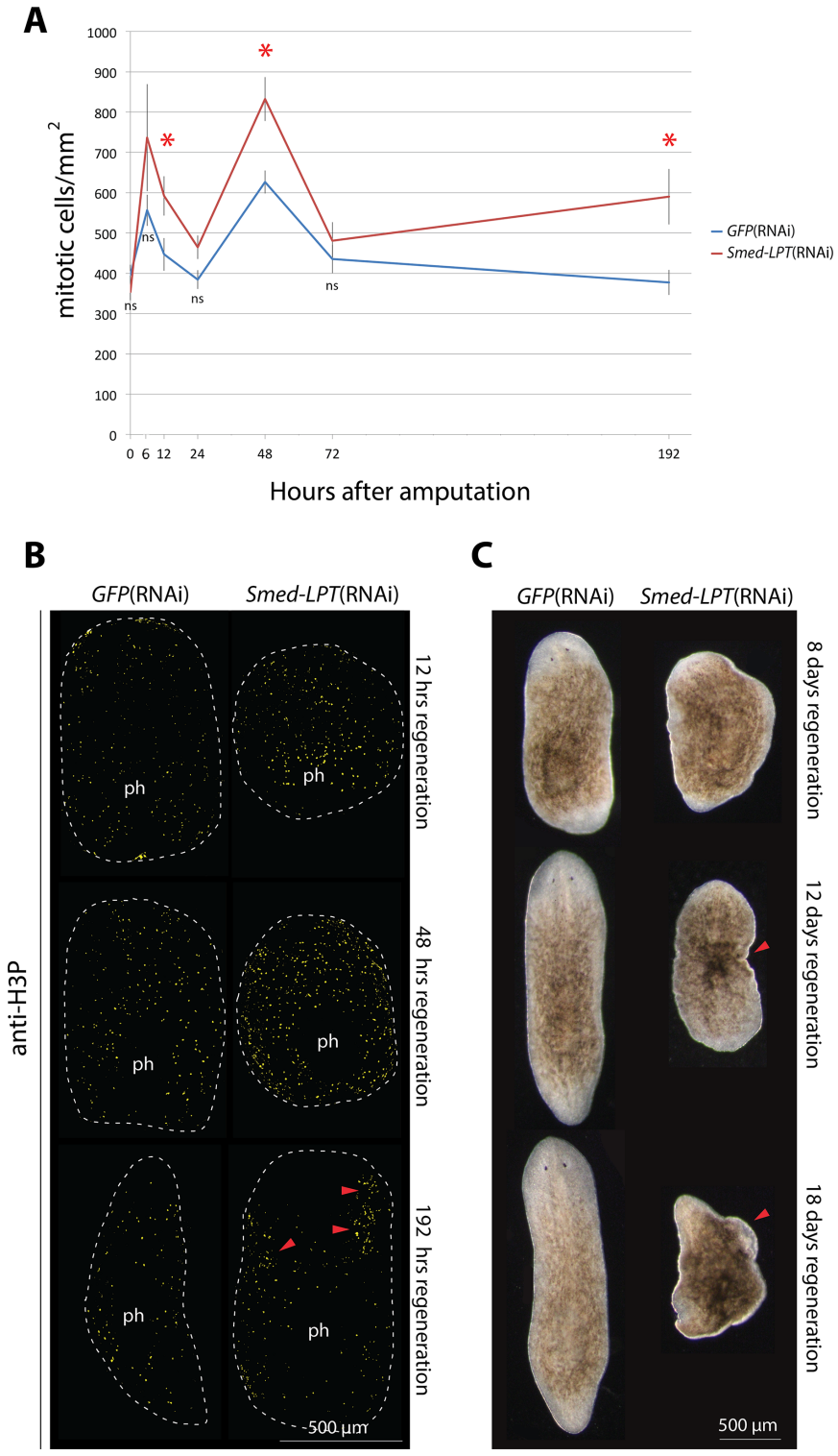


Figure 5.3. Overproliferation defects in *Smed-LPT*(RNAi) animals. (**A**) Proliferation curve measuring mitotic response during regeneration via quantifying amount of anti-H3P-positive cells. Comparisons were performed with a 2-tailed ttest assuming unequal variance where * marks significance $p<0.05$ and 'ns' marks 'not significant'. (**B**) Representative anti-H3P stained samples from timepoints significantly different from controls. Red arrows point to clusters of mitotic cells and 'ph' stands for 'pharynx'. (**C**) A typical progression of *Smed-LPT*(RNAi) phenotype where red arrows show outgrowth formation.

5.4.2. OUTGROWTHS IN SMED-LPT(RNAI) ANIMALS CONTAIN STEM CELLS

The next line of enquiry was whether stem cells are present in outgrowth compartments. Evidence showed that mitotic cells could be found at the boundary of outgrowth and normal tissue in a *Smed-LPT* knockdown animal (**Figure 5.4A**). A double staining experiment (*Smed-H2B*+anti-H3P) revealed that stem cells were entering body regions where they were not usually found, i.e. in the periphery of the normal stem cell pool (**Figure 5.4B**). These findings resemble the *Smed-smg-1*(RNAi) phenotype described in González-Estévez et al. (2012). The experiment in **Figure 5.4B** had 7 outgrowth-containing animals and 7 controls. An important finding was that in 7/7 cases, for both controls and experimental samples, every anti-H3P-positive cell was also *Smed-H2B*-positive. This means that all mitotic cells (as labeled by anti-H3P) are stem cells and the possibility of differentiated cells acquiring proliferative abilities as a consequence of *Smed-LPT*(RNAi) can be excluded.

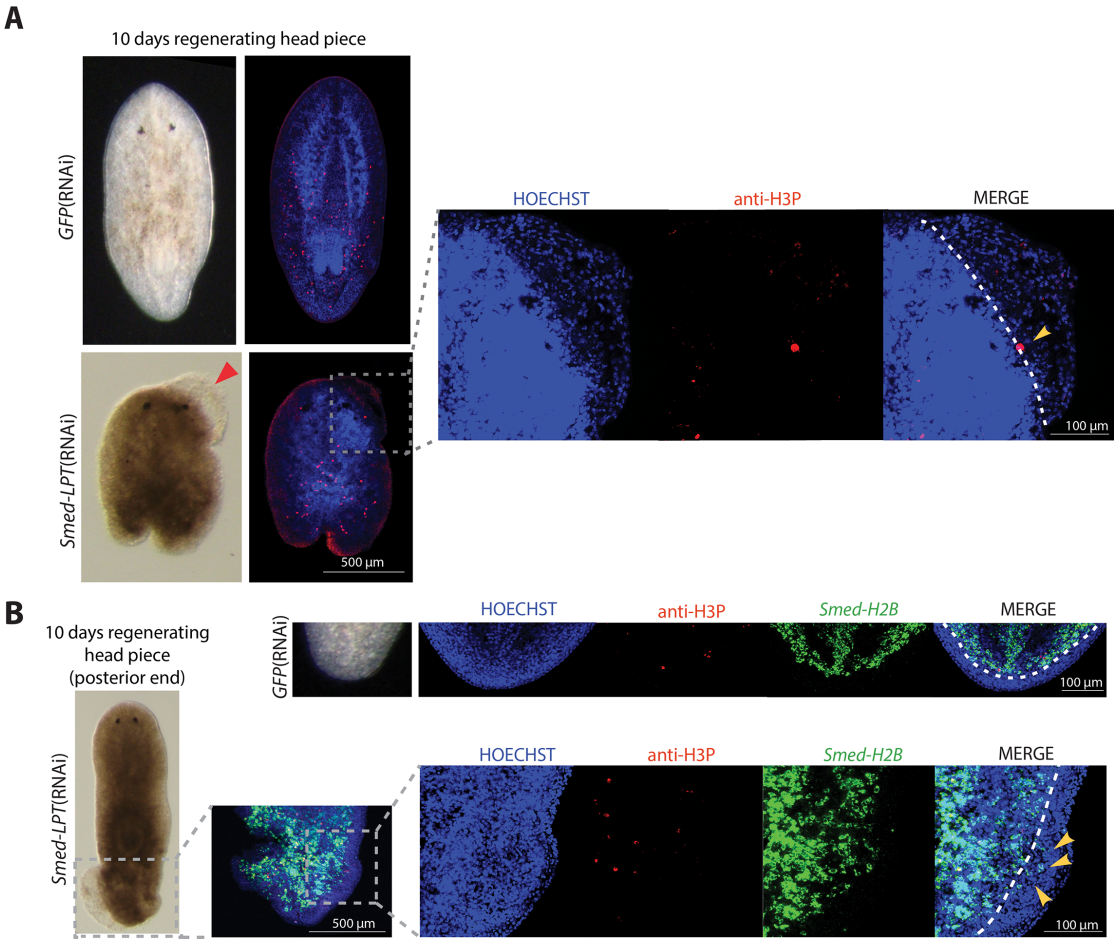


Figure 5.4. Stem cells outside of their usual compartments in *Smed-LPT* knockdown animals. (**A**) Brightfield and anti-H3P staining images of 10-day-regenerating head *Smed-LPT*(RNAi) and *GFP*(RNAi). The red arrow points towards an outgrowth. Zoomed in image shows a mitotic cell entering the outgrowth (yellow arrow). White dashed line shows where the outgrowth starts. (**B**) A double staining experiment with anti-H3P and *Smed-H2B* with accompanying brightfield images. Dashed line indicates where the stem cell compartment would end in a control worm. Yellow arrows point towards cells going outside of that compartment.

Further evidence that mitotic stem cells are present in outgrowth tissue is presented in a *Smed-LPT* knockdown regenerating tail piece forming an outgrowth with some evidence of possible cephalisation (as indicated by Hoechst staining) (**Figure 5.5**). Zooming in on the outgrowth revealed the presence of multiple mitotic cells. It would be really exciting to find out whether the stem cells present in such outgrowths have the properties of cancer stem cells. This could potentially be done via transplanting stem cells present in an outgrowth into a naïve planarian host and observing whether tumourigenesis in the host would follow. It would be equally intriguing to see whether transplantation of differentiated cells from the tumour could lead to tumourigenesis upon transplantation in a new host. If so, this would be indicative of cancer cell plasticity and the fluid fate of cancer stem cells. Since there is a high probability that healthy stem cells in the transplantation host would prevail over the transplanted cancerous cell, the host should ideally be lethally irradiated, as described in Wagner et al. (2011).

Another helpful experiment to unravel the properties of the outgrowths would be to examine whether they contain sigma neoblasts (cNeoblasts).

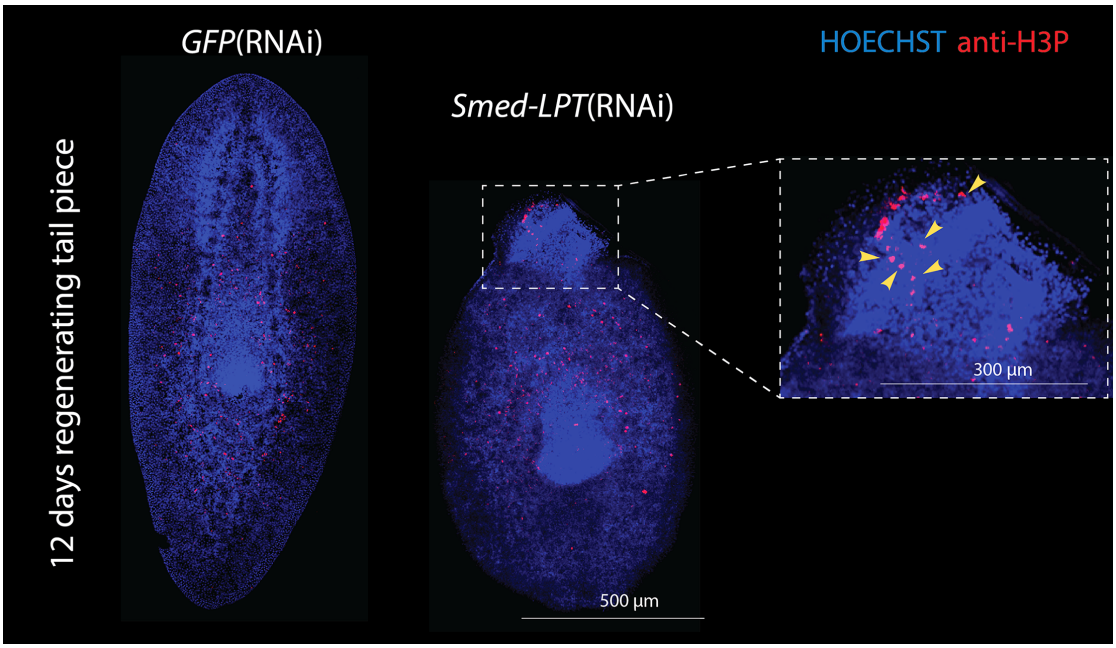


Figure 5.5. Mitotic stem cells present in an outgrowth of a *Smed-LPT*(RNAi) animal. 12-day regenerating tail pieces are stained with anti-H3P. The zoomed in image of the outgrowth formation in the *Smed-LPT* knockdown animals shows presence of mitotic cells, as indicated by the yellow arrows.

5.4.3. SMED-LPT(RNAI) LEADS TO NUCLEAR DEFORMATION

As is the case in mammalian cancer cells (Zink et al., 2004), some nuclei in *Smed-LPT*(RNAi) animals were often misshapen. Nuclei of *Smed-LPT*(RNAi) animals were frequently elongated and unlike the nuclei in their surroundings (**Figure 5.6A**). This could be explained by two possibilities. Either the odd-shaped nuclei belong to cells that are a different type from the surrounding cell type, or they belong to the same cell type as the cells around them but have suffered a morphological change. It was demonstrated that, at least some of these misshapen nuclei belonged to stem cells in an area where stem cells were found (**Figure 5.6B**). Cancer-associated nuclear defects are usually associated with changes in expression of lamins and/or chromosome abnormalities (Capo-chichi et al., 2011; Zink et al., 2004). Alternatively, nuclear elongation was observed as an adaptation facilitating motility inside capillaries during cancer metastasis (Yamauchi et al., 2005). Since interaction with the vasculature is not a concern for planarian tumourigenesis, the nuclear elongation seen in some cells could be a result of the accumulated mass of proliferative cells exerting pressure on other cells against the animal boundary and thus leading to their nuclear deformation.

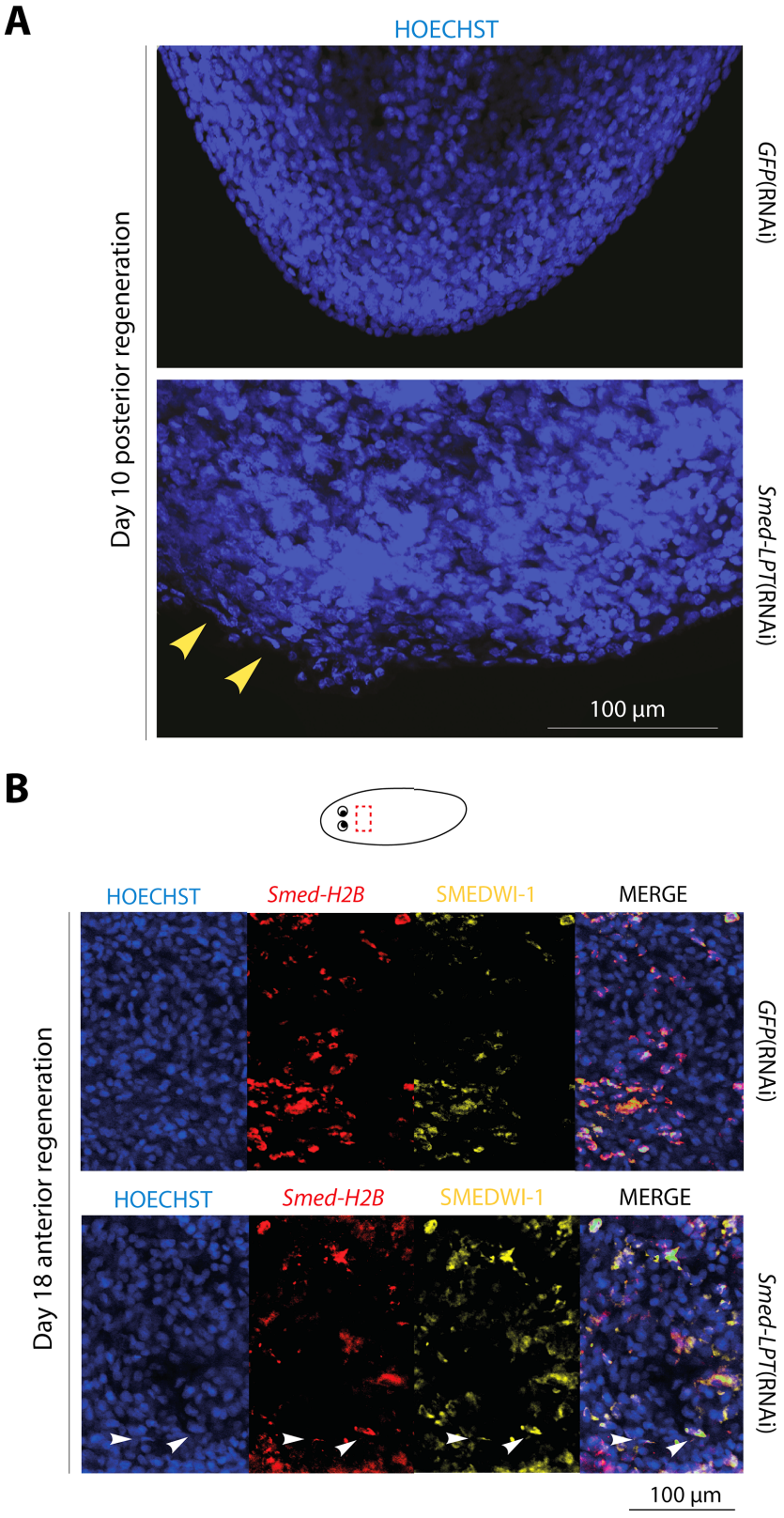


Figure 5.6. Abnormal nuclear morphology in *Smed-LPT*(RNAi) animals. (**A**) The posterior end of a regenerating head piece is compared between 10-day *Smed-LPT*(RNAi) and *GFP*(RNAi) regenerates. Yellow arrows point towards elongated nuclei. (**B**) Pre-pharyngeal regions were imaged. *Smed-H2B* and anti-SMEDWI-1 antibody staining of 10-day regenerating tail pieces. White arrows indicate stem cells with misshapen nuclei.

5.4.4. SMED-LPT(RNAI) ANIMALS EXHIBIT EPIDERMAL ABNORMALITIES

There are clear defects in the epidermis of *Smed-LPT*(RNAi) animals. The epithelial layer in *Smed-LPT* knockdown animals was found to be disorganised and looked as if the cells comprising it had lost their polarity (**Figure 5.7**). Such epidermal features might be related to de-regulation of the zeta population of stem cells. A stage of benign tumours turning into malignant lesions involves epithelial disorganisation and changes of epithelial-mesenchymal interaction (Nelson and Bissell, 2006; Partanen et al., 2009). Furthermore, a significant increase in average epidermal nuclear size (hyperplasia) was observed in *Smed-LPT*(RNAi) worms, another common nuclear malformation in cancer (Zink et al., 2004), similar to that resultant from *Smed-smg-1*(RNAi) (González-Estévez et al., 2012). The epidermal phenotype reported here also bears resemblance to the multi-layered epidermis following *Smed-PTEN-1/Smed-PTEN-2* double knockdown (Oviedo et al., 2008).

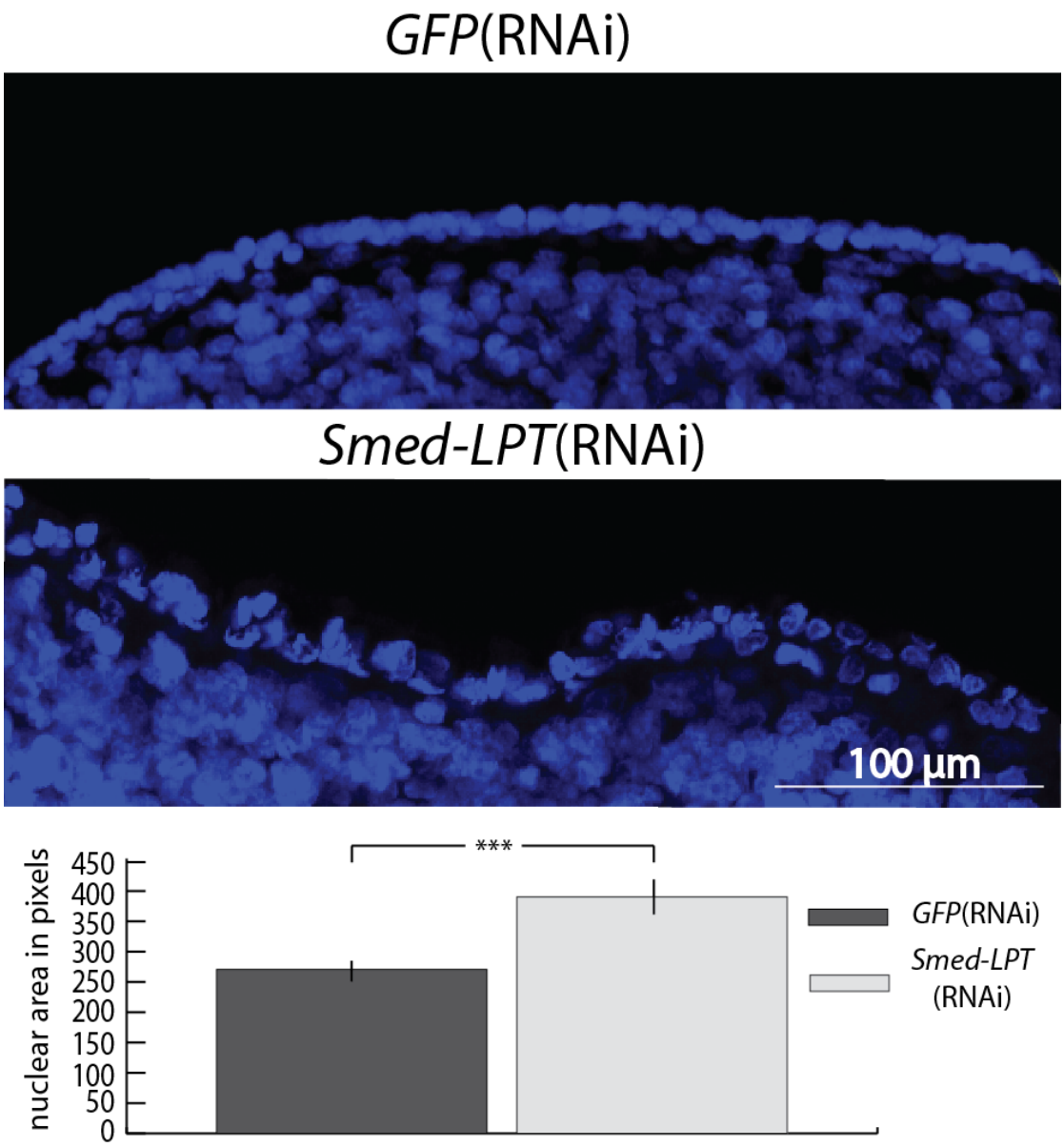


Figure 5.7. Epidermal disorganisation and nuclear hyperplasia in *Smed-LPT*(RNAi) animals. A zoomed in area of *Smed-LPT*(RNAi) animal's epidermis is compared to control (*GFP*(RNAi)). The graph shows a comparison of average epidermal nuclear sizes of *Smed-LPT*(RNAi) and *GFP*(RNAi) animals. The area of 20 nuclei was quantified for each condition. A 2-tailed ttest assuming unequal variance revealed a significant difference in nuclear size ($p<0.001$, marked with ***). Error bars represent SEM.

As reported in **Chapter 4**, the transcription of *Smed-NB.21.11e* and *Smed-AGAT-1* during regeneration was down-regulated following *Smed-LPT*(RNAi). However, this data was collected using day 8 regenerating animals that did not have any outgrowths at the time. Interestingly, the transcription of these genes seemed to be renewed exclusively within the area of the outgrowth once it had appeared (**Figure 5.8**). *Smed-NB.21.11e* and *Smed-AGAT-1* were expressed in a cellular disarray in the outgrowth region. Considering that these two genes had been associated with labeling epidermal progenitors, their transcriptional revival could be related to the epithelial disorganisation events reported earlier. Alternatively, their expression could be in cells unrelated to the epidermis. Since *Smed-AGAT-1* is a homologue of L-arginine:glycine amidinotransferase, known to be involved in the ATP/ADP conversion process (Lee et al., 1998), it is possible that the gene's renewed expression is indicative of tumour-dependent modification of tissue energetics, one of the hallmarks of cancer (Hanahan and Weinberg, 2011). Of course, more experiments would be needed to test this conjecture.

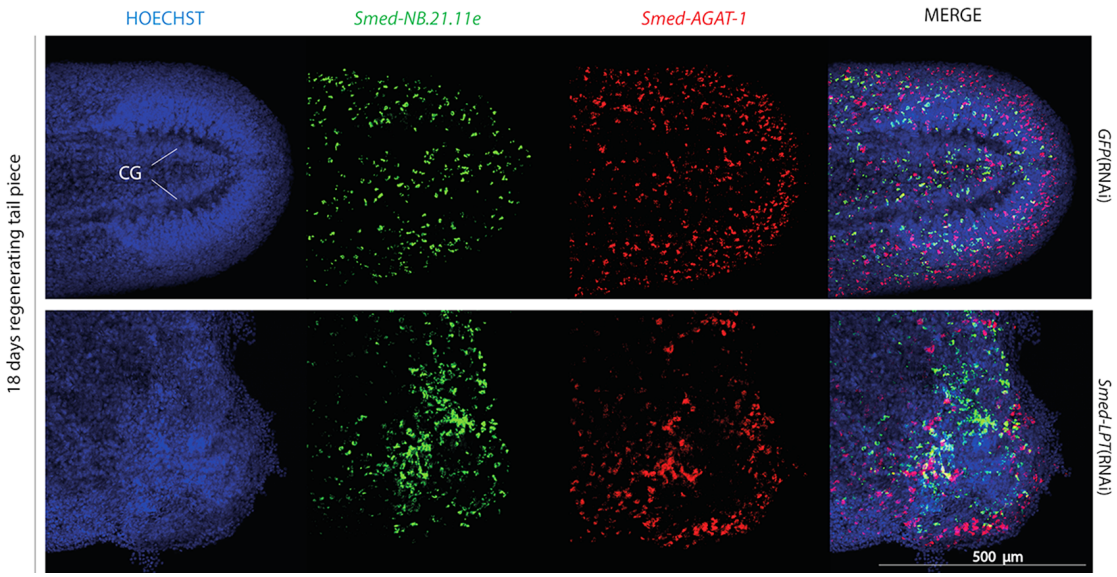


Figure 5.8. Renewed *Smed-NB.21.11e* and *Smed-AGAT-1* expression in outgrowth tissue following *Smed-LPT* knockdown. The anterior end of 18-day regenerating tail pieces are shown in the figure. CG marks the cephalic ganglia.

5.4.5. SMED-LPT CONTROLS STEM CELL MIGRATION

The next question of interest was whether stem cells following *Smed-LPT* knockdown conditions had a change in their migration response. For this purpose, a mild knockdown (Section **5.3.2**), combined with a shielded X-ray irradiation protocol was used (Section **5.3.3**; shielded irradiation was performed by Dr. Kosaka and Dr. Abnave). Twenty-four hours after irradiation the mild knockdown animals were cut underneath the eyes. This was done in order to trigger an active stem cell migration response, as described by Guedelhoefer and Alvarado (2012). Ten *Smed-LPT* and 10 control knockdown animals were left for observation. The phases the animals went through are shown in **Figure 5.9**. Following amputation, the animals formed a regenerative blastema, probably mostly comprised of already existing anterior structures' progenitor cells. The planarians looked healthy until day 7 post-irradiation (6 days post-amputation). The lack of stem cells in the vicinity of the amputation became apparent when animals started undergoing anterior regression. This phenomenon was due to the fact that migrating stem cells from the irradiation-protected area (lead-shielded area) still hadn't reached the amputation site. This regression phenotype was present until 20 days post-irradiation (19 days post-amputation). Surprisingly, at day 18 post-irradiation (17 days post-amputation) 4/10 of *Smed-LPT*(RNAi) animals appeared healthier than controls and had started recovering their missing anterior structures. This result was interpreted as due to faster stem cell migration following *Smed-LPT*(RNAi) when compared to control. Almost a month after irradiation and cutting, all animals had recovered their missing anterior structures (eye malformations were sometimes evident). One out of ten *Smed-LPT*(RNAi) animals had developed a small outgrowth by that time.



Figure 5.9. Brightfield observations following shielded irradiation and amputation of *Smed-LPT*(RNAi) animals. A schematic is showing the area protected by the lead shield and the area of amputation. Representative brighfield images are shown from different stages of recovery following shielded irradiation and amputation (4 days post-irradiation/3 days post-amputation; 7 days post-irradiation/6 days post-amputation; 11 days post-irradiation/10 days post-amputation; 18 days post-irradiation/17 days post-amputation; 29 days post-irradiation/28 days post-amputation). Ten worms per condition were used. White numbers at the top of each image indicate the number of animals per condition showing a particular phenotype. Red arrow points towards an outgrowth formation.

The speculation that *Smed-LPT*(RNAi) animals showed faster recovery due to increased stem cell migration proved to be correct, as demonstrated by the samples stained with a stem cell marker 11 days post-irradiation and 10 days post-amputation (**Figure 5.10**). The experiment was performed on ten animals per condition and the trends illustrated in **Figure 5.10** were consistent among all animals within a condition (10/10). Dashed line shows where the shield-protected area of stem cells was. *Smed-LPT*(RNAi) animals appeared to have a much more abundant stem cell migration response towards the amputation site when compared to controls. This explains why in brightfield observations *Smed-LPT* knockdown animals were, on average, recovering quicker.

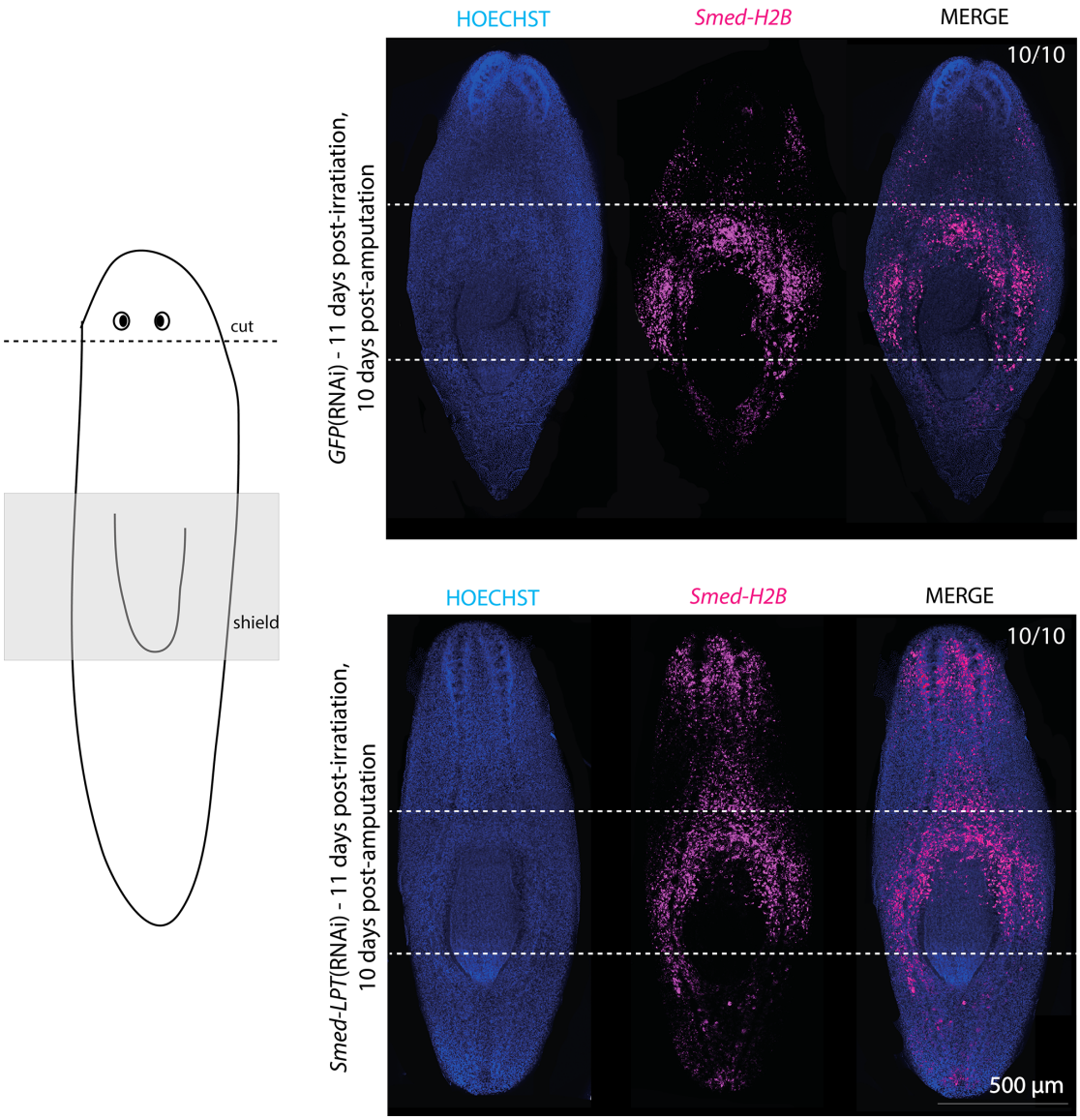


Figure 5.10. *Smed-LPT*(RNAi) leads to increased stem cell migration 10 days post-amputation and 11 days post-irradiation. The schematic shows where the lead shield was positioned and where the amputation was performed. Ten control and 10 *Smed-LPT*(RNAi) animals were stained with *Smed-H2B* 11 days after irradiation and 10 days after amputation. Dashed line shows where the shield-protected area was located in each animal. 10/10 worms within condition showed similar trends.

The stem cell behaviour observed in **Figure 5.10** was also seen later in the recovery process (**Figure 5.11**). At 18 days post-irradiation and 17 days post-amputation. The accumulation of stem cells around the wound site was once again much more obvious following *Smed-LPT*(RNAi) than control knockdown. Interestingly, the amount of passive migratory spread (towards the unwounded tail) was also much more prominent in the *Smed-LPT*(RNAi) animals than in controls.

These findings led to the conclusion that *Smed-LPT* has a regulatory function of active stem cell migration. Its role in passive stem cell migration could be due to increased neoblast proliferation passively pushing away the already existing stem cells.

The finding that *Smed-LPT*(RNAi) enhanced stem cell migration tentatively suggests that MLL3 in mammals could be important for cancer metastasis.

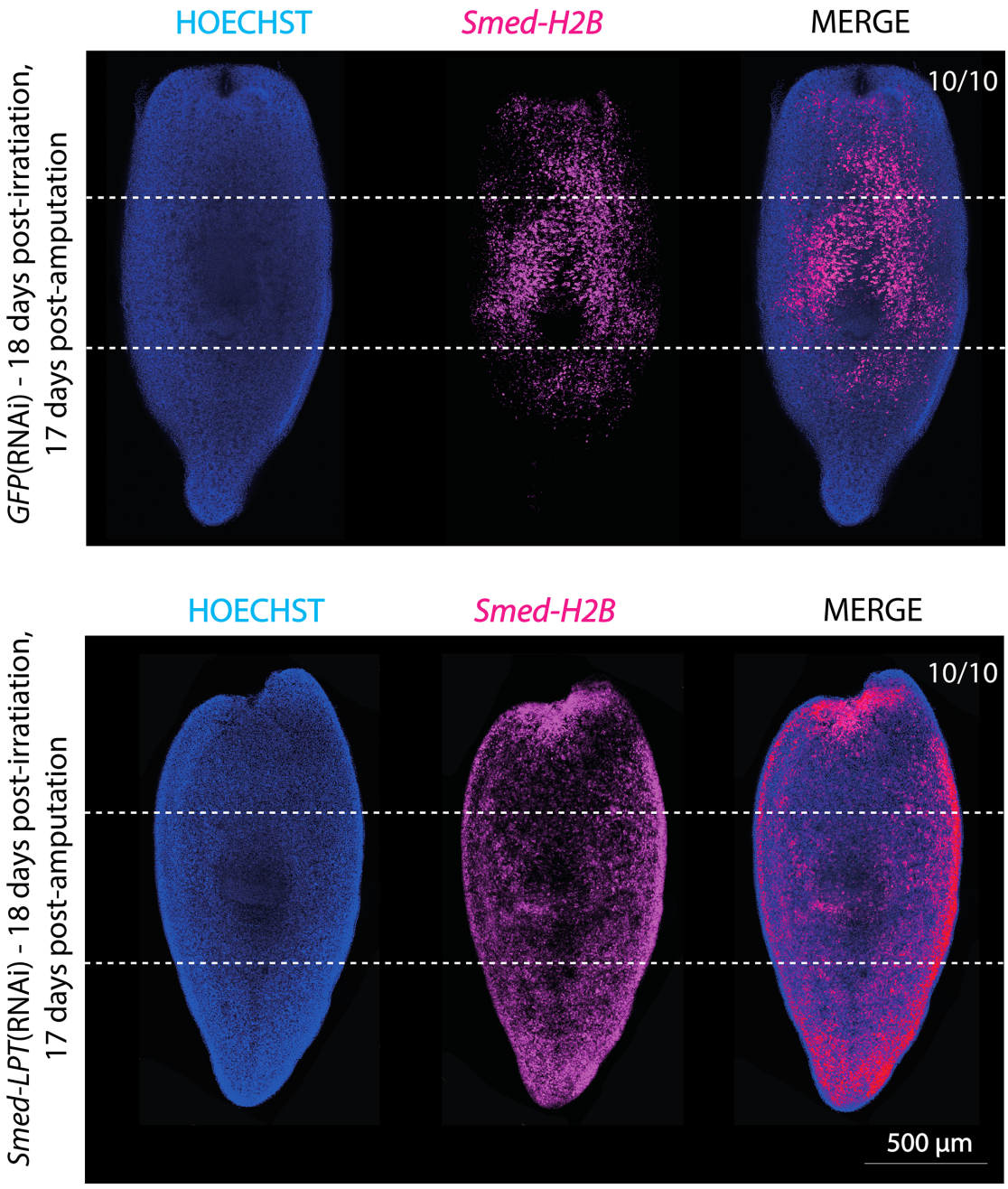


Figure 5.11. *Smed-LPT*(RNAi) leads to increased stem cell migration 17 days post-amputation and 18 days post-irradiation. Ten control and 10 *Smed-LPT*(RNAi) animals were stained with *Smed-H2B* 18 days after irradiation and 17 days after amputation. Dashed line shows where the shield-protected area was located in each animal. 10/10 worms within condition showed similar trends.

The stem cell migration behaviour observed for *Smed-LPT*(RNAi) animals is very similar to unpublished data from our lab on *Smed-smg-1*(RNAi) animals. Experiments performed by Ellen Aboukhatwa demonstrated that knockdown of the tumour suppressor *Smed-smg-1* was manifested in increased stem cell migration towards the wound site following shielded irradiation and amputation. A potential connection could be made between *Smed-LPT* and *Smed-smg-1*. It is possible that trithorax-related genes in planarians regulate *Smed-smg-1*. It was previously reported that *SMG1* is down-regulated via promoter hypermethylation in Acute Myeloid Leukaemia (Du et al., 2014). As mentioned in the introduction to this chapter, many AML cases are characterised by heterozygous deletion of *MLL3*, resulting in perturbation of histone methylation patterns (Chen et al., 2014). In a hypothetical scenario, an animal, like *S. mediterranea*, that does not have detectable DNA methylation (Jaber-Hijazi et al., 2013), could be regulating *Smed-smg-1* expression via SMED-LPT-TRR-mediated histone methylation. ChIP-seq experiments on *Smed-LPT*(RNAi) animals might help answer this question.

5.5. CONCLUSION

Knockdown of *Smed-LPT* leads to increased stem cell proliferation preceding and accompanying the onset of outgrowth formation. Dividing stem cells are present in those outgrowths and transcription of previously suppressed genes (*Smed-NB.21.11E* and *Smed-AGAT-1*) could be reinstated in the tumour. The epidermal tissue after *Smed-LPT*(RNAi) shows signs of disorganisation, associated with malignant transformation. Abnormal stem cell nuclei shape is also observed, a feature common

in cancer. Finally, stem cell migration is enhanced under *Smed-LPT* knockdown conditions, potentially pointing towards a metastatic advantage provided by suppression of *Smed-LPT*. All of these findings point towards a conserved role of *Smed-LPT* (and trithorax-related genes) in tumour suppression. A summary of the findings in context of cancer hallmarks is presented in **Figure 5.12**.

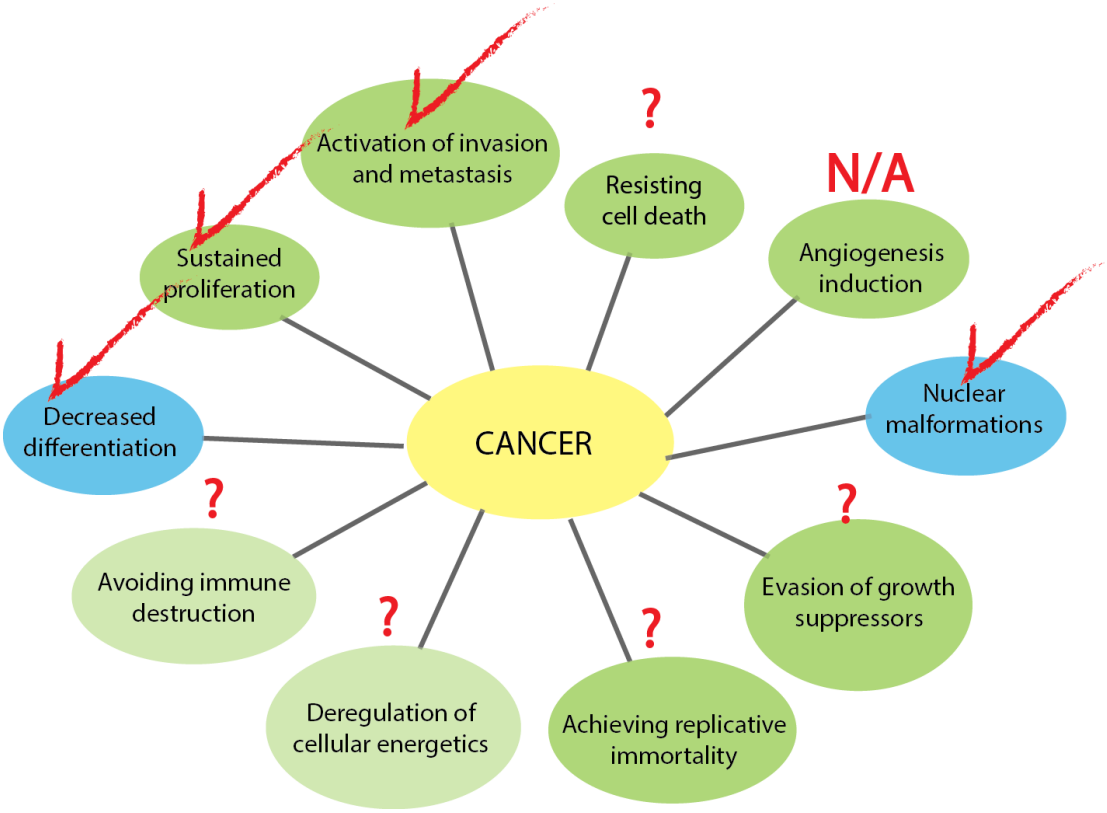


Figure 5.12. *Smed-LPT*(RNAi) phenotype in the context of the hallmarks of cancer. Cancer hallmarks, as defined by Hanahan and Weinberg (2011) are shown in darker green; lighter green shows 'emerging hallmarks' and 'unofficial' cancer hallmarks are shown in blue. A red 'tick' sign indicates that there is evidence for this feature in *Smed-LPT*(RNAi) animals. 'N/A' stands for 'not available'. The unexplored features in the context of *Smed-LPT*(RNAi) are marked with a question mark.

1. CELLULAR DYNAMICS AND HISTONE MODIFICATION CHANGES FOLLOWING KNOCKDOWN OF TRITHORAX-RELATED GENES

6.1. ABSTRACT

Methylation of histone lysine residues is one of the most credited mechanisms for passing epigenetic information through the cell cycle, as well as independently from it. Studies in the past few years have demonstrated the importance of LPT-Trr/MLL3/4 in conveying such information via methylating to varying degrees lysine residue 4 on histone 3 (H3K4). LPT-Trr/MLL3/4 is involved in transcriptional activation via selective hormone-dependent trimethylation of H3K4, as well as enhncer regulation and transcriptional repression via monomethylation of H3K4. Experiments demonstrated that in planarians *Smed-LPT, Smed-trr-1* and *Smed-trr-2* regulates H3K4me1, H3K4me3 and H3K27me3. Intriguingly, there appeared to be a different histone modification response to regenerative and homeostatic conditions under knockdown of trithorax-related genes. *Smed-trr-1* and *Smed-trr2* were also found to compensate for each other's H3K4 monomethylation capabilities after knockdown. The results presented in this chapter lay a solid foundation for future in-depth studies of changes in transcriptional regulation via methods such as ChIP sequencing.

6.2. INTRODUCTION

6.2.1. HISTONE MODIFICATION INHERITANCE DURING REPLICATION

The cell cycle poses a challenge for epigenetically inherited changes. The most trying event is the survival at the replication fork where chromatin organisation is perturbed and subsequently recovered. Faithful recovery of established chromatin patterns is important for maintenance of cell fate (Probst et al., 2009). However, the replication events also present a perfect opportunity for altering chromatin features in order to allow differentiation and development as necessary (Probst et al., 2009).

6.2.1.1. Models of histone modification inheritance

The DNA polymerases involved in the synthesis of the leading and the lagging strand during DNA replication are aided in their function by the proliferating cell nuclear antigen (PCNA) protein (Moldovan et al., 2007). It was suggested that, at least in yeast, PCNA might be the link between genetic and epigenetic inheritance during DNA replication (Zhang et al., 2000). PCNA interacts with a plethora of chromatin modifying complexes (Corpet and Almouzni, 2009; Zhang et al., 2000) and recruits to the replication fork proteins like HDACs (Milutinovic et al., 2002), SET8 (Huen et al., 2008) and chromatin assembly factor 1 (CAF1) (Shibahara and Stillman, 1999). Another suspected mediator between genetic and epigenetic inheritance at the replication fork is minichromosome maintenance complex (MCM). MCM is a replicative helicase reported to interact with anti-silencing function 1 protein (ASF1),

known to be a crucial chaperone for histones incorporated into parent and daughter DNA strands (Groth et al., 2007).

Histone variant H3.1 and H4 incorporation in chromatin is DNA-synthesis dependent, while the incorporation of H3.3 and CenH3 (centrosome specific H3), H2B and H2A are not dependent on the cell cycle (Kimura and Cook, 2001). For these reasons, histones H3 and H4 and their respective modifications are the perfect candidates for transmitting information from one DNA replication event to the next (Probst et al., 2009). Histones H3 and H4 were found to be deposited as dimers on the DNA (English et al., 2005; Tagami et al., 2004).

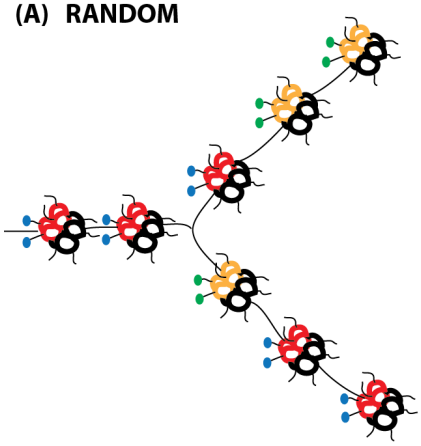
There are three potential models (Probst et al., 2009) to explain histone modification inheritance at the replication fork (**Figure 6.1**). The first model assumes that the distribution of old (parental) and new H3 and H4 histones happens randomly (**Figure 6.1A**). In this scenario, modifications are re-established on the new histones using old histones in neighbouring nucleosomes as a template. This happens when a histone modification Reader protein recognises a parental modification, following which a Writer protein (bound to the Reader) performs the modification on a new histone in the vicinity. This model was suggested to explain the maintenance of heterochromatin protein 1 (HP1) via H3K9 methylation propagation at pericentric heterochromatin (Bannister et al., 2001; Lachner et al., 2001). Furthermore, it was speculated that this mechanism could underlie the maintenance of the repressive H3K27me3 mark deposited by Polycomb repressive complex 2 (PRC2) and, during replication, also recognised by PRC2, independently from PRC1 (Hansen et al., 2008).

Probst et al. (2009) hypothesised that this model might be applicable for repetitive regions where many nucleosomes carry the same mark.

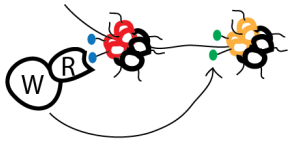
The second model proposes a semi-conservative distribution of new and parental histones (**Figure 6.1B**). Each nucleosome on both the leading and the lagging strand contains an old and a new H3-H4 dimer. This 'intraparticle' maintenance of histone modifications entails a Reader-Writer protein complex recognising a modification on a parental dimer and modifying the new histone dimer. This suggested mechanism bears resemblance to the way hemimethylated DNA on one strand serves as a template for DNA methylation on the opposite strand (Probst et al., 2009).

The third model poses that old and new histone dimers are deposited asymmetrically between the lagging and the leading strand (**Figure 6.1C**). In order for histone modification maintenance to happen, crosstalk between the two strands is necessary. Not much evidence is available for this model, but it might provide a window of opportunity for changing the cell fate on one of the two DNA strands (Probst et al., 2009).

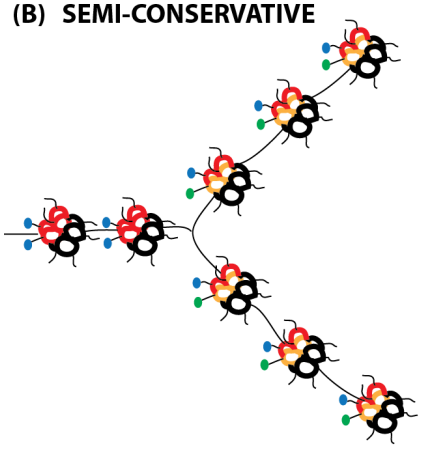
(A) RANDOM



'Interparticle' - maintenance via using neighbouring mark as template



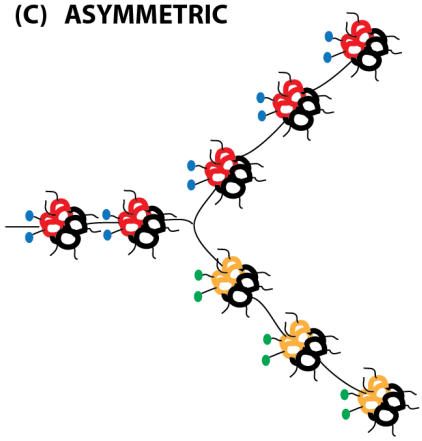
(B) SEMI-CONSERVATIVE



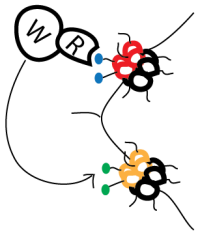
'Intraparticle' - maintenance via using hemimodified old nucleosome as a template



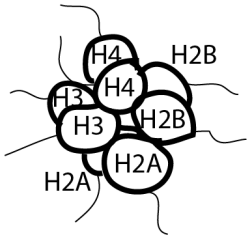
(C) ASYMMETRIC

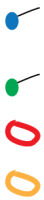


'Interstrand' - maintenance via inter-strand crosstalk



Histone octamer





Parental mark

New mark

Old H3-H4 dimer

New H3-H4 dimer



Reader

Writer

Figure 6. 1. Models of parental and newly synthesised histone dimers deposition at the replication fork. (**A**) Random histone dimer distribution – parental (old) and new H3-H4 histone dimers are distributed randomly between the leading and the lagging strand. A histone modification Reader/Writer complex binds to a parental modification and deposits a new modification of the same type on the neighbouring nucleosome containing a new H3-H4 dimer ('interparticle' inheritance). (**B**) Semi-conservative histone dimer distribution – old dimers are deposited evenly in every nucleosome on each of the two DNA strands, complemented by a new dimer. The Reader/Writer complex uses the modification on the old dimer as a template for modifying the new dimer ('intraparticle' inheritance). (**C**) Asymmetric histone dimer distribution – old dimers are distributed on one DNA strand, while new ones are on the other strand. The Reader/Writer complex needs to perform 'interstrand' recognition and modification functions. The figure is recreated from Probst et al. (2009).

It is also possible that histone modification inheritance is tightly connected to DNA methylation at regions rich in methylated cytosines. One example of such relation is the fact that methyl CpG-binding protein 1 (MBD1), in a complex with SETDB1 (Sarraf and Stancheva, 2004), interacts with histone chaperone CAF1 while DNA is being replicated (Reese et al., 2003). Furthermore, DNA methyltransferase 1 (DNMT1) interacts with the lysine methyltransferase G9a (Estève et al., 2006).

6.2.1.2. Histone variants inherited independently of the cell cycle

As mentioned above, histone variants like H3.3 are deposited in a replication independent pattern and thus carry replication-unrelated epigenetic memory (Henikoff et al., 2004; Loyola and Almouzni, 2007).

H3.3 is associated with active transcription and is usually found in areas with an abundance of activating histone modifications (Ahmad and Henikoff, 2002; Loyola et al., 2006). Often, nucleosomes containing H3.3 are much more unstable than those

with H3.1 (Jin and Felsenfeld, 2007), suggesting that H3.3 nucleosomes are more dynamic and mobile during transcription (Probst et al., 2009). H3.1 and H3.3 have not been observed in the same nucleosomes, however it was shown that when in close proximity to H3.3 nucleosomes, H3.1 nucleosomes tend to acquire activation modifications (Loyola et al., 2006). The dilution effect of H3.3 variants, resulting from them not participating in replication, is offset by the fact that chaperones like Hir-related protein A (HIRA) help H3.3 get incorporated in nucleosomes in a transcription-dependent manner (Tagami et al., 2004).

6.2.1.3. Histone lysine acetylation

Levels of histone lysine acetylation during replication were shown to be re-established to original levels as soon as 2 hours after deposition at the replication fork (Annunziato and Seale, 1983; Scharf et al., 2009; Sobel et al., 1995). Acetylation levels were rapidly adjusted between old and new histones (Scharf et al., 2009) due to the high turnover of acetyl groups (Chestier and Yaniv, 1979). The fast turnover rate, combined with the observation that histone acetylation events happened independently from one another, led researchers to speculate that acetylation is not a suitable mechanism to preserve epigenetic memory across cell divisions (Scharf et al., 2009).

6.2.1.4. Histone lysine methylation

Histone acetylation or phosphorylation have a much shorter half-life compared to methylation of histone lysines (Barth and Imhof, 2010; Pesavento et al., 2008). As

such, lysine methylation is the best candidate for having impact on epigenetic inheritance during the cell cycle (Goss, 2014). Newly made H3.1 histones incorporated in nucleosomes at the replication fork did not seem to acquire the same level of lysine methylation as old H3 histones, even after one whole cell cycle (Xu et al., 2012). The exception to this rule was H3K9me2 which was reinstated with no difference between old and new histones within one cell cycle (Xu et al., 2012). The authors speculated that this might be due to the fact that the lysine methyltransferase that catalyses H3K9me2, G9a, is localised at the very base of the replication fork (Estève et al., 2006).

It was also discovered that the lysine methylation dilution effect, due to uneven methylation of new histones, was amended by further methylation of old histones (Xu et al., 2012). In this way, the overall level of the specific modification within a nucleosome did not change significantly. The authors proposed that this was, in essense, a quick fix for the slow lysine methylation maturation of newly synthesised histones (Xu et al., 2012). These results were confirmed for the cell cycle inheritance of H3K9me3 and H3K27me3 by a more recent study (Alabert et al., 2015). It was found that the deposition of trimethylation marks on new histones continued for up to three subsequent cell divisions. The researchers also demonstrated that H3K27me1, H3K36me1 and H3K27me2 were among the first modifications to be re-established on the new histones in a nucleosome within one cell cycle (Alabert et al., 2015).

A question that arises is whether old histones are devoid of modifications upon deposition in the newly replicated DNA or whether they carry their modifications

with them. In the first scenario, in order to match the findings reported by Xu et al. (2012), old histones would have to have an affinity for being modified faster than newly synthesised histones. Alternatively, if the second scenario were true, the difference in histone methylation levels would simply be due to the delayed methylation of the new histones. Evidence from *Drosophila* studies pointed towards erasure of histone marks on old histones upon formation of new nucleosomes (Petruk et al., 2012, 2013). Alabert et al. (2015), however, reported that in human somatic cells there were no signs of modification erasure on old histones upon deposition on newly synthesised DNA.

Since Xu et al. (2012) did not investigate inheritance of histone lysine methylation marks associated with transcriptional activation (ex: H3K4me3), their conclusions were centered on mechanisms of epigenetic repression. The researchers proposed a 'buffer' model for inheritance of silencing methylation marks (Xu et al., 2012). The model builds on the fact that silenced regions are usually broadly spread across a few kilobases of DNA, rather than affecting a selected few nucleosomes (Lee et al., 2006b; Schwartz et al., 2006). Therefore, a certain threshold of methylation is what is important, rather than uniformly matched methylation between old and new histones (Xu et al., 2012). Hence, some degree of fluctuations in lysine methylation could be tolerated. This speculation was attested by studies failing to see a derepression of repressed targets during the initial dip of lysine methylation at the beginning of S phase (Bar-Joseph et al., 2008; Tachibana et al., 2002).

If this 'buffer' model is correct, then the semi-conservative and asymmetric models of nucleosome assembly at the replication fork (**Figure 6.1**) would not be appropriate

for maintaining the repression of a silenced region after the second round of replication. More specifically, the first round of replication might not have detrimental effects on histone lysine methylated regions. However, due to the delayed maturation of lysine methylation events on new histones, severe derepression of silenced genes would occur during the second round of replication if the semi-conservative or the asymmetric models were followed.

Attempts at studying H3K4 methylation inheritance at the cell cycle were largely unsuccessful due to the weak signal produced by these peptides in the pSILAC + mass spectrometry assays used (Scharf et al., 2009).

6.2.1.5. Histone modifications and different stages of the cell cycle

Goss (2014) looked at the histone modification patterns on genes characteristically expressed during different stages of the cell cycle in ES cells. Not surprisingly, she found that there was an enrichment of active mark(s) at the stage of the cell cycle pre-empting the gene's expression or at the stage coinciding with the gene's expression (Goss, 2014). The dynamic shifts in histone modification patterns throughout the cell cycle in ES cells were unlike the heritable nature of those modifications during the cell cycle of differentiated cells. The marks H3K4me3 and H3K27me3 were particularly good example of this heritability in differentiated cells. It was also demonstrated that the most universal shift in the tested histone modification marks between ES and differentiated cells was that the enrichment for H4K8ac at G1/S phase in ES cells was replaced by enrichment of H3K9ac at G1/S. This

phenomenon was observed for a plethora of pluripotency and house-keeping genes (Goss, 2014). However, the significance of this event is unknown.

The emerging pattern from these studies is that heritability (in the classic sense of the word) throughout the different stages of the cell cycle is not present in ES cells (Goss, 2014). Heritable patterns from one cell cycle stage to the next are much more obvious in differentiated cells (Goss, 2014). These observations could be related to the fact that chromatin proteins bind loosely to ES cells's chromatin and are in a state of hyperdynamic activity (Meshorer et al., 2006). During differentiation, chromatin-binding proteins bind tighter to chromatin (Meshorer et al., 2006) and potentially bring about more lasting changes related to lineage commitment.

6.2.2. CHROMATIN STATES

Chromatin organisation is extremely diverse and mobile. Different histone modification and non-histone protein binding patterns make some regions behave as a gene, while others as gene regulatory elements. The most widely investigated regulatory elements are promoters and enhancers. Therefore, it is of primary importance to understand their chromatin signature. It was described that in the human genome the H3K4me2/3-rich pattern was characteristic of active promoters, while a H3K4me1-rich pattern usually defined enhancers (Heintzman et al., 2007). The researchers also reported that promoter-distal (more than 2.5 kb from the 5' end of a gene) sites of p300 (histone acetyltransferase) binding were also found in enhancers (Heintzman et al., 2007). The H3K4me2/3-H3K4me1 promoter-enhancer dichotomy is not, however, as clear-cut. Pekowska et al. (2011) found that the

presence of all three H3K4 methylation states on an enhancer was reflective of activation of downstream genes. The problem was studied in the context of T-cell development in mice. It was observed that gain of H3K4me2/3 at enhancers was associated with activation of genes involved in T-cell development, while depletion of H3K4me2/3 in those enhancer regions led to repression of those same genes. The accumulation of H3K4me2/3 on enhancers was found to be related to co-occurrence of RNA-polymerase II's (Pol II) (Pekowska et al., 2011). However, Pekowska et al. (2011) recognised that the presence of H3K4me2/3 on enhancers might not be a universal requirement for enhancer activation, pointing out that H3K27ac might be an alternative/overlapping mark of enhancer activation (Creyghton et al., 2010).

An incredibly thorough effort was undertaken by Kharchenko et al. (2011) to describe different chromatin states associated with various functions in *Drosophila*. The main investigative tools were 18 antibodies against histone modifications. A machine-learning technique was used to detect combinatorial patterns for these 18 modifications in S2 and BG3 cells (Kharchenko et al., 2011). This endeavor resulted in the characterisation of 9 chromatin states summarised in **Figure 6.2**.

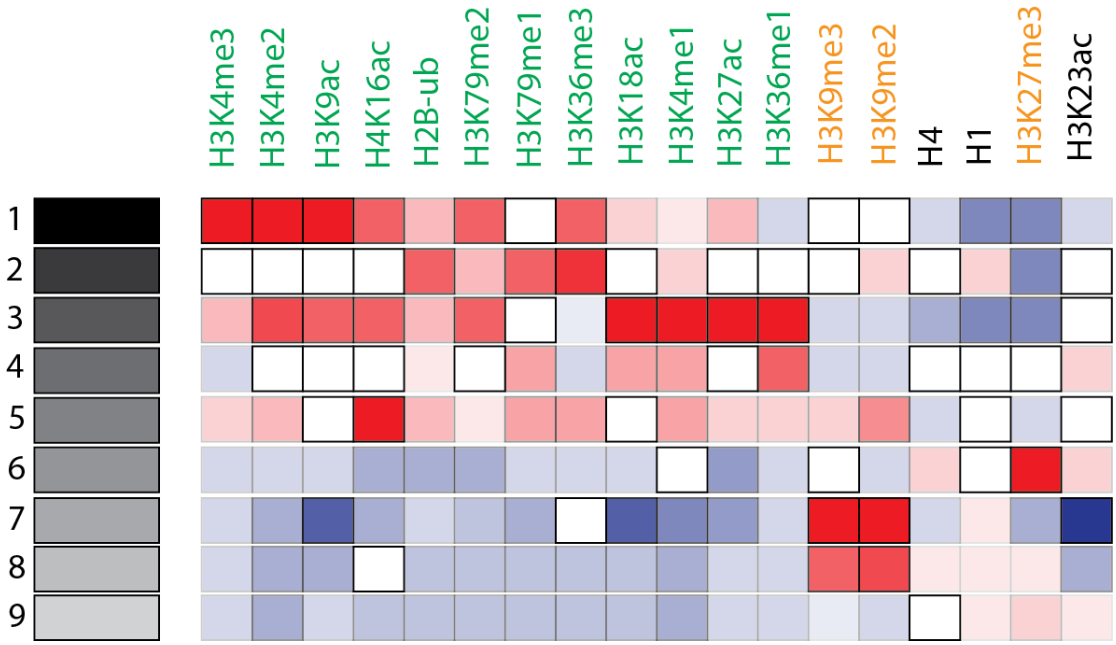


Figure 6.2. Kharchenko's 9 chromatin states. Individual states are marked with different nuances of gray. The presence of 18 histone modifications (active marks are in green; repressive marks are in orange) in each chromatin state is marked via shades of red (enrichment) or blue (depletion). The darker the colour, the more enriched/depleted a certain mark is. The figure is recreated from Kharchenko et al. (2011).

State 1 was found to be characteristic of active promoters and transcriptional start site-related regions enriched for almost all active histone modifications, especially H3K4me2/3 and H3K9ac. Around 8% of the *Drosophila* genome showed this pattern. State 2 was seen in 13% of the genome, usually in exon regions and it characterised transcriptional elongation patterns defined mostly by enrichment in H3K36me3. State 3 was associated with enhancers and enrichment for H3K18ac, H3K4me1, H3K27ac and H3K36me1, seen in 8% of the genome. State 4 showed enrichment of H3K36me1, but not H3K27ac and was seen in 11% of the genome, mostly intronic sequences/enhancers. State 5 demonstrated high levels of H4K16ac and was associated with dosage compensation in male cells and comprised 6% of the genome. State 6 was representative of Polycomb-mediated transcriptional repression (6% of the genome). State 7 was associated with enrichment for H3K9me2/3 in pericentromeric heterochromatin, comprising around 5% of the genome. Another type of heterochromatic region (state 8) was designated by slightly lower enrichment of H3K9me2/3 and comprised 2% of the genome. Finally, state 9 represented 'background' silent domains seen in 40% of the genome (Kharchenko et al., 2011).

As expected, with changes in organismal demands, there was a certain level of transition from one state to the other (Kharchenko et al., 2011).

Intriguing findings regarding enhancer heterogeneity were reported by Heintzman et al. (2009). The group looked at five human cell lines – HeLa, GM06690, K562, embryonic stem cells and BMP4-induced embryonic stem cells and discovered that, while modifications at gene promoters and insulators did not tend to vary between

cell lines, the same was not true for enhancer signatures (Heintzman et al., 2009). It was reported that the distribution of p300 binding sites and histone modifications patterns (most notably H3K4me1) across enhancers was highly cell type-specific (Heintzman et al., 2009; Koch et al., 2007). These findings suggest that enhancers are probably more important in establishing cell type-specific transcriptional patterns than previously thought (Heintzman et al., 2009).

6.2.3. LPT-Trr/MLL3/4 AND HISTONE MODIFICATIONS

6.2.3.1. Trimethylation of H3K4 on promoters

As discussed in the **Chapter 1**, the ancestral SET1 protein complex in basal species and its two orthologous complexes in mammals, SET1A and SET1B, are the main H3K4 di- and tri-methylating agents in the genome (Ardehali et al., 2011; Mohan et al., 2011; Wu et al., 2008). However, a small part of the H3K4 promoter trimethylation events is shared between Trx/MLL1/2 and LPT-Trr/MLL3/4 COMPASS-like complexes. Trx/MLL1/MLL2 is responsible for the fine-tuned regulation of HOX genes via regulating H3K4 trimethylation on their promoters (Wang et al., 2009a). LPT-Trr/MLL3/4, on the other hand, trimethylates H3K4 on gene promoters involved in hormone signaling in their role as hormone receptor co-activators (Chauhan et al., 2012; Goo et al., 2003; Lee et al., 2006a; Sedkov et al., 2003). Since these functions are hormone-related, different amounts of promoters might be controlled by LPT-Trr/MLL3/4 at different stages or organismal development (Chauhan et al., 2012; Sedkov et al., 2003). See **Figure 6.3A** for a graphical representation. LPT-

Trr/MLL3/M4's hormone-signaling related functions were discussed previously in **Chapter 4**.

6.2.3.2. Monomethylation of H3K4 on enhancers

Recently, the distinction between active and 'poised' (inactive) enhancers has become clearer. It was discovered that active enhancers showed H3K27ac enrichment, poised enhancers contained H3K27me3 instead, and both types of enhancers had H3K4me1 (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011). It was reported that the LPT-Trr complex in *Drosophila* was heavily involved in depositing H3K4 monomethylation marks on enhancers (Herz et al., 2012). The study demonstrated that loss of LPT or Trr led to a drastic decrease in H3K4me1 and H3K27ac levels predominantly on enhancers in S2 cells, unlike the loss of Set1, which mainly affected H3K4me2/3 levels. The effect on H3K27ac could be explained by the speculation that perturbing the trithorax-related complex by depleting LPT or Trr could lead to suboptimal functioning of the H3K27 demethylase UTX, also found in the trithorax-related complex. More specifically, in wild type circumstances, UTX demethylates H3K27me3 (repressive mark) and thus facilitates H3K27ac (activation mark) by CBP/p300 acetyltransferase (Herz et al., 2010, 2012; Tie et al., 2012). It was also reported that depletion of UTX led to increases in H3K27me3 (Herz et al., 2012) and decreases in H3K4me1 (Herz et al., 2010). These findings suggest that LPT, Trr and UTX all share an intertwined responsibility for the histone modification patterns on active enhancers.

Importantly, *trr*(RNAi) resulted in decrease in LPT, UTX and Trr protein levels, while *LPT*(RNAi) or *UTX*(RNAi) did not lead to a decrease in Trr levels (Herz et al., 2012). Thus, it was concluded that Trr had a stabilising role in the COMPASS-like complex in *Drosophila* (Herz et al., 2012).

Genome binding patterns of Trr, LPT and UTX proteins were assessed via ChIP-seq (Chromatin Immunoprecipitation sequencing). The three proteins showed greatest affinity for promoters and transcribed introns (enhancers) (Herz et al., 2012), chromatin states 1 and 3 respectively from Kharchenko et al.'s (2011) classification. There was significant (but not exclusive) overlap in genomic targets for the three proteins (Herz et al., 2012). This phenomenon could be reflective of either some degree of functionally different occupancy of the three proteins or noise in the experiment.

Similar associations of MLL3 with enhancer H3K4me1 and H3K27ac could also be found in mouse embryonic fibroblasts (MEFs) (Herz et al., 2012). Further studies demonstrated that, albeit partially functionally redundant, MLL3 and MLL4 were responsible for histone modification patterns on enhancer regions in human HCT116 cells and mouse embryonic stem cells (Hu et al., 2013). It was shown that in mammalian cells 70% of all enhancers suffered H3K4me1 depletion upon double knockout of *Mll3* and *Mll4* (Hu et al., 2013).

ChIP-seq experiments on S2 *Drosophila* cells after Trr depletion revealed that H3K4me1 levels were mostly affected in Kharchenko's chromatin states 3, 4 and 6 (Herz et al., 2012). The authors speculated that state 6 (Polycomb-repressed regions) could, at least partially, contain inactive/poised enhancers – with lower presence of

H3K4me1, depletion of H3K27ac and enrichment for H3K27me3 (Herz et al., 2012). State 3 had high enrichment for H3K4me1 and H3K27ac and as such was, most likely, representative of active anhancers (Herz et al., 2012; Kharchenko et al., 2011). State 4 had enrichment for H3K4me1 and not so much H3K27ac (but more than state 6), which could mean that state 4 is a transition state between inactive and active enhancers (Herz et al., 2012; Kharchenko et al., 2011). These findings led to the construction of a model where the trithorax-related complex combines the H3K27 demethylating activity of UTX and the H3K4 monomethylating abilities of LPT-Trr/MLL3/MLL4 to facilitate the acetylation of H3K27 by CBP/p300 on enhancers and thus lead to their activation (Herz et al., 2012). Refer to **Figure 6.3B** for a graphical summary.

MLL3/4's involvement in cancer could be related to aberrant H3K4me3-dependent promoter activation, but it most certainly has a lot to do with perturbed histone modification patterns on enhancers (Akhtar-Zaidi et al., 2012; Herz et al., 2014).

6.2.3.3. Monomethylation of H3K4 on promoters

MLL3/4-dependent H3K4 monomethylation was also found on gene promoters (Cheng et al., 2014). Cheng et al. (2014) investigated the histone modification landscape on promoter regions (3 kb upstream and downstream from the TSS) of skeletal muscle (C2C12) myoblasts and differentiated myotubes. Based on H3K4me1 and H3K4me3 distribution patterns, the researchers observed three distinct groups of promoters. Group 1 was characteristic of all active promoters and it had the above-mentioned pattern of TSS-proximal H3K4me3 and H3K4me1 distributed in the

areas around the H3K4me3 enrichment. Group 2 were promoters marked exclusively with H3K27me3 or lacking activating and repressive marks altogether. Interestingly, Group 3 exhibited lack of TSS-related H3K4me3, but had a widely distributed (spanning the TSS site) pattern of H3K4me1 instead, often combined with H3K27me3 expression. The authors presented evidence that Group 2 was, most likely, comprised of constitutive chromatin, while Group 3 contained regions of facultative chromatin (conditionally repressed genes). It was observed that MLL3 binding sites coincided with the H3K4me1-rich regions on Groups 1 and 3 and that depletion of *Mll3* led to the de-repressed transcription of genes belonging to Group 3. Even more genes belonging to Group 3 were de-repressed upon simultaneous depletion of *Mll3* and *Mll4*, resulting in increased myogenic differentiation (Cheng et al., 2014). An illustration of this model of MLL3/4 function on promoters is shown in **Figure 6.3C**.

The involvement of MLL3/4-mediated H3K4 monomethylation on repressed promoters was confirmed in embryonic fibroblasts, macrophages and embryonic stem cells (Cheng et al., 2014).

The researchers then focused on a set of Group 3 genes regulated by MLL3/4 and known to be involved in muscle differentiation (Cheng et al., 2014). Upon initiation of differentiation, there was a sharp decrease in H3K4me1 on TSS sites, which was replaced by H3K4me3 on TSS. Authors speculated that this occurrence might be related to the localisation of the H3K4me1 demethylase LSD1 (Lysine specific histone demethylase 1) on the TSS of those genes. It was shown that after the depletion of monomethylated H3K4 and addition of trimethylated H3K4 on TSS, ING1 and Sin3A also accumulated on TSS (Cheng et al., 2014). H3K4me3 promoter regions are

characterised by the recruitment of a great amount of chromatin Reader and Reader/Writer complexes binding to the H3K4me3 modification (Yun et al., 2011). One such example is the Reader ING1, which does not have any catalytic domains, but, once bound to H3K4me3, attracts the Sin3A/HDAC complex (Peña et al., 2008). The recruitment of a repressive complex to actively transcribed genomic regions has been widely reported, but its significance is still debated.

Cheng et al. (2014) concluded that H3K4me1 regions act as a spatial boundary setter for the binding of Readers and Writers of the histone code (ING1 and Sin3A specifically) on the TSS.

Interestingly, MLL3/4's effect on histone methylation patterns on promoters of repressed genes did not seem to be related to the amount of H3K27me3 on those promoters (Cheng et al., 2014). This finding is surprising in the light of Herz et al.'s (2014) study implicating LPT-Trr/MLL3/4's H3K27 demethylating partner UTX in maintaining the distinction between active and inactive enhancers. This discrepancy could be explained by the speculation that UTX acts in concert with MLL3/4 only on enhancers, but not on promoters. Tentative support for this interpretation is provided by the fact that UTX can be recruited by proteins other than MLL3/4 (Seenundun et al., 2010) and that ChIP-seq experiments showed that LPT, Trr and UTX did not overlap completely in their targets (Herz et al., 2012). These findings point towards UTX having a dual identity – the first – as part of the trithorax-related complex and the second – as an independent agent.

The modes of LPT-Trr/MLL3/4's action proposed by the studies mentioned in this chapter are summarised in **Figure 6.3.**

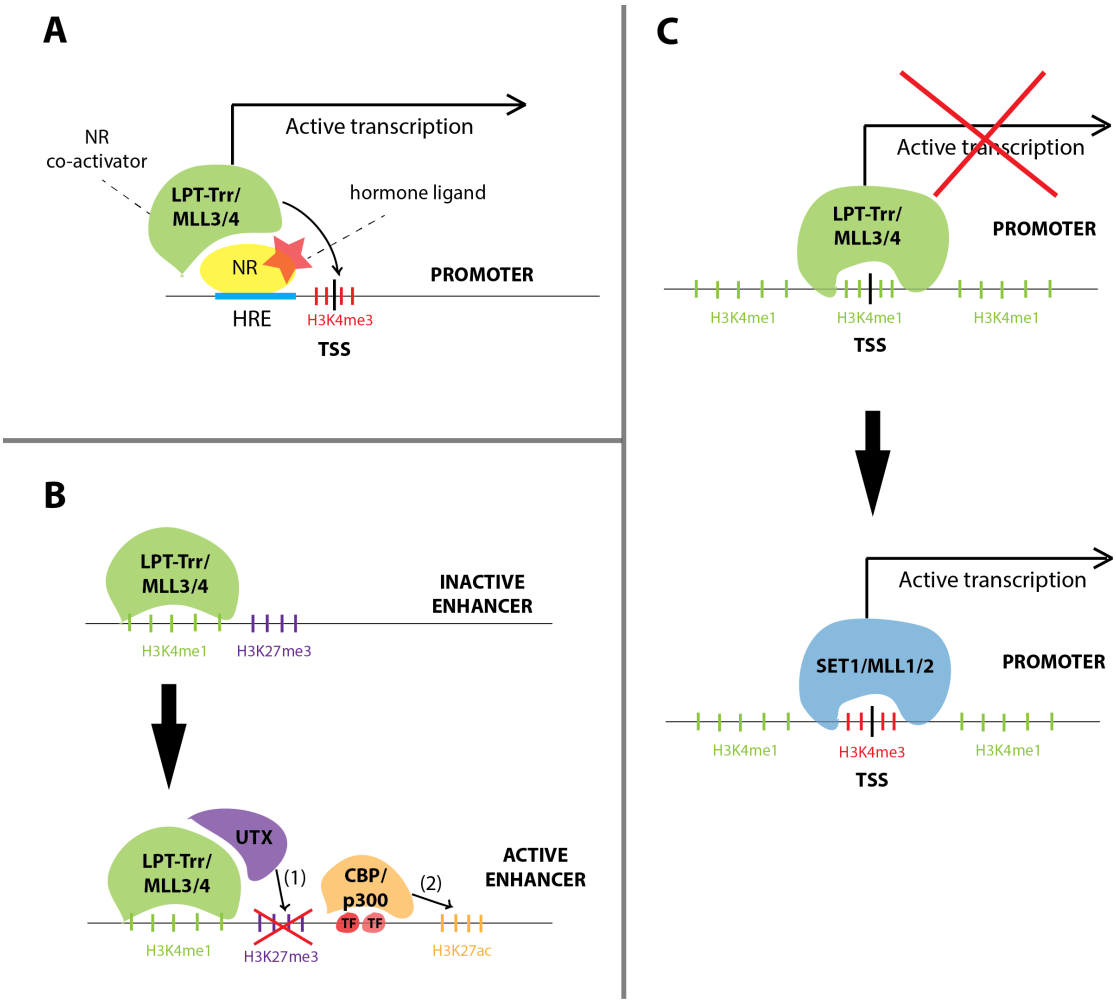


Figure 6.3. The three proposed modes of action of LPT-Trr/MLL3/4. (**A**) In response to hormone signaling, the nuclear receptor (NR) binds to the hormone response element (HRE), located on a gene's promoter. In the presence of a hormone, like *Drosophila*'s ecdysone, the LPT-Trr/MLL3/4 COMPASS-like complex acts as a co-activator to the nuclear receptor. The COMPASS-like complex trimethylates H3K4 in the TSS region and thus activates transcription. (**B**) LPT-Trr/MLL3/4 complex performs H3K4 monomethylation on both active and inactive enhancers. Upon activation, LPT-Trr/MLL3/4's H3K27 demethylating partner UTX removes the H3K27me3 mark (event 1) and then allows for the transcription factor-bound CBP/p300 to perform acetylation of H3K27 (event 2) and thus activate the enhancer. (**C**) LPT-Trr/MLL3/4-dependent H3K4 monomethylation over the TSS of a promoter prevents transcriptional activation. The promoter is activated when the H3K4me1 mark is removed from the TSS and LPT-Trr/MLL3/4 is replaced by SET1, MLL1 or MLL2, which trimethylates H3K4 and activates transcription.

6.2.3. Aim of chapter

The goal of the current chapter is to observe the relationship between histone modification changes and cellular dynamics under trithorax-related knockdown conditions. Histone modifications are investigated on both whole organismal level and G2/M stem cells. Regeneration and homeostatic conditions following knockdown are investigated. Finally, the role of functional redundancy of *Smed-trr-1* and *Smed-trr2* is explored in the context of histone lysine methylation.

6.3. Materials and Methods

6.3.1. Reagents

The reagents used for experiments in this chapter are listed in **Appendix A**.

6.3.2. RNAi of Smed-set1 and Smed-utx

DsRNA was dissolved to 2 μg/μl in PCR-grade water. Animals were injected for 4 consecutive days, after which they were either left whole or amputated on the next day.

6.3.3. Double RNAi

Smed-LPT, Smed-trr-1, Smed-trr-2 and *GFP* dsRNA were each dissolved to 4 μg/μl. For each of the double knockdown conditions a working solution was made containing 1:2 of each desired dsRNA. For example, *Smed-trr-1/Smed-trr-2* dsRNA working

solution contained 1 part *Smed-trr-1* and 1 part *Smed-trr-2* dsRNA. This resulted in each dsRNA in the working solution being at a concentration of 2 µg/µl with a final solution concentration of 4 µg/µl.

6.3.4. Fluorescence Activated Cell Sorting (FACS)

Romero et al.'s (2012) protocol was used. Cells were kept on ice until the FACS procedure started. Immediately before sorting, propidium iodide was added (1 µg/ml) to each cell vial in order to separate the living from the dead cells.

A Becton Dickinson FACSCalibur machine with a UV laser was used for the cell sorting. Hoechst 33342 was detected in blue (461 nm; nuclear size measure) and red (617 nm; cytoplasmic complexity measure). The gating logic was very similar to the one described in Romero et al. (2012). The final gates set for the sorts are shown in **Appendices J** and **K**.

6.3.5. Protein extract from whole worms

Protein was extracted from 5-20 animals per condition. Animals were homogenised with an electric homogeniser in a solution containing cOmplete protease inhibitors, PMSF and PBS. 2x Loading Buffer and DTT were added and the samples denatured at 100°C.

Protein concentration was measured with Qubit, then samples were snap-frozen on dry ice and stored at -20°C or used straightaway in a Western blot procedure.

6.3.6. WHOLE PROTEIN EXTRACT FROM X1 SORTED CELLS

The sorted cells (minimum $1x10^5$ cells) collected in PBS in a protein low-binding microtube were pelleted and the same method of protein extraction and quantification as in **6.3.6.** was used.

6.3.7. WESTERN BLOT

Since all of the histone modification antibodies used in this chapter's experiments were in the same size range (15-20 kDa), the antibody against unmodified histone H3 could not be used as a loading control on the same gel as the histone modification of interest. Therefore, for each experiment reported in this chapter, the first step was to run a calibration procedure. The same amount of whole protein (5-30 μg) from each knockdown condition (according to Qubit measures) was run on a 4-12% Bis/Tris gel, the PVDF membrane containing the transferred proteins was stained with Ponceau S whole protein stain (if protein material met the detection limit of Ponceau S) and histone H3 unmodified was detected. Based on these initial results, the amount of whole protein was modified between conditions (if necessary) so that even H3 bands (and Ponceau S staining, if used) could be achieved. When Western blots on x1 cell extract were performed, the amount of whole protein was too low for detection with Ponceau S. In this case, only a calibration H3 (unmodified) staining was used and amounts loaded were modified (if necessary) for subsequent procedures.

The Western blot procedure used 4-12% Bis/Tris gels run at 200 V/125 mA for 30-50 minutes (depending on whether one or two gels were run simultaneously). The gels were run in 1x NuPAGE® MES SDS Running Buffer using XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen). A wet transfer on a PVDF membrane (soaked for 1 minute in Methanol) was done using a Mini Trans-Blot Electrophoretic Transfer Cell (BioRad). The transfer was run for 1.5 hours in chilled conditions (ice packs used) at 125 V/300 mA. If the Western blot was on whole animal protein extracts, the membrane was washed with distilled water and stained for 5 minutes at RT (shaking) with Ponceau S solution. A couple of washes were performed with distilled water in order to increase the definition of Ponceau S staining pattern. In order to reverse the staining, 0.1M sodium hydroxide treatment was used, followed by abundant distilled water washes. If the Western blot procedure was on cell extracts, this step was skipped. Membrane was blocked for 1 hour in 3% skimmed milk in PBT (PBS, 0.05% Tween-20). Block solution was changed with 3% skimmed milk in PBT containing 1:1000 of the primary antibody of interest. The samples were left at 4°C on a rocker ON. Membrane was washed for 1 hr at RT on a rocker in Block solution and anti-rabbit- or anti-mouse-HRP (as necessary) antibody was added to the Block solution to a final concentration of 1:4000. The membrane was left in this solution for 3 hours at RT. The membrane was washed for 1 hour in PBT. Samples were developed with SuperSignal West Pico Trial (ThermoFisher Scientific) as per manufacturer's instructions. The membrane was imaged using GeneSys software on G:BOX Syngene gel doc. Exposure times varied depending on the samples.

6.4. RESULTS AND DISCUSSION

Numerous studies on planarians have demonstrated that there are important differences between planarian homeostatic tissue maintenance and regeneration. Alterations are associated with rates of stem cell renewal and differentiating neoblast progeny production. In order to investigate the difference between regenerative and homeostatic conditions under different knockdown conditions, the following markers were chosen – *Smed-H2B*, anti-SMEDWI-1 antibody, *Smed-AGAT-1* and *Smed-NB.21.11e.* Antibodies against H3K4me3, H3K4me1 and H3K27me3 were used to study histone modification dynamics associated with cellular changes. In order to check the specificity of the chosen antibodies, *Smed-set1* (RNAi) and *Smed-utx*(RNAi) were used. A brightfield view of the *Smed-set1*(RNAi) phenotype characterised by a smaller regenerative blastema and delayed eye formation, is shown in **Figure 6.4A** and it is in agreement with previously reported data (Hubert et al., 2013).

SET1 is known to perform the bulk of H3K4me3 and small portions of the H3K4me2 and H3K4me1 (Mohan et al., 2011). Hence, *Smed-set1*(RNAi) was used as a positive control in each experiment. UTX, on the other hand, is known to demethylate H3K27me3. Knockdown of *Smed-utx* resulted in a phenotype, similar to that observed following *Smed-set1*(RNAi) (**Figure 6.4B**). It was hypothesised that, if the chosen anti-H3K27me3 antibody was specific, a Western blot should detect increased H3K27me3 levels following *Smed-utx*(RNAi). This expectation was confirmed for *Smed-utx*(RNAi) homeostatic, but not regeneration conditions (**Figure 6.4C**). Reasons for this discrepancy are discussed later in this chapter.

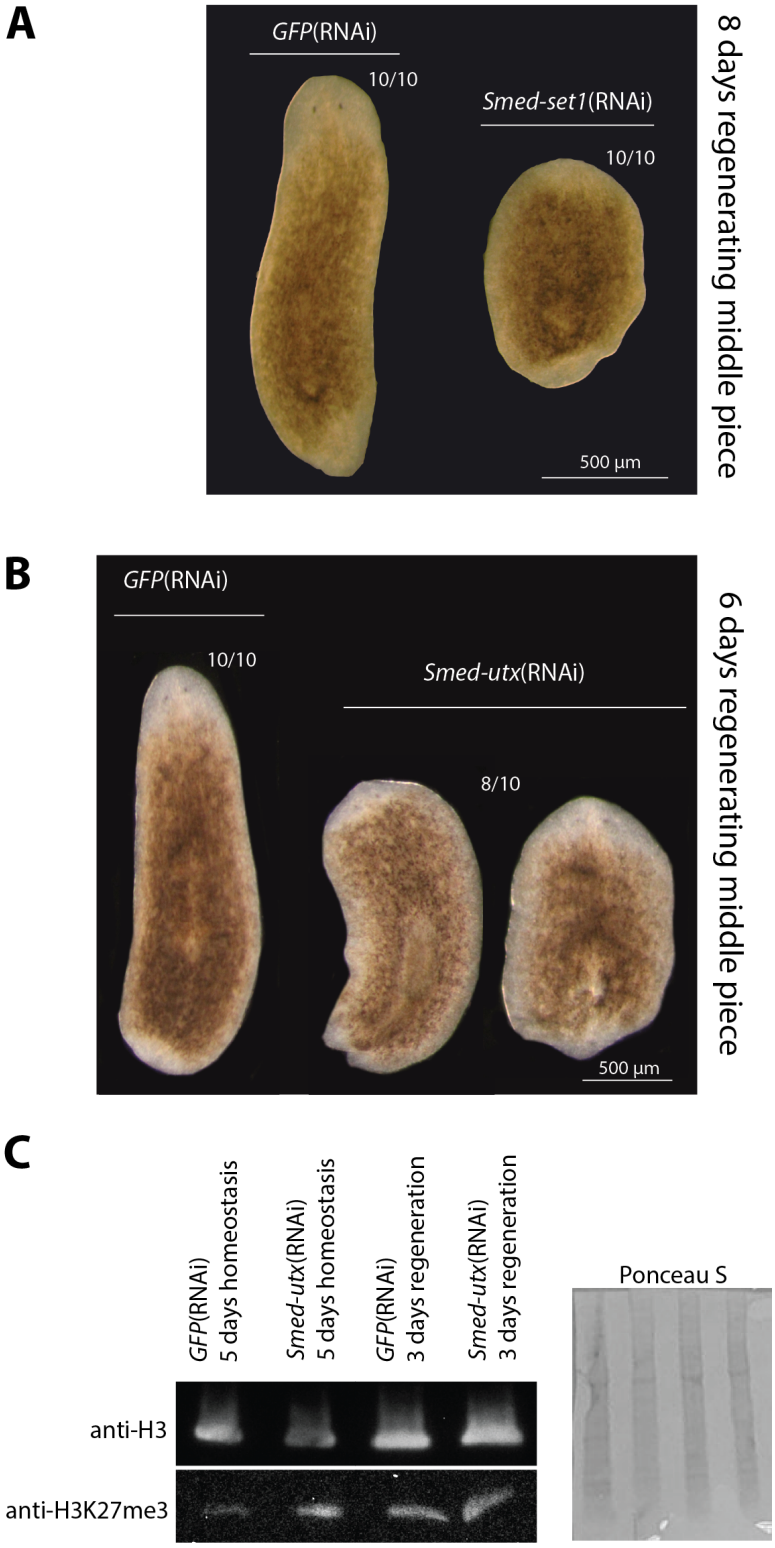


Figure 6.4. *Smed-set1*(RNAi) and *Smed-utx*(RNAi). (**A**) Brightfield images of 8-day regenerating middle pieces following *GFP*(RNAi) or *Smed-set1*(RNAi). Ten worms per condition were used. (**B**) Brightfield images of 6-day regenerating middle pieces following *GFP*(RNAi) or *Smed-utx*(RNAi). Ten worms per condition were used. (**C**) Western blot with anti-H3 (unmodified) and anti-H3K27me3 antibodies on 5-day homeostatic and 3-day regenerating animals. Ponceau S whole protein staining is also presented.

Smed-mbd2/3(RNAi) was also added to these experiments. This gene was reported to control differentiation (Jaber-Hijazi et al., 2013). It was reasoned that it could be possible to draw parallels between differentiation-related histone modification changes caused by *Smed-mbd2/3* and trithorax-related genes knockdown.

An important remark regarding this chapter is that it aims to find general patterns of cellular changes associated with general patterns of histone modification changes. It is not within the scope of the present text to elucidate epigenetic changes on specific genes. This chapter serves mostly as a foundation for future work.

6.4.1. Trithorax-related genes control H3K4me1 and H3K27me3 levels during homeostatic conditions

Day 5 homeostasis was chosen for these experiments because it was assumed that five days after the final dsRNA injection would be sufficient time to overcome proliferative alterations due to wounding. There were no significant differences found in stem cell (*Smed-H2B*-positive) and early stem cell progeny (*Smed-H2B*-negative/SMEDWI-1-positive) dynamics between conditions (**Figure 6.5**).

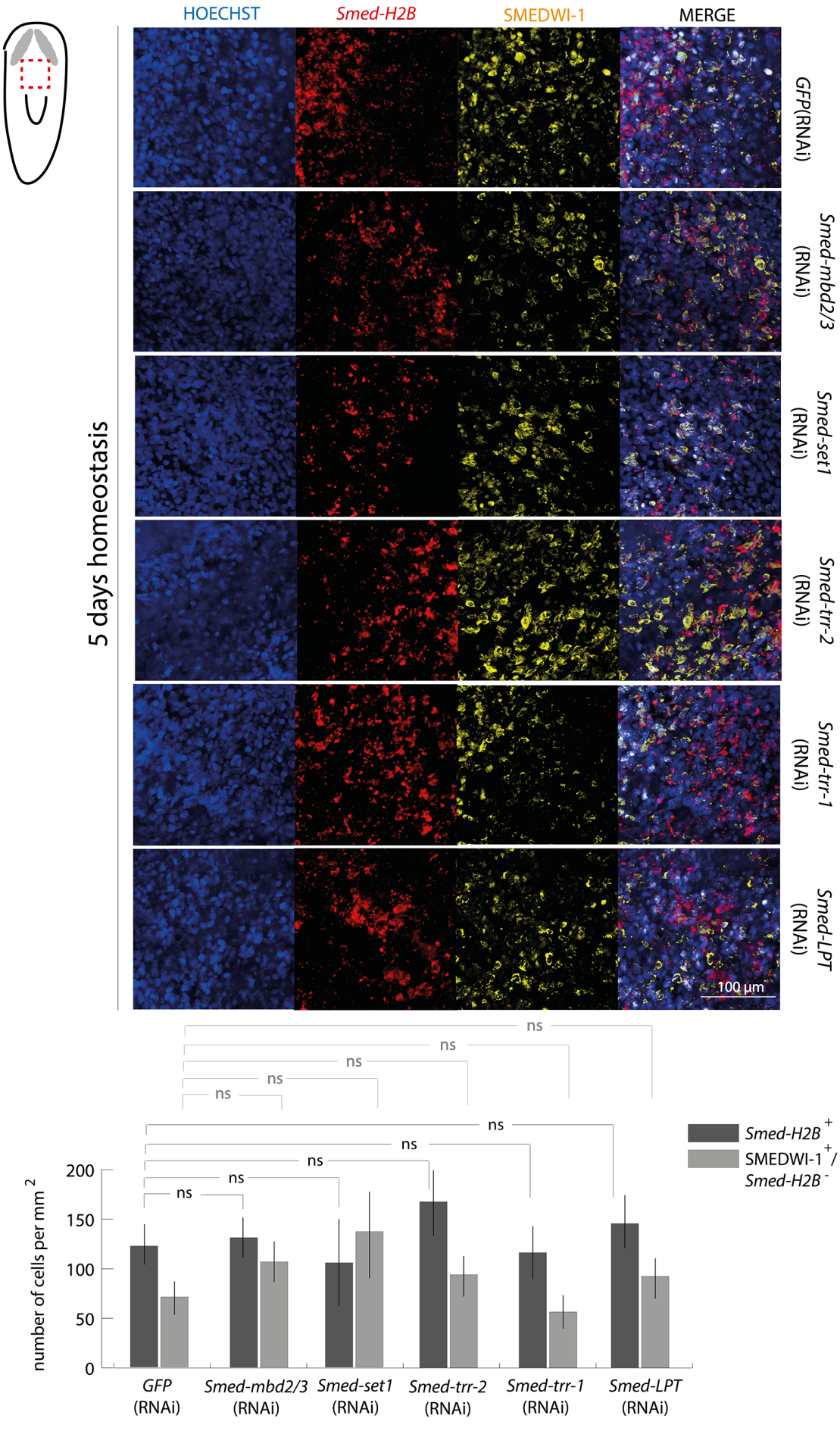

HOECHST
Smed-H2B
SMEDWI-1
MERGE
5 days homeostasis
GFP(RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)
100 µm
ns
Smed-H2B+
SMEDWI-1+/ Smed-H2B-
number of cells per mm2
200
150
100
50
0
GFP (RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)

Figure 6.5. Stem cell and early stem cell progeny dynamics at day 5 homeostasis following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1*. Knockdown animals were stained with *Smed-H2B* RNA probe and anti-SMEDWI-1 antibody. Representative pre-pharyngeal area staining for each condition is shown. Ten worms per condition were used. The bar graph shows mean average cell count per condition with error bar representing SEM. A 2-tailed ttest assuming unequal variance was used where 'ns' stands for 'not significant'.

Next, the state of epidermal progenitor cells was examined using the markers *Smed-NB.21.11e* and *Smed-AGAT-1.* The only significant changes noted at this point were that in *Smed-LPT*(RNAi) and *Smed-mbd2/3*(RNAi) conditions *Smed-AGAT-1*-positive cells were decreased (**Figure 6.6**). The results for *Smed-mbd2/3*(RNAi) were in agreement with previous work by Dr. Jaber-Hijazi who demonstrated that at one week of homeostasis *Smed-NB.21.11e* expression did not appear different to control, but *Smed-AGAT-1* expression was decreased (Jaber-Hijazi et al., 2013). Interestingly, *Smed-LPT*(RNAi) samples at 5 days homeostasis were significantly different from controls when it came to *Smed-AGAT-1* expression. This is in conflict with the data presented in **Chapter 4** where *Smed-AGAT-1*-positive cells in *Smed-LPT*(RNAi) were not significantly decreased at 2 weeks of homeostasis. There was, however, a non-significant decrease. This discrepancy could be explained by the variability of *Smed-LPT*(RNAi) phenotype. Specifically during homeostasis, phenotype progression had been noticed to show considerable variations. For example, different RNAi batches were shown to vary in their time of death quite considerably (any time between 2 weeks and 4 weeks homeostasis). Such inconsistencies were observed during the homeostatic, but not regeneration phenotype. It is possible that the act of amputation serves as a synchronising event.

Alternatively, the discrepancy between early and late homeostatic timepoint could be reminiscent of what was previously observed during regeneration and tumour formation. **Chapter 4** demonstrated that there was a significant down-regulation of *AGAT-1*-positive cells at day 8 regeneration, but the presence of such cells was re-established at later tumourigenic stages of the phenotype (**Chapter 5**).

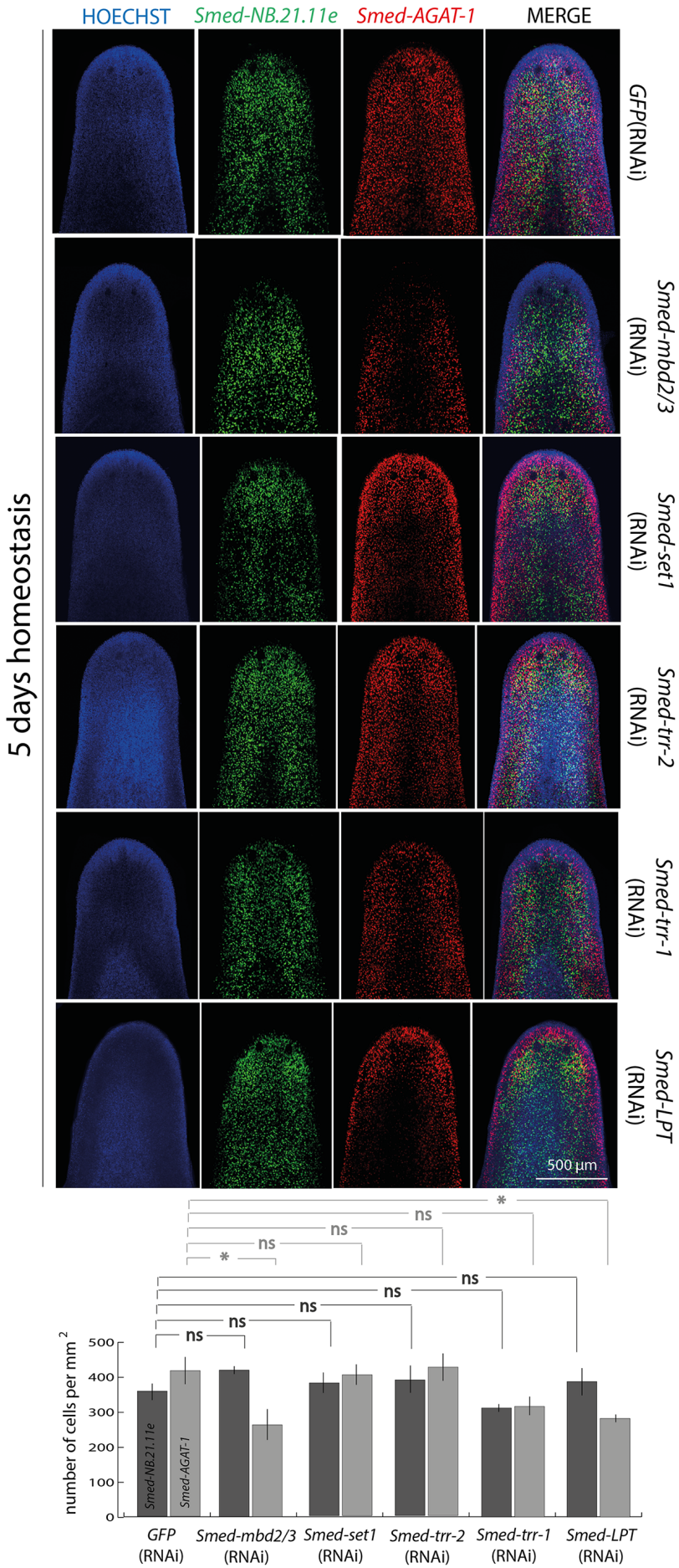

HOECHST
Smed-NB.21.11e
Smed-AGAT-1
MERGE
5 days homeostasis
GFP(RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)
500 µm
number of cells per mm 2
0
100
200
300
400
500
Smed-NB.21.11e
Smed-AGAT-1
ns
*
GFP (RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)

Figure 6.6. Early and late epidermal neoblast progeny dynamics at day 5 homeostasis following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1*. Knockdown animals were stained with *Smed-NB.21.11e* and *Smed-AGAT-1* RNA probes. Representative anterior area staining for each condition is shown. Ten worms per condition were used. The bar graph shows mean average cell count per condition with error bar representing SEM. A 2-tailed ttest assuming unequal variance was used; $p<0.05$ is marked with an asterisk and 'ns' stands for 'not significant'.

Having gained some insight into stem cell and progeny dynamics at day 5 homeostasis, the next step was to establish the connection between these changes and histone modifications. The three modifications of interest (H3K4me1, H3K4me3 and H3K27me3) were used for Western blot analysis on x1-sorted cells (G2/M) and whole animals.

Worms from the same injection batch as the experiments described above (**Figure 6.5** and **Figure 6.6**) were used for FACS-sorting of x1 population of stem cells. It is known that neoblasts are present in both x1 (G2/M cells) and x2 (G1 cells) FACS populations. However, due to the fact that x2 contains stem cell progeny as well, x1 was deemed a 'cleaner' choice for focusing on stem cells. The dot plots showing the gating during the FACS-sorting are shown in **Appendix J**. Protein was extracted from the sorted x1 populations for each knockdown and Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 were performed (**Figure 6.7A**). *Smed-set1*(RNAi) and *Smed-LPT*(RNAi) led to a decrease in H3K4me1 in x1 cells, with *Smed-set1*(RNAi) also leading to an expected decrease in H3K4me3. Interestingly, *Smed-mbd2/3*(RNAi), *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) resulted in an increase in H3K4me1. Two hypotheses might be drawn from these results. First,

the decrease in *Smed-AGAT-1*-positive cells following *Smed-LPT*(RNAi) and *Smed-mbd2/3*(RNAi) are governed by different mechanisms, as shown by the opposing effects on H3K4me1 levels. Second, *Smed-trr-1* and *Smed-trr-2* are, most likely, overcompensating for each other's loss of H3K4 methylating function, as shown by the dramatic increase in H3K4me1 following both *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi).

Finally, an intriguing observation was that H3K27me3 levels seemed to be up-regulated in all tested conditions (apart from *Smed-set1*(RNAi), for which there was insufficient material for the procedure). This could mean that, in the absence of a trithorax-related gene (*Smed-LPT*, *Smed-trr-1* or *Smed-trr2*), the H3K27me3 demethylase located in the same complex (UTX) cannot function optimally, thus leading to an accumulation in H3K27me3. This interpretation is supported by previous reports of Trr having a stabilising role of the trithorax-related complex and depletion of *trr* leading to depletion of UTX in *Drosophila* (Herz et al., 2012).

This experiment was repeated using whole animals' protein extract instead of sorted cells' (**Figure 6.7B**). In the context of the whole animal, it could be seen that knockdown of either of the three trithorax-related genes led to a decrease in H3K4me1. This suggests that the overcompensation of H3K4 monomethylation in x1 cells from *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) animals did not apply to every cell type. The decrease in H3K4 monomethylation following *Smed-LPT*(RNAi), however, seemed to affect the whole animal.

It is important to point out that, these experiments do not give us resolution into whether H3K4me1 changes are affecting promoters or enhancers or both. It could be

helpful to see the dynamics of H3K27ac (characteristic of active enhancers), known to be down-regulated upon *LPT/trr/MLL3/4* depletion (Herz et al., 2012).

H3K4me3 on the whole organismal level was affected only under *Smed-set1*(RNAi) conditions, but not as drastically as when looking at the level of x1 population alone.

The increased amount of H3K27me3 in all experimental samples was in full agreement with the results from x1 cells.

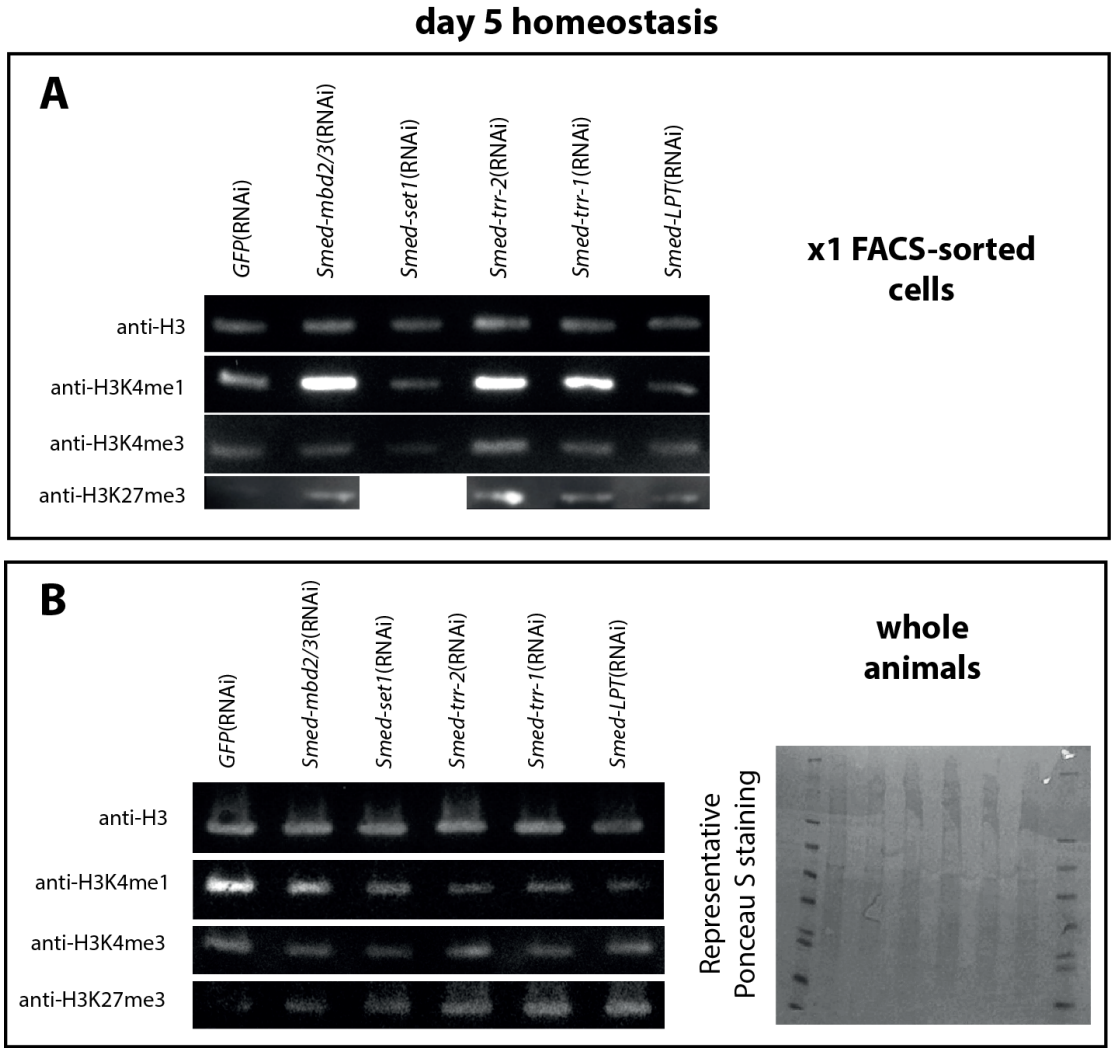


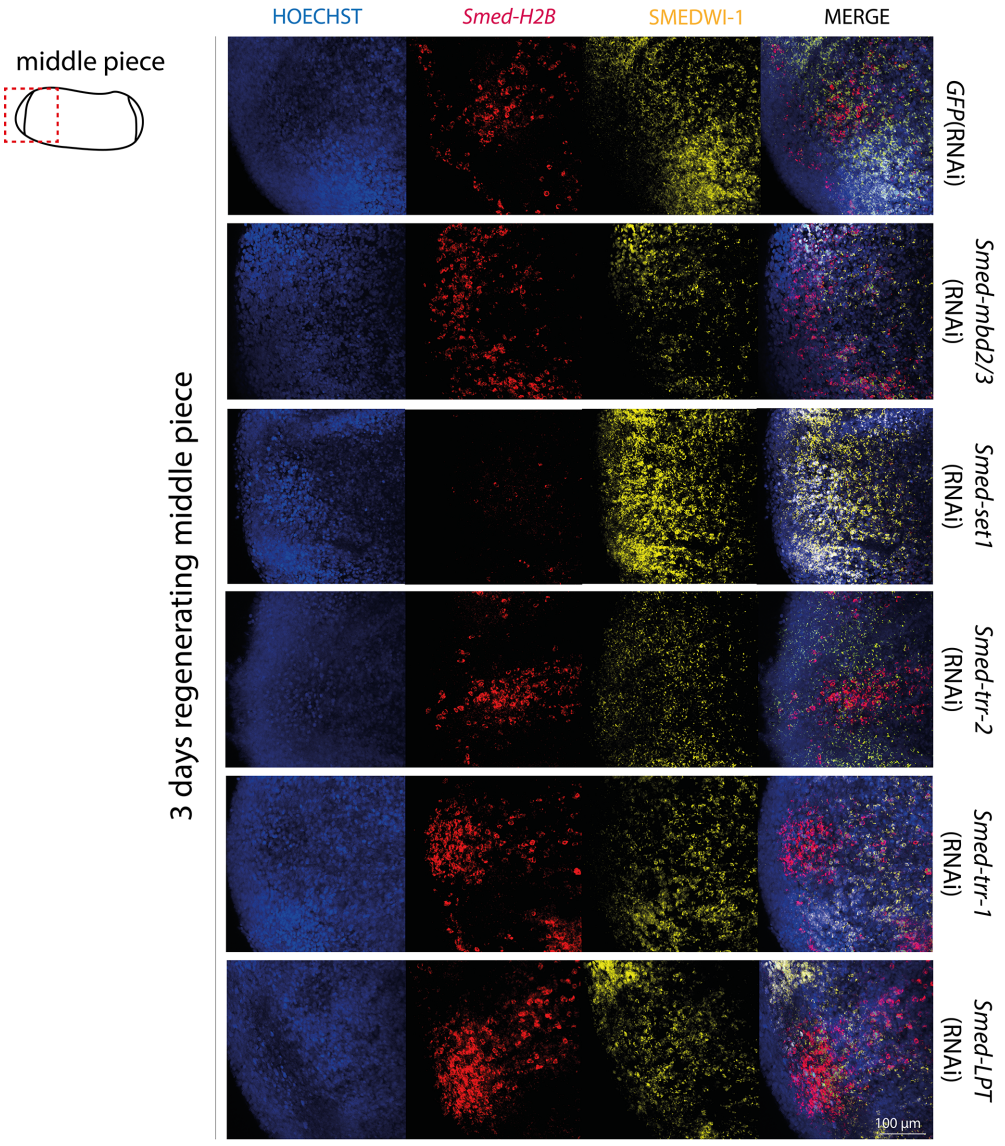
Figure 6.7. Histone modification changes at day 5 homeostatic knockdown phenotypes. (**A**) Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from x1 FACS-sorted cells. Three technical replicates were done for H3K4me1. (**B**) Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from whole animals. An example Ponceau S-stained PVDF membrane prior to antibody incubation is shown.

In conclusion, it was observed that the decrease in the *Smed-AGAT-1*-positive epidermal progenitor cells in *Smed-LPT*(RNAi) and *Smed-mbd2/3*(RNAi) at day 5 homeostasis was reflected in differential effects on H3K4me1, but similar up-regulation of H3K27me3 (both in x1 cells and whole animals). Thus, it could be speculated that H3K27me3 levels influenced late epidermal progenitor cell production. One potential mechanism for this is increased H3K27me3-mediated repression of epidermal fate commitment genes following knockdown.

6.4.2. Trihtorax-related genes regulate histone modifications differently during regeneration compared to homeostasis

At day 3 of regeneration, none of the chosen knockdowns showed an obvious brightfield phenotype yet. Hence, this timepoint was selected in the hope that early (causative) cellular and histone modification changes could be observed without the impact of secondary ('contaminating') changes. Stem cell and early stem cell progeny dynamics were assessed in regenerating middle pieces (**Figure 6.8**). There was an overall decrease in stem cell early progeny production in *Smed-mbd2/3*(RNAi) and *Smed-trr-1*(RNAi). Noteworthy was also the increased average (not significantly different) amount of early stem cell progeny following *Smed-set1*(RNAi).

The amount of stem cells was significantly increased following *Smed-LPT*(RNAi), but decreased following *Smed-set1*(RNAi). The increased number of neoblasts in the *Smed-LPT*(RNAi) condition is in agreement with previously reported (**Chapter 5**) increase in the number of proliferative cells. The observed stem cell depletion in *Smed-set1*(RNAi) confirms previous data (Hubert et al., 2013).


middle piece
HOECHST
Smed-H2B
SMEDWI-1
MERGE
3 days regenerating middle piece
GFP(RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)
100 µm

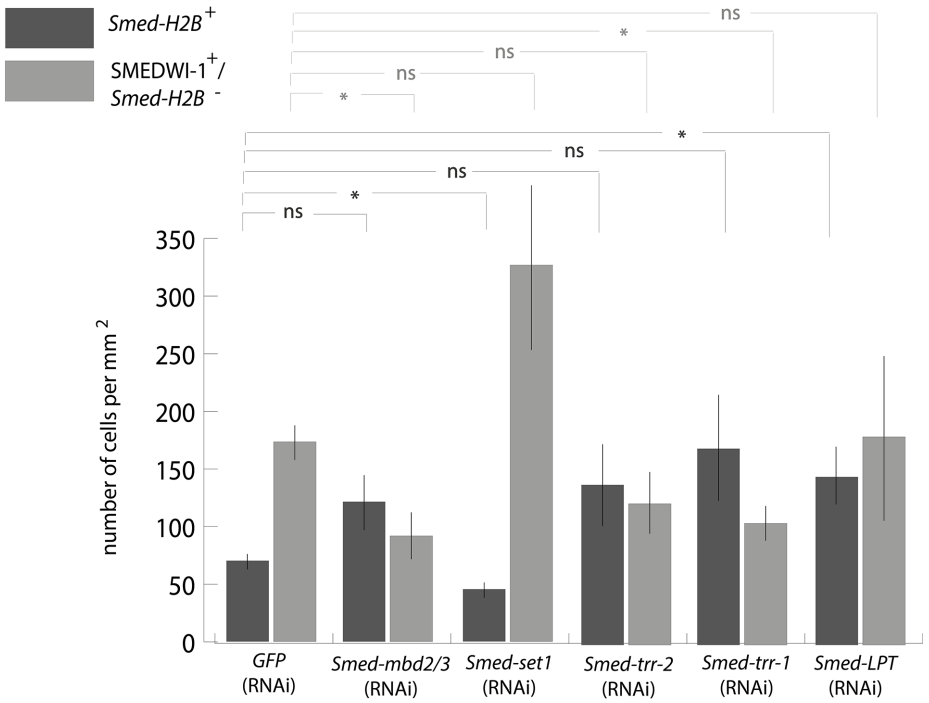

Smed-H2B+
SMEDWI-1+/ Smed-H2B -
number of cells per mm 2
0
50
100
150
200
250
300
350
ns
*
ns
ns
*
ns
*
ns
ns
*
GFP (RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)

Figure 6.8. Stem cell and early stem cell progeny dynamics at day 3 regeneration following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1*. Knockdown regenerating middle pieces were stained with *Smed-H2B* RNA probe and anti-SMEDWI-1 antibody. Representative anterior area staining for each condition is shown. Ten worms per condition were used. The bar graph shows mean average cell count per condition with error bar representing SEM. A 2-tailed ttest assuming unequal variance was used. Significant changes ($p<0.05$) are marked with an asterisk and non-significant ones – with 'ns'.

When the early and late epidermal progenitor production was addressed using *Smed-NB.21.11e* and *Smed-AGAT-1* markers respectively, it was found that *Smed-trr-2*(RNAi) and *Smed-LPT*(RNAi) animals showed a decreased production of early epidermal progenitors (**Figure 6.9**). These two knockdown conditions, together with *Smed-mbd2/3*(RNAi) animals, also showed a decrease in late epidermal progenitors. The *Smed-LPT*(RNAi) findings are in line with previously reported (**Chapter 4**) decrease in epidermal progenitor production at later stages of regeneration (day 8). *Smed-trr2*(RNAi) animals were significantly affected in their production of epidermal progenitors, which is surprising, considering the fact that this was not the case at later stages of regeneration such as day 8 (**Chapter 4**). This could mean the initial delayed speed in the production of these cells was rectified as regeneration progressed. Alternatively, the lower rate of epidermal progenitor production under *Smed-trr-2*(RNAi) conditions could be resulting in making longer lived, but fewer epidermal progenitors, which slowly accummulate and thus mask the defect at later stages of regeneration. In the case of *Smed-mbd2/3*(RNAi), the results for day 3 regeneration match Dr. Jaber-Hijazi's unpublished observations where there was an obvious downregulation of *Smed-AGAT-1*-positive cells, but not *Smed-NB.21.11.e*.

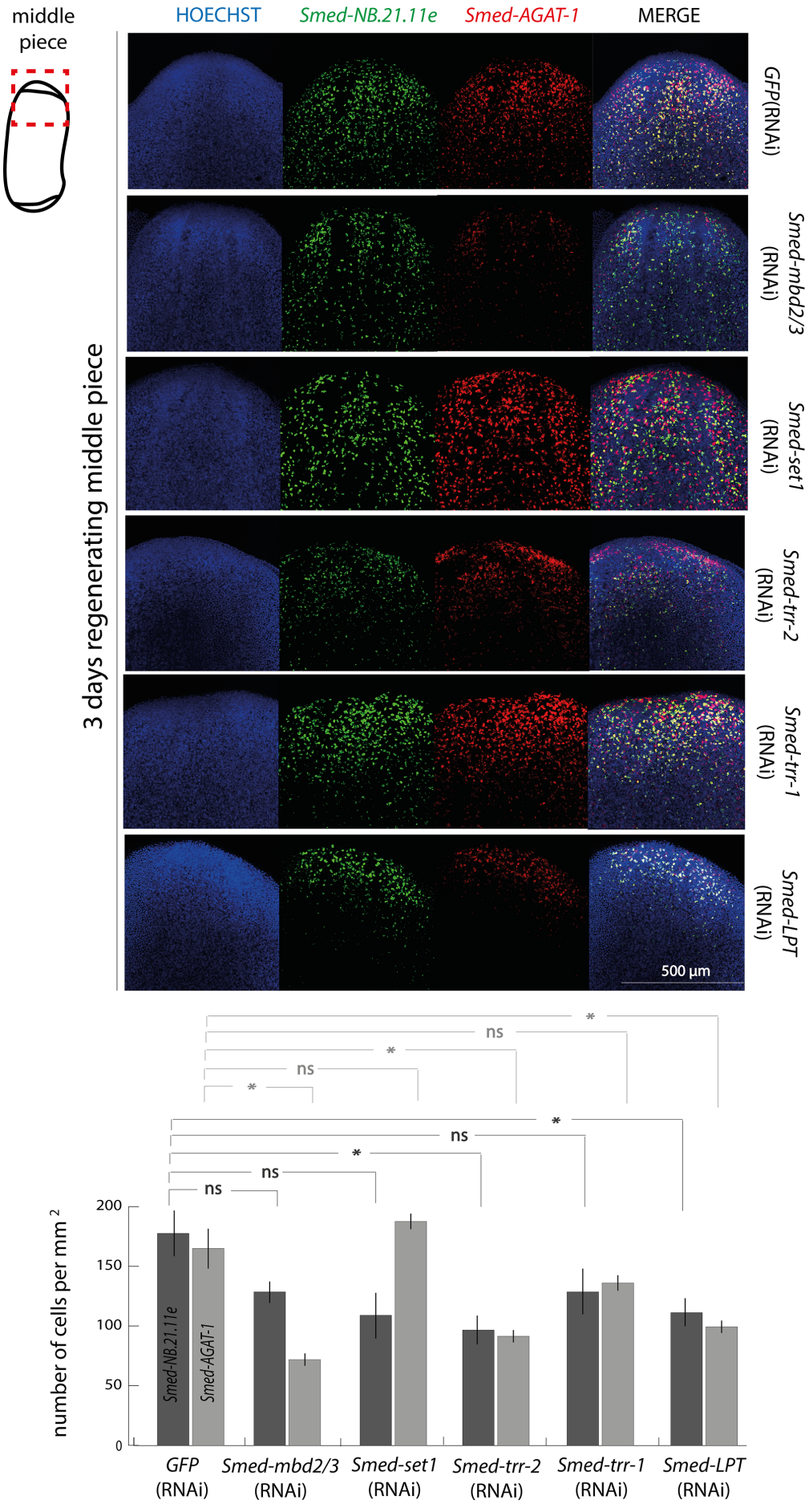

middle piece
HOECHST
Smed-NB.21.11e
Smed-AGAT-1
MERGE
3 days regenerating middle piece
GFP(RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)
500 µm
number of cells per mm 2
0
50
100
150
200
Smed-NB.21.11e
Smed-AGAT-1
ns
*
GFP (RNAi)
Smed-mbd2/3 (RNAi)
Smed-set1 (RNAi)
Smed-trr-2 (RNAi)
Smed-trr-1 (RNAi)
Smed-LPT (RNAi)

Figure 6.9. Early and late epidermal neoblast progeny dynamics at day 3 regeneration following knockdown of trithorax-related genes, alongside with knockdowns of *Smed-mbd2/3* and *Smed-set1*. Knockdown regenerating middle pieces were stained with *Smed-NB.21.11e* and *Smed-AGAT-1* RNA probes. Ten worms per condition were used. Representative anterior area staining for each condition is shown. The bar graph shows mean average cell count per condition with error bar representing SEM. A 2-tailed ttest assuming unequal variance was used; $p<0.05$ is marked with an asterisk and 'ns' stands for 'not significant'.

The next step in the analysis was to correlate the changes in cell dynamics to histone modifications. For this purpose, knockdown regenerates from the same injection batch that was used for the experiments in **Figure 6.8** and **Figure 6.9**, were used for FACS-sorting of x1 population. The sorting gates used for this procedure are shown in **Appendix K**. Protein was extracted from the sorted x1 cells and Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 were performed (**Figure 6.10**). Results showed that the increased amount of H3K4me1 in *Smed-mbd2/3*, *Smed-trr-1* and *Smed-trr-2* knockdown animals observed at day 5 homeostasis was not present at 3 days of regeneration and no clear difference to control was seen for these samples (**Figure 6.10A**). There was a slight decrease in H3K4me1 under *Smed-LPT*(RNAi) conditions, but the effect was minor. This finding was surprising, considering the fact that planarian regeneration is a demanding process where, often, phenotypic defects become more obvious than in homeostatic conditions.

There was a clear decrease in the level of H3K4me3 in *Smed-trr-2*(RNAi) and *Smed-set1*(RNAi) x1 stem cells (**Figure 6.10A**). The effect on H3K4me3 in the *Smed-set1*(RNAi) was expected. The drastic effect seen in *Smed-trr-2*(RNAi), but not in

Smed-trr-1(RNAi) or *Smed-LPT*(RNAi), however, was surprising. There was also a sharp decrease in H3K27me3 levels under *Smed-trr-2*(RNAi) conditions. What these findings could indicate is that the decrease in epidermal progenitor production following *Smed-trr-2*(RNAi) could be related to lack of activation of epidermal lineage commitment genes (lack of H3K4me3) and decreased repression of stem cell-specific genes (decrease in H3K27me3). This interpretation is supported by the elevated average number of stem cells (not significantly) after *Smed-trr-2*(RNAi) (**Figure 6.8**). Nevertheless, these hypotheses warrant further investigation.

One clear finding was that the pattern of the H3K27me3 mark under regenerative conditions for all experimental knockdowns was unlike the pattern seen under homeostatic conditions. In homeostasis, x1 cells (and the whole animal) showed increased amounts of H3K27me3 following trithorax-related genes knockdown. However, during early regeneration, H3K27me3 was either unchanged or reduced under those knockdown conditions in x1 cells.

Investigation of histone modifications at the level of the whole regenerating organism, revealed deepening complexity (**Figure 6.10B**). The inverse H3K27me3 pattern discussed above was more obvious in whole animals. These findings would suggest that UTX (or a H3K27me3-demethylating substitution mechanism) is acting under regenerative, but not homeostatic conditions after trithorax-related genes knockdown. Alternatively, another H3K27 demethylating agent might be replacing or helping UTX during regeneration. One potential candidate could be LSD1, which usually demethylates mono- and dimethylated lysines (Huang et al., 2007). The findings shown in Figure 6.4C are in agreement with such interpretation.

There was a minor increase in the amount of H3K4me1 following *Smed-LPT*(RNAi) and *Smed-trr-1*(RNAi), potentially reflective of the increase of this modification in cells other than x1. This could be due to the activation of some compensatory H3K4 monomethylation mechanisms that were activated during regeneration. For example, if the trithorax-related complex was perturbed by the knockdown of one of its subunits, it is possible that other histone methylases, like SET1, would act as substitutes.

One of the few easily reconcilable findings was that, like in x1 cells, H3K4me3 was downregulated in whole regenerating organisms under *Smed-set1*(RNAi) conditions. A possible emerging model for SMED-SET1's function is that the protein's H3K4 trimethylating activity in stem cells is related to transcriptional activation of stem cell maintenance genes. Upon knockdown of *Smed-set1*, H3K4me3 decreases, as a result of which existing stem cells, no longer having a stable stem cell maintenance mechanism in place, transition into neoblast progeny cells (as seen in **Figure 6.8**).

A somewhat opposite pattern could be deduced regarding *Smed-LPT*(RNAi) and *Smed-trr-2*(RNAi). The only similarity between them, in terms of histone modification changes at 3 days of regeneration, was the decrease in H3K27me3 in x1 cells compared to control. The relationship between these events and the observed cell dynamics could be as follows: relieved repression on stem cell maintenance genes could be leading to increase in stem cell number and decreased commited progeny production (**Figure 6.8** and **Figure 6.10**). A similar phenomenon was observed in the context of Glioblastoma multiforme tumour where decreased H3K27me3 was

associated with an invigorating effect on cancer stem cells, specifically via increased Wnt signaling (Rheinbay et al., 2013).

Finally, whole animal protein extract Western blots were performed on day 8 regenerates (**Figure 6.10C**). These experiments revealed a new pattern, different from day 5 homeostasis or day 3 regeneration. There did not appear to be a profound difference in H3K4me1 levels between experimental conditions and controls. However, there seemed to be opposing patterns in H3K4me3 and H3K27me3. Knockdown of trithorax-related genes was reflected in decreased H3K4me3 and increased H3K27me3 at this stage. It should be pointed out that, as described in **Chapter 5**, day 8 regeneration marks the onset of tumour formation in *Smed-LPT*(RNAi). Therefore, a potential explanation is that the histone modification changes at day 8 are reflective of a mixture of both primary (causative) and secondary (non-causative) lysine methylation changes associated with tumour formation. Increases in H3K27me3 and decreases in H3K4me3 were reported in cancer before (Jones and Baylin, 2007; Roesch et al., 2010) and are probably associated with decreased activation/increased repression of lineage commitment genes.

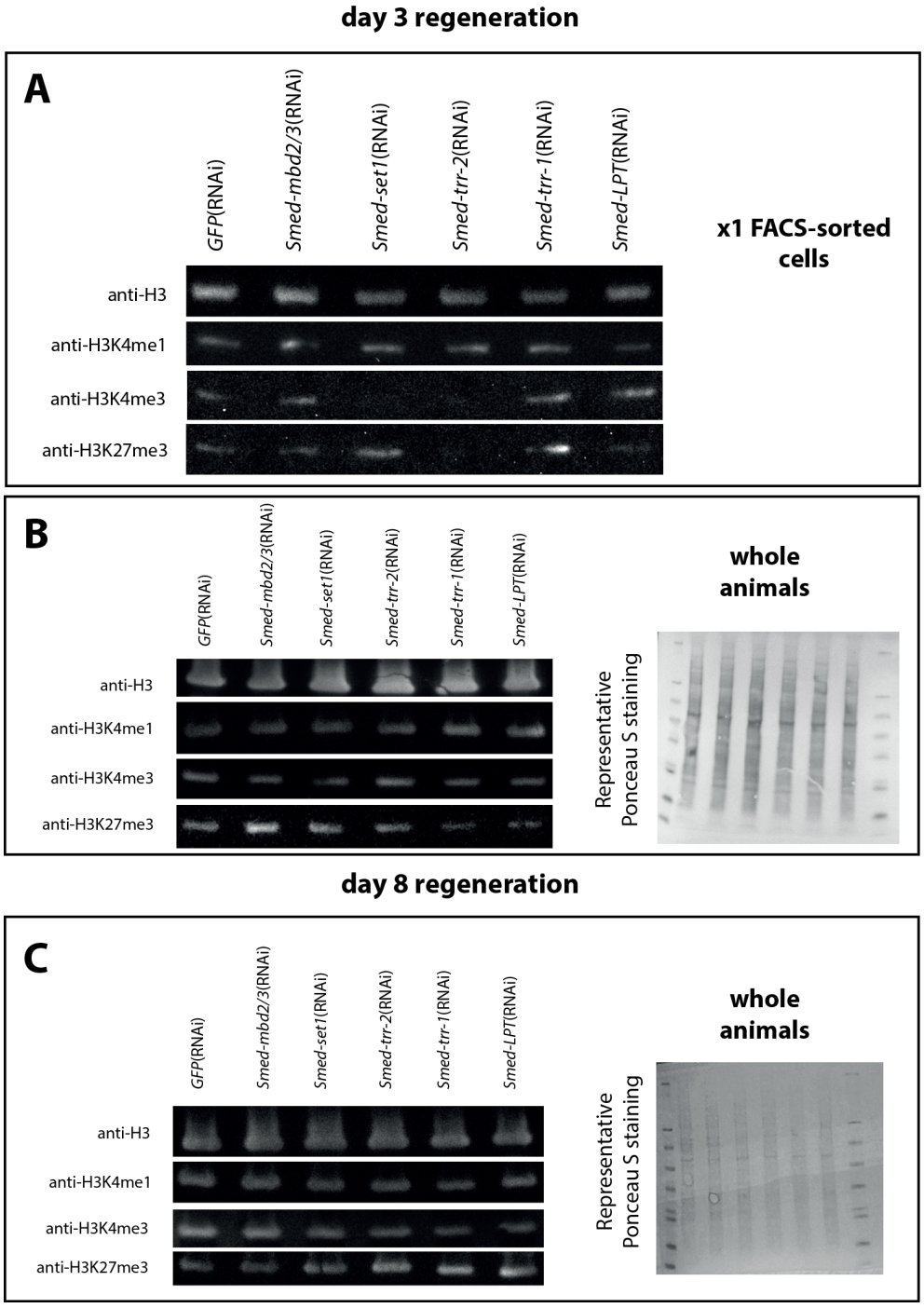


Figure 6.10. Histone modification changes at day 3 and day 8 regeneration knockdown phenotypes. (**A**) Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from x1 FACS-sorted cells from day 3 regenerates. There were two technical replicates of the H3K4me1 Western blot. (**B**) Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from whole animals from day 3 regenerates. An example Ponceau S-stained PVDF membrane prior to antibody incubation is shown. (**C**) Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from whole animals from day 8 regenerates. An example Ponceau S-stained PVDF membrane prior to antibody incubation is shown.

One conclusion that becomes clear from all of the presented cellular and histone modification dynamics studies is that animals with a trithorax-related gene knockdown seem to respond differently to homeostatic and regeneration conditions. The patterns of decreased H3K4me1 in *Smed-LPT*(RNAi) and increased H3K4me1 in *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) during homeostasis could be explained easily. Decrease in SMED-LPT levels leads to a decreased number of partner molecules for SMED-TRR-1 and SMED-TRR-2, hence less binding of the LPT-Trr complex to chromatin and less histone methylation. Based on previous findings (Herz et al., 2012), the reason why the effect was more obvious for H3K4me1 than H3K4me3 could be because the LPT-Trr complex is responsible mostly for H3K4 monomethylation and less so for trimethylation. The fact that x1 cells of *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi) samples showed an increase in H3K4me1 could potentially be explained by overcompensation on behalf of SMED-TRR-1 in *Smed-trr-2(*RNAi) and SMED-TRR-2 in *Smed-trr-1*(RNAi). Alternatively, upon sensing a disturbance in SMED-TRR-1's or SMED-TRR-2's function, other methyltransferases could be trying to compensate for this defect.

The increase in H3K27me3 following trithorax-related genes knockdown at 5 days homeostasis also seems logical if we assume that the H3K27me3 demethylating enzyme UTX fails to function properly if the core of the trithorax-related complex is perturbed (Herz et al., 2012). One of the few conclusions that could be drawn regarding the connection between histone modifications and changes in gene expression is that *Smed-AGAT-1*'s downregulation in *Smed-LPT*(RNAi) and *Smed-mbd2/3*(RNAi) animals was correlated with an increase in the repressive mark

H3K27me3 in both whole animals and stem cells specifically. This could be reflective of a greater pattern of increased repression of genes responsible for differentiation.

All of these observations, however, were reversed when we looked at early stages of regeneration (day 3). It appears that regeneration is a trigger for histone methylation compensatory mechanisms to turn on, in substitution of the perturbed trithorax-related complex. Potential examples of the result of such (over)compensatory mechanisms are the increased amount of H3K4me1 in whole protein extract from regenerating animals, as well as the decrease of H3K27me3 (possibly due to UTX being reactivated or another H3K27me3-demethylating mechanism being switched on). Further support for this interpretation is the observation that under *Smed-utx*(RNAi) conditions H3K27me3 was upregulated (as expected) in homeostatic conditions, but not during regeneration (**Figure 6.4C**). It is possible that such dire compensatory measures are a function of the extreme demands placed by the process of wound closure, healing and blastema formation. It is also plausible that these compensatory mechanisms are exhausted by the time late regeneration stages are reached (day 8), thus resulting in a less surprising pattern of H3K4me3 decrease and H3K27me3 increase following knockdown of one of the trithorax-related complex genes.

Interestingly, at no tested timepoint was H3K4me1 as dramatically reduced under trithorax-related gene knockdown as would be expected based on LPT-Trr's hypothesised primary function in H3K4 monomethylation. Two possible explanations of this phenomenon could be suggested. It is easily explained why that would be the case when one of the *trr* genes is knocked down – the other one compensates for

the loss and performs the missing gene's methylation properties. However, H3K4me1's lack of substantial down-regulation in x1 cells (or even increase in whole animals) following *Smed-LPT*(RNAi) during early regeneration, could have different causes. One explanation could be that in the absence of SMED-LPT, its partners - SMED-TRR-1 and SMED-TRR-2, bind to a different chromatin-docking protein and, hence, still perform some histone H3K4 monomethylation. Alternatively, in the absence of SMED-LPT, the two TRRs could bind chromatin on their own using their PHD-like zinc-binding domain. As discussed in **Chapter 2**, the PHD-like zinc-binding domain could have DNA-binding properties under some circumstances (Liu et al., 2014). Therefore, an interesting possibility is raised – the pathology following *Smed-LPT*(RNAi) could be not so much due to decrease in H3K4me1, but to changes in distribution pattern of H3K4me1. Conversely, the presented results leave open the possibility that the *Smed-LPT*(RNAi) phenotype is not related to H3K4me1 levels.

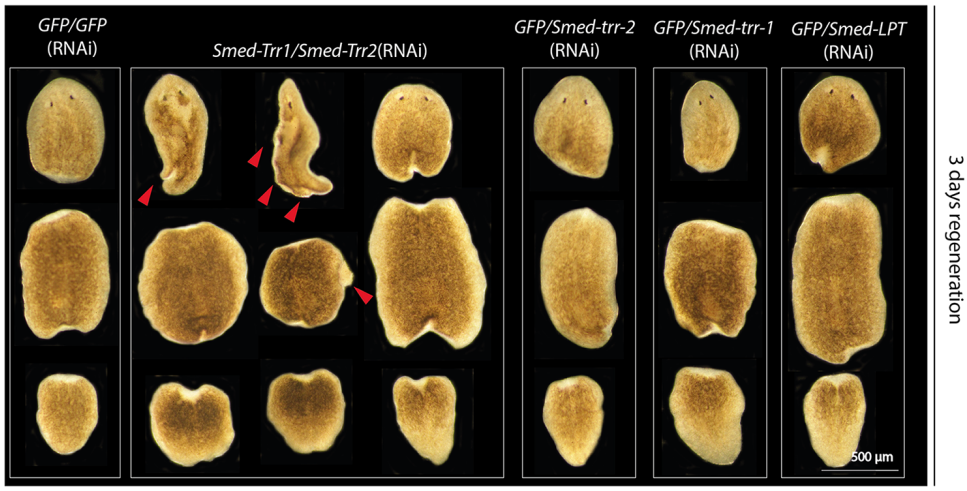
6.4.3. Functional redundancy between Smed-trr-1 and Smed-trr-2 masks trithorax-related proteins' involvement in H3K4me1

As previously shown in literature (Cheng et al., 2014; Herz et al., 2012, 2014; Hu et al., 2013), the LPT-Trr protein complex is heavily involved in monomethylation of H3K4. However, this has not been convincingly demonstrated by the Western blot experiments presented so far in this chapter. One of the main suspected causes for masking LPT-Trr's effect on H3K4 monomethylation is functional redundancy between *Smed-trr-1* and *Smed-trr-2,* previously reported by Hubert et al. (2013). A

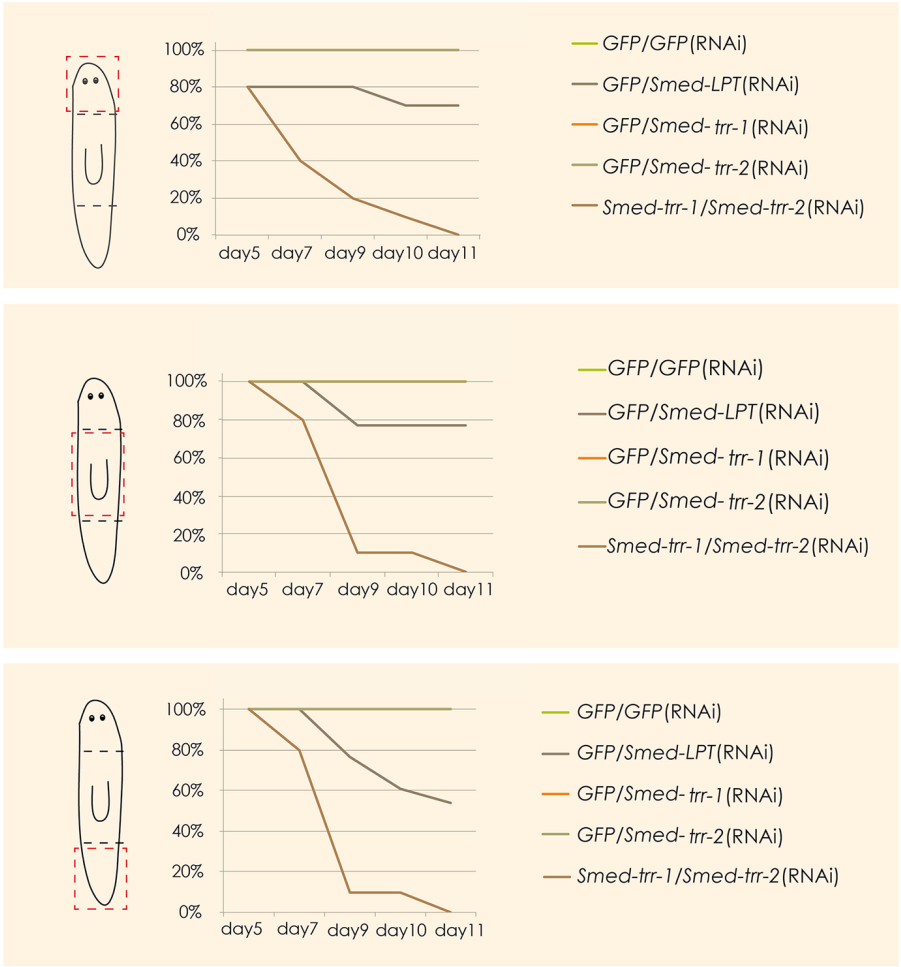
confirmation of these findings was provided by a double knockdown experiment. Regenerating *Smed-trr-1/Smed-trr-2*(RNAi) animals succumbed to similar defects like *Smed-LPT*(RNAi) animals. The progression and severity of the phenotype in the double knockdown was, however, dramatically increased. Three-day regenerating pieces had already started forming outgrowths in the double knockdown condition (**Figure 6.11A**) and dying commenced as early as 5 days following amputation (**Figure 6.11B**). Unlike *Smed-LPT*(RNAi) animals, 100% of all head, middle and tail *Smed-trr-1/Smed-trr-2*(RNAi) regenerating pieces developed outgrowths (**Figure 6.11C**) followed by death.

Hubert et al. (2013) observed diminished blastema formation in their double *trr* knockdown animals, similar to the present study, but failed to report the emergence of tumours. They also reported ventral curling prior to death, implying neoblast deficiency (Hubert et al., 2013). This was not observed in the current experiment. The discrepancy in results could be due to the different knockdown methods used in the two studies.

A



B



C

Regenerating piece	% *Smed-trr-1/Smed-trr-2*(RNAi) forming outgrowths
Head	100%
Middle	100%
Tail	100%

Figure 6.11. Double knockdown of *Smed-trr-1* and *Smed-trr-2* results in fast progressing outgrowth phenotype. (**A**) Brightfield images of 3 days regenerating head, middle and tail pieces following control knockdown (*GFP/GFP*) or a double experimental knockdown (*GFP/Smed-LPT, GFP/Smed-trr-1, GFP/Smed-trr-2, Smed-trr-1/Smed-trr-2*). Ten worms per condition were used. Red arrows point towards outgrowths. (**B**) Survival curves of head, middle and tail regenerating double knockdown pieces. The survival curves for *GFP/Smed-trr-1*(RNAi), *GFP/Smed-trr-2*(RNAi) and *GFP/GFP*(RNAi) fall on the same place (100%) in all three graphs. (**C**) Percentage of head, middle and tail double knockdown pieces developing outgrowths.

Double knockdown of *Smed-trr-1* and *Smed-trr-2* also resulted in a drastic decrease of H3K4me1 levels at day 3 regeneration (**Figure 6.12**), contrary to the results obtained for single knockdown with either of these genes (**Figure 6.10A** and **6.10B**).

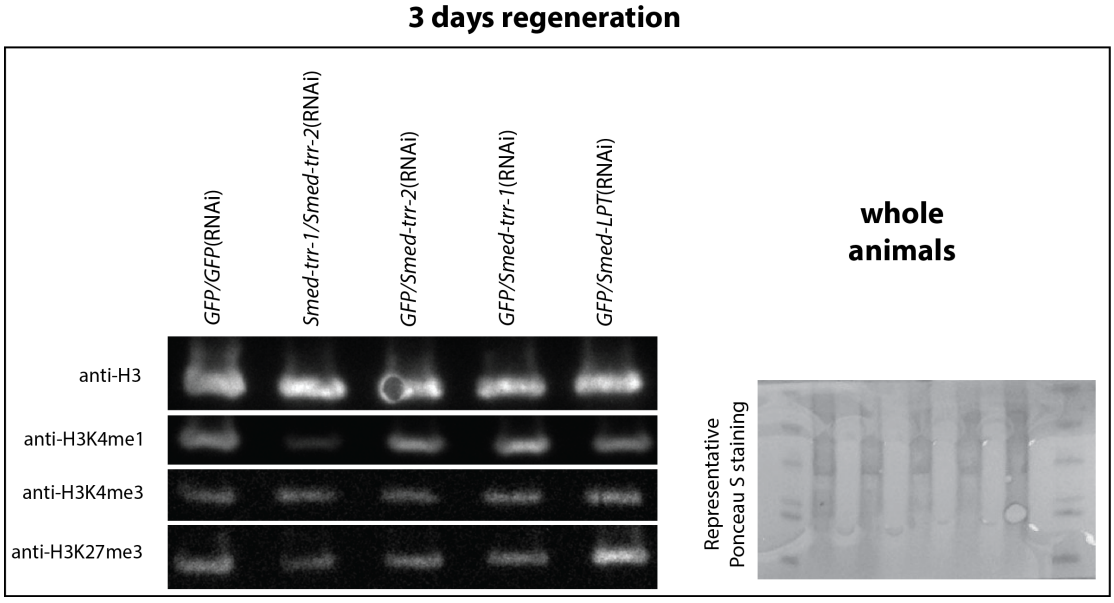


Figure 6.12. Double *trr* knockdown phenotype results in decreased H3K4me1 at 3 days of regeneration. Western blots with anti-H3 (unmodified), anti-H3K4me1, anti-H3K4me3 and anti-H3K27me3 antibodies on protein extract from whole regenerating animals at day 3 of regeneration. An example Ponceau S-stained PVDF membrane prior to antibody incubation is shown.

These results suggest that the LPT-Trr complex is involved mainly in H3K4 monomethylation and that TRR proteins are compensating for each other in single *trr* gene knockdown conditions.

6.5. CONCLUSION

The experiments in this chapter presented evidence that trithorax-related genes have an effect on H3K4me1, H3K4me3 and H3K27me3 under different regenerative and homeostatic conditions. Homeostasis revealed decrease of H3K4me1 and increase of H3K27me3 following *Smed-LPT* knockdown. *Smed-trr-1* and *Smed-trr-2* appeared to be overcompensating for each other's H3K4 monomethylation-associated abilities in G2/M stem cells. The compensation between these two genes was confirmed by the drastic down-regulation of H3K4me1 under *Smed-trr-1/Smed-trr-2* double knockdown conditions. Data on regenerating animals suggested that trithorax-related genes knockdown animals respond differently compared to homeostatic ones in terms of histone modifications. The changes in H3K4me1 were not as clear-cut and H3K27me3 was decreased, rather than increased. These observations suggest that under challenging circumstances, such as regeneration, alternative compensatory mechanisms could switch on. These mechanisms could involve other histone lysine demethylases and members of other COMPASS or COMPASS-like complexes.

At present, few specific conclusions could be drawn regarding the histone modification changes on particular genes influencing cellular dynamics under trithorax-related knockdown conditions. It can be speculated that loss of stem cells

at 3 days regeneration in *Smed-set1*(RNAi) animals could be caused by decreased H3K4me3-mediated activation of neoblast fate genes, thus pushing existing stem cells into differentiated fates. Similarly, it appears that the decrease in *Smed-AGAT-1*-postive cells in day 5 homeostasis *Smed-mbd2/3* and *Smed-LPT* knockdown worms is related to increased H3K27me3-mediated repression of epidermal fate genes.

This chapter presents evidence that *Smed-LPT, Smed-trr-1* and *Smed-trr-2* are indeed involved in histone lysine methylation changes. These findings will serve as foundation for future ChIP-seq experiments in order to gain a clearer understanding of trithorax-related genes' function.

7. Final Discussion

7.1. The ancestral Mll3/4 gene homologue underwent multiple gene fission events throughout evolution

Previous studies reported that the ancestral homologue of the vertebrate *MLL3/4* gene had undergone a fission event within arthropods in the orders of Diptera and Hymenoptera (Mohan et al., 2011). This event gave rise to two genes – *LPT* (corresponding to the N-terminus of *Mll3/4* homologue) and *trr* (homologous to the C-terminus of *Mll3/4*) in those orders. The present work experimentally confirmed the existence of a split *Mll3/4* homologue in planarians via finding the full sequence of a transcript orthologous to the N-terminus part of mammalian *Mll3* and *Mll4* genes. This transcript was named *Smed-LPT.* Hubert et al. (2013) had already reported the existence of two *trr* homologues named *Smed-trr-1* and *Smed-trr-2*.

Phylogenetic analysis (based on available genomic data) across phyla revealed a turbulent evolutionary history for the ancestral *Mll3/4* homologue. The analysis presented in this thesis expanded our knowledge of the evolution of trithorax-related genes. The current findings demonstrated that the *Mll3/4* gene homologue was un-split in the superphylum Radiata, but had likely undergone multiple fission events in Bilateria. No common bilaterian ancestor containing an original fission event was found. Instead, splitting of the *Mll3/4* homologue appeared to had happened independently in some deuterostome and protostome phyla. In deuterostomes, the gene was split in hemichordates and cephalochordates, but not

echinoderms and chordates. In protostomes, the fission event was identified as basal to all Lophotrochozoa. In Ecdysozoa, however, nematodes had one gene homologous to the vertebrate *Mll3/4*, while three independent fission events affecting the orders of Diptera (1) and Hypenoptera (2) (Mohan et al., 2011), as well as Amphipoda and Cladocera (3), but not Coleoptera and Lepidoptera had occurred.

The reason for the numerous fission events that had happened throughout the evolution of the *Mll3/4* homologue is unknown. The findings are surprising, considering the fact that gene fusion and fission events can usually be traced back to one common ancestor (Kummerfeld and Teichmann, 2005). One hypothesis could be that the fission event could be advantageous under certain evolutionary circumstances. The fission event might occur frequently since the two resultant products should be capable of reassembling as proteins (Kummerfeld and Teichmann, 2005; Mohan et al., 2011). The scenario of two proteins originating from the same gene working together is not novel for the MLL family. Trithorax, MLL1 and MLL2 proteins are known to undergo protein cleavage by Taspase 1 (Hsieh et al., 2003). However, the gene split in *Mll3/4* homologues and the protein cleavage in Trx/MLL1/2 happen at a different location. *Mll3/4* is split anterior to the PHD-like zinc-binding, FYRN, FYRC and SET domain group, while Trx/MLL1/2 is cleaved between the FYRN and FYRC domains.

A noteworthy finding was that the only animals showing a duplication (and even triplicationin the case of *P. torva*) of *trr* genes were Platyhelminth species. This observation suggests that the duplication happened within the Platyhelminth phylum.

One obvious limitation of the phylogenetic experiments described in this thesis is that all of the data used (apart from *Smed-LPT* sequence) was not experimentally verified via RACE (Rapid Amplification of cDNA Ends). As such, it is possible that some predicted split events were actually false positives due to anomalous genome assembly and/or protein predictions.

Mohan et al. (2011) hypothesised that *Drosophila* LPT and Trr proteins work together. It was proposed that LPT binds chromatin and thus directs the catalytic activity of its complex co-habitants – the H3K4 histone lysine methyltransferase Trr and the H3K27 histone lysine demethylase UTX.

Similar to its *Drosophila* orthologue, SMED-LPT was found to contain PHD-fingers known to be important for chromatin binding. Amino acid residue analysis revealed that the gene's PHD-fingers should have affinity for binging non-methylated or monomethylated lysine residue 4 on histone 3 (H3K4). Analysis of the SET domain amino acid composition of SMED-TRR-1 and SMED-TRR-2 revealed that this domain should have a preference for catalyzing monomethylation of H3K4 rather than di- or trimethylation, based on the substitution of phenylalanine with tyrosine at the C-terminus of the domain. The same substitution is also present in *Drosophila* Trr's SET domain. The amino acid residue analysis on PHD and SET domains of SMED-LPT, SMED-TRR-1 and SMED-TRR-2 proteins is in agreement with research placing greater emphasis on LPT-Trr performing the bulk of H3K4 monomethylation in *Drosophila*, while MLL3 and MLL4 have similar functions in mammalian cells (Herz et al., 2012; Hu et al., 2013). The amino acid analysis presented in this thesis does not necessarily refute SMED-LPT-TRR's involvement in H3K4 promoter trimethylation. It is possible

that under changing developmental and/or environmental circumstances, the protein complex adopts alterations in its conformation and is, in fact, capable of performing limited amounts of H3K4me3. This seems plausible, considering the fact that LPT-Trr/MLL3/4 was found to be involved in hormone-dependent transcription, associated with different developmental stages such as *Drosophila* moulting (Chauhan et al., 2012; Sedkov et al., 2003) and circadian rhythms in mammals (Kim et al., 2015; Valekunja et al., 2013).

An intriguing feature of the amino acid analysis presented in this thesis was the PHD-like zinc-binding domain found at the N-terminus of SMED-TRR-1 and SMED-TRR-2. It was demonstrated that SMED-TRR-1 had all characteristic for this domain HC5HC2H residues, while SMED-TRR-2 had a substitution of the final C. It was previously reported that PHD-like zinc-binding domains could have some dsDNA-binding properties (Liu et al., 2014). Hence, it is possible that SMED-TRR-1 (and possibly SMED-TRR-2) has some limited DNA (or histone tail) binding capabilities independently of SMED-LPT. This interpretation is supported by data showing that LPT and Trr (which also contains PHD-like zinc-binding domain) protein binding sites on the DNA do not overlap completely in *Drosophila* (Herz et al., 2012). Western blot data in **Chapter 6** showing presence of H3K4me1 in *Smed-LPT*(RNAi) animals, but absence of H3K4me1 in *Smed-trr-1/Smed-trr-2*(RNAi) animals could also point towards the same interpretation.

Another characteristic that makes SMED-TRR-1 different from SMED-TRR-2 is that SMED-TRR1 was found to be the only one out of the three identified planarian trithorax-related proteins to contain a nuclear receptor (NR) binding motif. In

Drosophila, both LPT and Trr have several such motifs associated with performing hormone-regulated transcriptional activation (Chauhan et al., 2012; Sedkov et al., 2003).

The findings that SMED-TRR-1 has an NR motif and a better preserved PHD-like zinc-binding domain than SMED-TRR-2, could indicate that the two proteins have some level of functional divergence, despite evidence of partial redundancy described in this thesis and by Hubert et al., (2013). It is possible that the two proteins have shared functions. However, hormone-related activities are possibly relegated to SMED-TRR-1, which could perform lysine methylation either via binding to SMED-LPT first, or independently.

7.2. PLANARIAN TRITHORAX-RELATED GENES ARE EXPRESSED IN STEM CELLS AND STEM CELL PROGENY

A combination of *Smed-H2B*(RNAi), gamma irradiation and *Smed-H2B* co-localisation studies demonstrated that trithorax-related genes in *S. mediterranea* were expressed in neoblasts, as well as neoblast progeny cells. *Smed-trr2* and *Smed-LPT* were also expressed in the differentiated gut (*Smed-trr-2* more so than *Smed-LPT*) and CNS, while *Smed-trr-1* was absent from the gut. All three genes were also expressed during regeneration, most prominently in recovering brain and pharynx. Similar results were previously demonstrated for *Smed-trr-1* and *Smed-trr-2* (Hubert et al., 2013). They also fit well with findings reported by Brun et al. (2006) that mouse *mll3* was strongly expressed in the CNS during embryonic development.

This incomplete similarity in gene expression patterns further supports the possibility that the three trithorax-related genes have developed some degree of independent functional specialisation. The hypothesis that split gene parts of the *Mll3/4* ancestral homologue could acquire novel functions is also supported by the fact that some Lophotrochozoa (the mollusk *C. gigas* and Platyhelminth *C. sinensis*) had obtained novel domains (cadherin, GH9 and Tyrosinase) in their LPT genes (**Chapter 2**).

Nevertheless, the three trithorax-related genes co-localised with each other, suggesting that, even if some independent functions were acquired, their combined function is still likely to be present. These findings support Mohan et al.'s (2011) discovery that *Drosophila* LPT co-immunoprecipitates with Trr protein suggesting that they work together. Undoubtedly, it would be useful to see if trithorax-related proteins also co-localise in planarians.

The fact that planarian trithorax-related genes were found to be present in over 90% of neoblasts suggests that they have a crucial role in stem cell regulation. This hypothesis was supported by the results presented in **Chapters 4** and **5**. The genes' role in stem cell regulation is further validated by studies in zebrafish reporting increased expression of *mll3a, mll4a* and *mll4b* in the embryonic blastoderm (Sun et al., 2008). An exciting future experiment would be to investigate the distribution of trithorax-related genes' expression across different known subpopulations of neoblasts like sigma ('true' pluripotent neoblasts) and zeta (neoblasts committed to the epidermal lineage) (van Wolfswinkel et al., 2014).

7.3. TRITHORAX-RELATED GENES CONTROL STEM CELL DIFFERENTIATION OF SELECTED LINEAGES

Functional studies presented in this thesis confirmed that trithorax-related genes are involved in stem cell regulation. More specifically, they proved to be responsible for proper stem cell differentiation down certain lineages. Stem cell maintenance was not deteriorated following knockdown of trithorax-related genes, hence the differentiation defects were not due to a depleting pool of stem cells not producing enough progeny cells.

All described differentiation phenotypes were strongest during regeneration and not homeostasis. This is not surprising, considering that recovering missing tissues is a more stressful and demanding process than the process of homeostatically maintaining these tissues. For all differentiation defects uncovered, the malformations were most prominent in *Smed-LPT*(RNAi) animals, often absent altogether in *Smed-trr-1*(RNAi) and/or *Smed-trr-2*(RNAi). These findings point towards functional redundancy between *Smed-trr-1* and *Smed-trr-2* reported previously by Hubert et al. (2013) and also confirmed in this thesis (**Chapter 6**).

The specific differentiation defects observed heavily affected the regeneration of anterior structures. Optic cup regeneration was perturbed under both *Smed-LPT* and *Smed-trr-1* knockdown conditions. Problems with eye formation in association with *LPT/Trr/Mll3/4* depletion were reported in *Drosophila*, mouse and planarians (Hubert et al., 2013; Kanda et al., 2013; Lee et al., 2006a; Sedkov et al., 2003). Furthermore, the total number of every terminally differentiated neuronal subtype examined (acetylcholinergic, GABAergic, serotonergic, dopaminergic) was decreased

following *Smed-LPT*(RNAi), as well as the overall size of the regenerating brain. A notable discovery was that *Smed-LPT*(RNAi) regenerating animals had an excess of *Smed-pitx*-positive serotonergic neuronal precursor cells. This suggests that in the absence of *Smed-LPT*, precursors of serotonergic neurons cannot complete their differentiation process and, as a consequence, accumulate. The irregularities observed in the recovery of the serotonergic system under *Smed-LPT* knockdown also led to motility difficulties, alleviated via treatment with ectopic serotonin.

The observed anterior regeneration problems following *Smed-LPT*(RNAi) could be explained by the fact that planarian trithorax-related genes were all expressed in the regenerating CNS, possibly regulating the differentiation of stem cells down different neuronal lineages.

Other differentiated structures affected by knockdown of trithorax-related genes and, more specifically, *Smed-LPT*, included the pharynx, the gut and the epidermis (early and late epidermal precursors). However, the regeneration and homeostatic maintenance of other structures, like protonephridia and dorsal and ventral cilia, was unaffected. These findings suggest that the trithorax-related complex in planarians is responsible for the differentiation of some tissues (brain, gut, pharynx, eyes, epidermis), but not others (cilia, protonephridia).

The tissue differentiation experiments described in this thesis provided an answer mostly to the question of whether the terminal differentiation of a structure was affected after disturbing the function of trihtorax-related genes. In the case of serotonergic neurons, the addition of the *Smed-pitx* experiment led to a better understanding of the regulation of the serotnergic system by the COMPASS-like

complex. It would be useful to gain similar knowledge for other lineages as well. One example would be to explore the cells positive for the genes *Smed-hnf-4* and/or *Smed-gata4/5/6* known to be expressed in the progenitors of the planarian gut (Scimone et al., 2014) or *Smed-FoxA* (Adler et al., 2014), expressed in pharyngeal progenitors. Such experiments would allow us to better understand, in the context of different tissues, at which point of cell commitment the trithorax-related complex is involved.

The differentiation defects ensuing after knockdown of trithorax-related genes could be related to disturbance of the genes' nuclear receptor co-activator functions associated with selective H3K4me3 on hormone-induced promoters (Chauhan et al., 2012; Sedkov et al., 2003). However, they could equally well be due to disturbances in H3K4me1 deposition on enhancers leading to altered balance between active and inactive enhancer regions (Herz et al., 2012; Hu et al., 2013). A third scenario is also possible where knockdown of trithorax-related genes leads to reduction of H3K4me1 on some gene promoters, thus leading to derepression of those genes (Cheng et al., 2014). It will be difficult to distinguish between these three scenarios and to what extent they contribute to the phenotype, until ChIP seq data with histone modification antibodies (H3K4me1, H3K4me3 and H3K27me3) is available.

7.4. STEM CELL OVERPROLIFERATION RESULTS IN OUTGROWTH FORMATION IN *SMED-LPT*(RNAI) ANIMALS

The interest of using planarians as model organisms to understand cancer pathology originated in the second half of the twentieth century. A number of studies exposed regenerating and homeostatic planarians to a variety of teratogens and carcinogens. Data on the effects of these compounds on planarians was in strong agreement with mammalian studies (Best and Morita, 1982; Schaeffer, 1993), thus establishing planarians as a viable model for understanding tumourigenesis. It was proposed that the neoblast overproliferation resultant from carcinogen exposure was reminiscent of the processes leading to embryonic stem cell teratoma and other mammalian tumours (Best and Morita, 1982; Foster, 1963, 1969; Schaeffer, 1993). More recently, using planarians as a model to understand basic mechanisms relevant to cancer biology has entered the genetics era and some valuable findings have already been made available to the cancer community. The most notable example is *Smed-smg-1*, identified for the first time as a tumour suppressor in planarians (González-Estévez et al., 2012) and shown to have conserved functions in mammals (Han et al., 2014; Roberts et al., 2013).

The greatest appeal of using planarians to unravel cancer mechanisms is related to two facts. The first one is that this animal undergoes a high amount of tissue turnover throughout its lifetime, associated with an incredible amount of stem cell divisions and, yet, somehow, it remains robust and protected from cancer. Therefore, mutations or treatments that lead to 'tricking' this animal into developing cancer are extremely valuable to help us understand why this worm is so resilient to

malignancy. Moreover, the mammalian hematopoietic and gastrointestinal systems are among the tissues with the highest turnover rates in mammals. Therefore, their stem cells could share similar regulatory mechanisms with planarian neoblasts (Oviedo et al., 2008). Flatworms provide the benefit of an *in vivo* model system to study the cancer regulatory mechanisms of stem cells in a highly proliferative context. Therefore, there is great value in exploring the mechanisms by which one of the most frequently mutated genes in all cancers, *Mll3,* is causing cancer in planarians.

Apart from compromised stem cell differentiation abilities, which is one of the characteristics of cancer, *Smed-LPT*(RNAi) planarians also showed high prevalence of outgrowth formation. As with differentiation defects, the outgrowth phenotype was much stronger in regenerates. It was demonstrated that increased levels of stem cell proliferation preceded and accompanied the development of tumours. There is a discrepancy between the findings on stem cell numbers (*Smed-H2B*-positive cells) reported in **Chapter 4** and those on mitotic cell numbers (anti-H3P-positive cells) cited in **Chapter 5**. At day 8 of regeneration the number of *Smed-H2B*-positive cells was not significantly different between *Smed-LPT*(RNAi) and controls. At the same regeneration timepoint, there were significantly more mitotic cells in *Smed-LPT*(RNAi) animals than in controls. Since the notion of dividing differentiated cells was excluded in **Chapter 5**, the most likely explanation for this finding is that dividing stem cells at day 8 of regeneration were producing more progenitor cells belonging to specific lineages. We know this was not the case for the epidermal lineage (labeled by *Smed-NB.21.11e* and *Smed-AGAT-1*), but production of other progenitor cells, like *Smed-pitx*-positive serotonergic neuronal precursors, was increased.

Moreover, the average number of early stem cell progeny cells (SMEDWI-1-positive/*Smed-H2B*-negative) was also elevated (although not significantly) in *Smed-LPT*(RNAi) compared to control.

Stem cells were shown to be present in *Smed-LPT*(RNAi) animals' outgrowths. Such neoblasts could potentially bear similarities with cancer stem cells and it would be of extreme interest to study their properties further. A technically challenging experiment aiding their characterisation would be FACS-sorting stem cells from *Smed-LPT*(RNAi) tumours and performing RNA-seq and/or ChIP-seq with different histone modifications. The control for this experiment would be sorting neoblast from a matched *GFP*(RNAi) control. An even further stretch would be to do single cell sequencing on these cells. Since the experiments presented in this thesis demonstrated that some stem cells after *Smed-LPT*(RNAi) had a misshapen nucleus (a common feature of cancerous cells), it would be intriguing to see gene expression differences between these cells and cells with unchanged nuclear shape.

A cancer hallmark that would be relatively easy to study in planarians is resistance to cell death. The TUNEL technique (terminal transferase-mediated dUTP Nick end labelling) as a method of labelling apoptotoc cells had been successfully used in planarians before (Pellettieri et al., 2010). It would be important to see if *Smed-LPT*(RNAi) animals show resistance to apoptosis and, if so, how this resistance is temporally related to tumour formation.

Another feature associated with tumour development and also seen in *Smed-LPT*(RNAi) animals is epithelial disorganisation. Similar to the *Smed-smg-1*(RNAi) cancer phenotype (González-Estévez et al., 2012), signs of epidermal hyperplasia

were observed, as well as disturbance of the normal interaction between the epithelium and the mesenchyme. Changes in epithelium-mesenchyme interaction are often associated with the onset of malignancy and metastasis (Nelson and Bissell, 2006; Partanen et al., 2009). A candidate proxy measurement of metastatic capacity, the cell migration assay using partial irradiation, confirmed that stem cells in *Smed-LPT*(RNAi) animals had enhanced migration capacity. Unpublished data in the Aboobaker lab, obtained by Ellen Aboukhatwa, showed that stem cells behaved in a similar way in *Smed-smg-1*(RNAi) planarians. It would be exciting to explore if the enhanced migratory phenotype in *Smed-LPT*(RNAi) animals is due to epigenetically mediated suppression of *Smed-smg-1.* Another potential avenue for learning more about the cell migratory behaviour following *Smed-LPT* knockdown is to examine the behaviour of other cell types using the partial irradiation assay. One good candidate are *Smed-pitx*-positive serotonergic neuronal precursors, which, like dividing stem cells, are more abundant in *Smed-LPT*(RNAi) animals. Furthermore, it would be informative to look at the migratory capabilities of sigma and zeta populations of stem cells. It could be the case that the increased stem cell migration (as marked by *Smed-H2B*) observed in the experiments in this thesis is underpinned by disproportionate response on behalf of sigma or zeta stem cells.

A surprising feature of the *Smed-LPT*(RNAi) phenotype was the renewed production of *Smed-NB.21.11e* and *Smed-AGAT-1* transcripts in tumours, having previously severely down-regulated their transcription in pre-tumourous stages. Since the two transcripts are used to label epidermal progenitors, it is possible that their reinstated presence in the tumour region was reflective of the observed epithelial disorganisation. An equally plausible explanation is that the genes' expression was

related to changing cellular energetics in cancerous conditions. In order to learn more about the energetic and metabolic changes happening within the *Smed-LPT*(RNAi) tumours, metabolomics profiling could be performed. This could be done on the whole tumour tissue as well as on different FACS-sorted cell populations. Such experiment would also shed light on whether hormone regulation is perturbed (as would be expected from LPT-Trr/MLL3/4 studies in *Drosophila* and mammals) under trithorax-related knockdown conditions and whether it is associated with tumour formation.

7.5. TRITHORAX-RELATED GENES AND PLANARIAN CELL DYNAMICS

Experiments presented in **Chapters 4**, **5** and **6** allowed us to deduce information regarding the role of trithorax-related genes in stem cell maintenance and differentiation. It was demonstrated that neoblasts, as labeled by *Smed-H2B*, showed increased numbers following *Smed-LPT*(RNAi) at day 3 regeneration. Labeling of proliferative neoblasts (anti-H3P) also showed that *Smed-LPT*(RNAi) animals had significantly elevated numbers of dividing stem cells at 12 hours, 2 days and 8 days following amputation. Similar over-proliferative effects were also seen for *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi). These experiments led to the conclusion that trithorax-related genes control the self-renewal capacity of neoblasts. Studies in planarian biology have shown the existence of two types of neoblasts – cNeoblasts (also known as the sigma population; true pluripotent stem cells) (van Wolfswinkel et al., 2014; Wagner et al., 2011, 2012) and neoblasts committed to different lineages (Adler et al., 2014; Cowles et al., 2013; Currie and Pearson, 2013; Lapan and

Reddien, 2011; März et al., 2013; Rink et al., 2011; Scimone et al., 2011; van Wolfswinkel et al., 2014). For at least one type of committed neoblasts – the epidermal precursors (zeta population), it was shown that they still possess limited self-renewal capability (van Wolfswinkel et al., 2014). The experiments described in this thesis do not allow a distinction between committed and non-commited neoblasts. Hence, whether trithorax-related genes regulate committed or non-committed stem cells' self-renewal is currently unknown.

It is established that *Smed-SP6-9* is expressed in optic cup precursor neoblast and neoblast progeny cells (Lapan and Reddien, 2011). *Smed-SP6-9's* down-regulation after *Smed-LPT*(RNAi) and *Smed-trr-1*(RNAi) could indicate that trithorax-related genes are involved in the commitment of cNeoblasts to optic cup fate. However, it could also mean that trithorax-related genes are responsible for the commitment of optic cup progenitors further down the line. In order to resolve this issue, a double *in situ* hybridisation with *Smed-SP6-9* and *Smed-H2B* probes could be performed on knockdown regenerates.

Similarly, *Smed-pitx* is expressed in neoblasts and neoblast progeny cells destined to become serotonergic neurons (Currie and Pearson, 2013; März et al., 2013). At this point, a distinction between trithorax-related genes' involvement in stem cell commitment or stem cell progeny differentiation cannot be made.

This thesis has demonstrated that early and late epidermal progenitors (*Smed-NB.21.11e*–positive and *Smed-AGAT-1*–positive cells respectively) are down-regulated following *Smed-LPT*(RNAi). However, it is not known whether the zeta neoblast population (epidermal precursors) is also affected. Therefore, the precise

time at which trithorax-related genes exert control on epidermal differentiation cannot be stated with certainty.

Proof of trithorax-related genes' involvement in terminal differentiation of structures was provided. This was achieved via the use of differentiated brain markers (*Smed-GAD, Smed-TPH*, *Smed-TH*, *Smed-chat*), as well as markers for gut (*Smed-porcupine-1*) and pharynx (*Smed-laminin*). Echoing the previously made points, we cannot currently state at which point exactly stem cell commitment and differentiation is perturbed following trithorax-related genes knockdown. Double *in situ* hybridisation with *Smed-H2B* and markers for neoblasts commited to the neuronal lineage, such as *Smed-sim* and *Smed-coe* (Cowles et al., 2013) would be one example of an experiment to further advance our conclusions.

A summary figure of the neoblast self-renewal, lineage commitment and differentiation findings, presented in this thesis, is shown below (**Figure 7.1**).

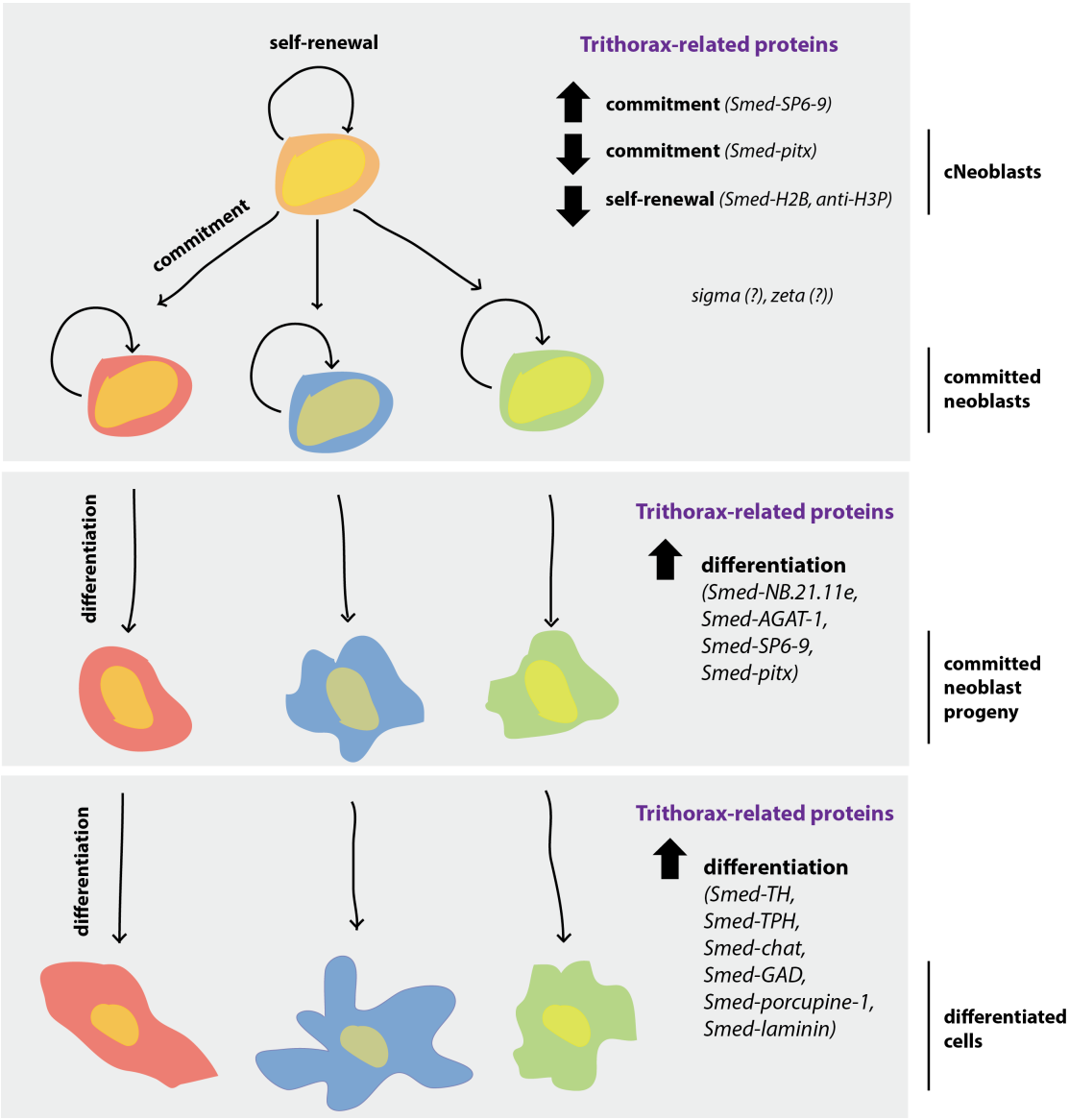


Figure 7.1. Trithorax-related genes' involvement in neoblast self-renewal control, commitment and progenitor differentiation. Circular lines represent self-renewal capacity, while straight lines indicate neoblast commitment or neoblast progeny differentiation. The differently coloured cells do not belong to specific lineages, but illustrate a concept. The thick black arrows represent trithorax-related proteins' involvement in activation (upward-facing thick arrow) of a specific process, as illustrated by an experiment with an *in situ* hybridisation marker or an antibody (all in italics), or suppression (downward-facing thick arrow) of a process.

7.6. TRITHORAX-RELATED GENES CONTROL H3K4ME1, H3K4ME3 AND H3K27ME3

This thesis confirmed previous findings (Chauhan et al., 2012; Cheng et al., 2014; Herz et al., 2012; Hu et al., 2013; Mohan et al., 2011; Sedkov et al., 2003) that trithorax-related proteins are associated with modifications on H3K4 and on H3K27 (via UTX). **Chapter 6** also demonstrated that the trithorax-related cell regulatory mechanisms summarised in **Figure 7.1** are underpinned by changes in H3K4me1, H3K4me3 and H3K27me3 on both stem cell level (x1 FACS-sorted cells) and on whole organismal level. We observed that during homeostasis, H3K4me1 was down-regulated in the whole organism following *Smed-LPT*(RNAi), *Smed-trr-1*(RNAi) and *Smed-trr-2*(RNAi), while H3K27me3 was up-regulated. This trend was reversed during early regeneration (day 3), possibly indicating some compensatory (or overcompensatory) mechanisms that get switched on in trithorax-related-deficient regenerative circumstances. Western blot studies also showed that increases in H3K27me3 and decreases in H3K4me3 in the whole organism during late regeneration are characteristic of tumourous phenotype onset in *Smed-LPT*(RNAi). These findings demonstrate that highly dynamic changes in histone modifications accompany shifts in cellular dynamics and organismal conditions. One way to solidify the presented conclusions would be to introduce more biological replicates. However, these experiments provided evidence that the antibodies used were specific, as demonstrated by H3K4me3 down-regulation following *Smed-set1*(RNAi) and up-regulation of H3K27me3 after *Smed-utx*(RNAi). Hence, the results shown in **Chapter 6** serve as a stable foundation for future ChIP sequencing experiments.

Intriguingly, experiments described in **Chapter 6** demonstrated that double knockdown of *Smed-trr-1* and *Smed-trr-2* resulted in a quicker, more severe phenotypic version of *Smed-LPT*(RNAi). Tumours appeared in all regenerating animals (starting as early as day 3 of regeneration) leading to their death within 11 days. Plausible explanations for these observation are discussed below.

First, the *Smed-LPT* knockdown could have been less efficient than the double knockdown of *Smed-trr1* and *Smed-trr2*. This could be tested with a qPCR experiment measuring transcript levels after knockdown. If such experiment were to confirm this speculation, this would mean that after *Smed-LPT*(RNAi), there was still a substantial amount of LPT protein left for TRR-1 and TRR-2 to bind to and methylate histones. According to this interpretation, the tumourous phenotype and the speed of its progression would be explained by varying amounts of correctly deposited H3K4 methylation.

Alternatively, the difference in phenotypic observations between *Smed-LPT*(RNAi) and *Smed-trr-1/Smed-trr2*(RNAi) could not be heavily reliant on difference in residual transcript and protein levels of the knockdown genes. Instead, what could be happening is that in the absence if LPT, TRR-1 and TRR-2 had some capabilities of binding chromatin on their own and performing histone methylation that way. Another possibility is that in the absence of LPT, they bound other chromatin-binding proteins to serve as their anchor to chromatin and to allow them to perform histone methylation. If TRR-1 and TRR-2 do indeed have some LPT-independent histone lysine methylating capabilities, the tumours in *Smed-LPT*(RNAi) could be explained by abberant H3K4 methylation patterns rather than an overall decrease in H3K4

methylation. On the other hand, a double knockdown of *Smed-trr-1* and *Smed-trr-2* and the severe phenotype that followed could be related to a drastic decrease in H3K4 methylation (specifically monomethylation).

One way to reveal what is happening on a chromatin level under *Smed-LPT*(RNAi) conditions would be to perform ChIP sequencing experiments with antibodies against histone modifications. This approach should resolve whether the cancerous phenotype of *Smed-LPT*(RNAi) animals is due to less H3K4 methylation or to abberantly distributed H3K4 methylation.

7.7. FUTURE DIRECTIONS

The most significant of the experiments planned to help deepen our knowledge of trithorax-related genes' cancer-related function is ChIP-seq with antibodies against H3K4me1, H3K4me3 and H3K27me3. Performing the experiment on whole animals would not greatly advance our knowledge, due to the noise produced by different modifications in various cell types. Therefore, x1 (G2/M neoblasts) FACS-sorted cells are to be used since this is the cleanest, best-defined population of stem cells available with current techniques. Using x1 would allow us to observe the histone modification changes happening in stem cells and associated with tumourigenesis.

Two phenotypes presented in this thesis gave rise to tumourous outgrowths – *Smed-LPT*(RNAi) and *Smed-trr-1/Smed-trr-2*(RNAi). Since we are interested primarily in the initial (causative) changes leading to outgrowth formation, day 3 regenerating *Smed-LPT*(RNAi) animals (when no phenotypic changes are visible in brightfield) are to be

used. This timepoint is chosen because none of the available studies on LPT-Trr/MLL3/4's cancer-inducing activity have focused on the early events leading to malignancy. Most of the studies observe *Mll3/4* mutations in conditions where the malignant tumour is already a fact. Therefore, it is difficult to pinpoint the primary changes in the noisy landscape of heavily mutated tumours.

Smed-LPT(RNAi) is preferred to double *trr* knockdown phenotype for two reasons. First, the phenotypic development in the double *trr* knockdown is extremely quick (tumours appear as early as day 3 of regeneration). Hence, this experimental condition might not be suitable for temporally separating causative from secondary cancer-related changes. Second, based on the Western blot experiments presented in **Chapter 6**, the outgrowth formation after double *trr* knockdown is associated with gross depletion of H3K4me1, while tumour formation after *Smed-LPT*(RNAi) is not. Hence, it would be intriguing to find out whether *Smed-LPT*(RNAi) tumours are the result of aberrantly distributed rather than decreased H3K4me1. If so, this would be evidence for epigenetic plasticity (as a result of either compensatory activation of other H3K4 methyltransferases, like SMED-MLL1/2 and SMED-SET1 or SMED-TRR-1/TRR-2-mediated rogue H3K4 methylation) leading to tumour formation.

No previous efforts have been made to demarcate enhancer regions in the planarian genome. This endeavor would be extremely difficult without chromatin conformation capture (3C) data. Therefore, it was decided that the planned ChIP-seq experiment is to focus on promoter region regulation in *Smed-LPT*(RNAi) conditions. Based on literature (Cheng et al., 2014), we expect to see promoter regions of the three types illustrated in **Figure 7.2**.

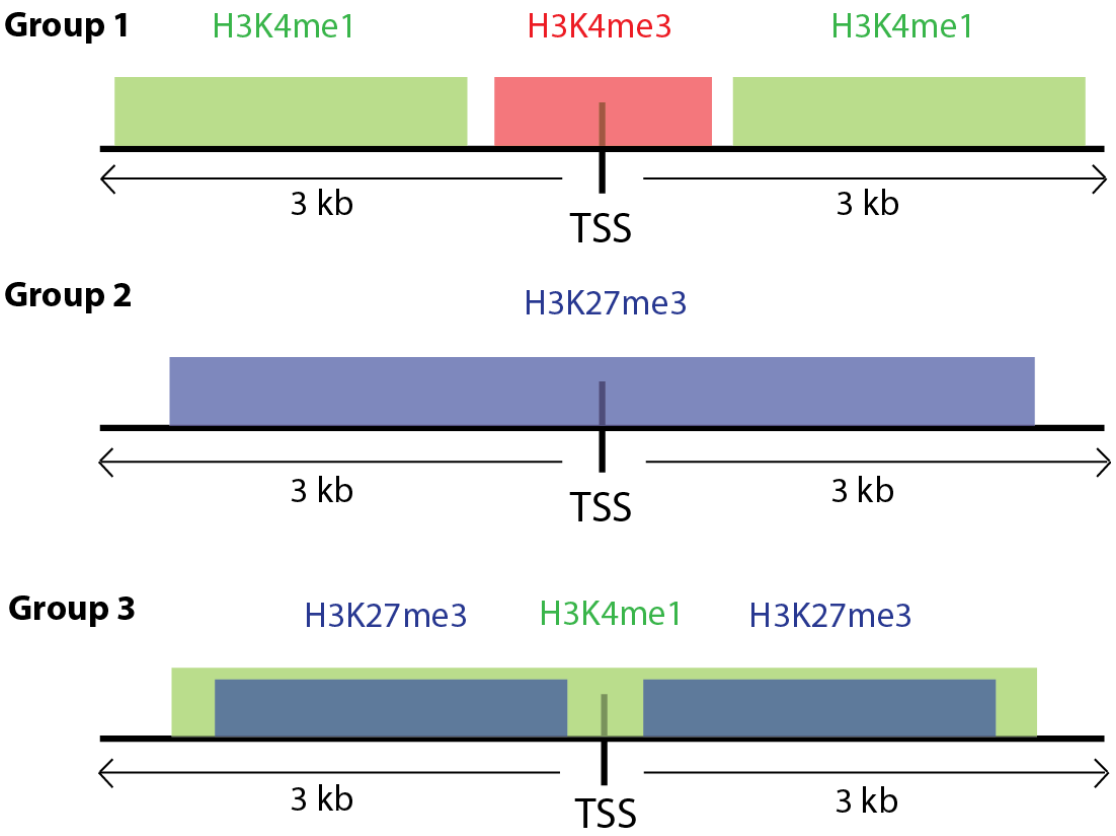


Figure 7.2. Three groups of gene promoter types. Group 1 is characterised by H3K4me3 at the TSS and neighbouring H3K4me1 and is seen in active promoters. Group 2 promoters are characteristic of constitutive chromatin with enrichment of H3K27me3. Group 3 is seen in facultative chromatin promoters with H3K27me3 on the two sides of the TSS and H3K4me1 spanning the whole promoter (including the TSS). The graph is based on data by Cheng et al. (2014).

After obtaining 2 biological replicates for the described ChIP-seq experiment, the mRNA levels of a small group of genes whose promoters are affected after *Smed-LPT*(RNAi) are going to be tested via qPCR. Such experiments would serve the purpose of validating the ChIP-seq results.

Provided that there was no shortage of time or resources, 3C-sequencing methods would be used in order to define enhancers in the planarian genome. Afterwards, the most comprehensive study of trithorax-related genes' role in planarian carcinogenesis would involve the following:

	Experiment 1	**Experiment 2**	**Experiment 3**
Experimental knockdown conditions	day 3 regeneration *Smed-LPT*(RNAi) versus *GFP*(RNAi) animals	day 8 regeneration *Smed-LPT*(RNAi) versus *GFP*(RNAi) animals	day 3 regeneration *Smed-trr-1/Smed-trr-2*(RNAi) versus *GFP/GFP*(RNAi) animals
ChIP-seq with antibodies against H3K4me1, H3K4me3 and H3K27me3 on FACS-sorted cells	x1, x2, x ins FACS-sorted cell populations	x1, x2, x ins FACS-sorted cell populations	x1, x2, x ins FACS-sorted cell populations

Table 7.1. The ideal future experiment.

Such experiments would allow us to observe the histone modification changes across different cell populations pre-tumourigenesis and at the onset of tumour formation after *Smed-LPT*(RNAi). Furthermore, the experiments will conclusively show whether there is a difference between the histone modification landscape of *Smed-LPT*(RNAi) and double *trr* knockdown animals. Because of the 3C-based sequencing, histone modifications on enhancers would also be observed, thus leading to a fuller understanding of trithorax-related genes' involvement in cancer.

8. APPENDICES

8.1. APPENDIX A

Reagents used for experiments described in Chapters 2, 3, 4, 5 and 6.

Reagent	Company	Catalogue number
Chapter 2		
dNTPs – dATP, dTTP, dCTP, dGTP	Roche	11934511001, 11934546001, 11934520001, 11051466001
FirstChoice® RLM-RACE Kit	Ambion, Invitrogen, ThermoFisher Scientific	AM1700
HyClone Water, Molecular Biology Grade	GE	SH30538.01
MinElute Gel Extraction Kit (50)	Qiagen	28604
Phusion® High-Fidelity DNA Polymerase	NEB	M0530S
Qubit® RNA BR Assay Kit	Invitrogen, ThermoFisher Scientific	Q10210
RNase H	NEB	M0297S
RNaseOUT™ Recombinant Ribonuclease Inhibitor	Invitrogen, ThermoFisher Scientific	10777-019
SuperScript® III Reverse Transcriptase	Invitrogen, ThermoFisher Scientific	18080-044
TRIzol® Reagent	Ambion, Invitrogen, ThermoFisher Scientific	15596-026
TURBO™ DNase (2 U/µL)	Ambion, ThermoFisher Scientific	AM2238
Chapter 3 and 4		
5-(and-6)-carboxytetramethylrhodamine succinimidyl ester (TAMRA)	Molecular Probes	C-1171
Digoxigenin-11-UTP	Roche	11209256910
DyLight 633 NHS ester	Thermo Scientific/Pierce	46414
DyLight 488 NHS Ester	Thermo Scientific/Pierce	46403

Fluorescein RNA Labeling Mix	Roche	11685619910
MinElute Gel Extraction Kit	Qiagen	28604
pGEM®-T Easy Vector System I	Promega	A1360
QIAquick PCR Purification Kit (50)	Qiagen	28104
QIAprep Spin Miniprep Kit	Qiagen	27104
REDTaq® ReadyMix™ PCR Reaction Mix	Sigma Aldrich	R2523-20RXN
RNaseOUT™ Recombinant Ribonuclease Inhibitor	Invitrogen, ThermoFisher Scientific	10777-019
RNTPs –UTP, ATP, GTP, CTP	Roche	11140949001, 11140965001, 11140957001, 11140922001
SP6 RNA Polymerase	NEB	M0207S
Subcloning Efficiency DH5alpha Competent Cells	Invitrogen, ThermoFisher Scientific	18265-017
TA Cloning® Kit, Dual Promoter, with pCR™II Vector, without competent cells	Invitrogen, ThermoFisher Scientific	K2070-20
T7 RNA Polymerase	Roche	10881767001
Yeast RNA (10 mg/ml)	Ambion	AM7118
Chapter 5		
Anti-H3P (phosphorylated serine 10 on histone h3) antibody; rabbit	Millipore	09-797; Lot: JBC1787322
1,1,1-Trichloro-2-methyl-2-propanol hemihydrate (Chloretone)	Sigma Aldrich	112054-50G
Alexa Fluor 488 goat anti-rabbit	Molecular Probes	A-11008
Chapter 6		
Anti-H3 (unmodified) rabbit polyclonal antibody	Abcam	AB1791
Anti-H3K4me1 rabbit polyclonal antibody	Abcam	AB8895
Anti-H3K4me3 rabbit polyclonal antibody	Abcam	AB8580
Anti-H3K27me3 mouse monoclonal antibody	Abcam	AB6002
anti-mouse IgG HRP linked antibody	Cell Signalling	7076P2
anti-rabbit IgG HRP linked antibody	Cell Signalling	7074P2

cOmplete™ Protease Inhibitor Cocktail Tablets provided in glass vials	Roche; Sigma Aldrich	000000011697498001
Deoxyribonuclease I from bovine pancreas (DNase I)	Sigma Aldrich	DN25-10MG
Hoechst (bisBenzimide H 33342 trihydrochloride)	Sigma Aldrich	B2261-25MG
NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 9-well	ThermoFisher Scientific; Invitrogen	NP0327BOX
NuPAGE® MES SDS Running Buffer (20X)	ThermoFisher Scientific	NP0002-02
Papain	Worthington Biochemical Corporation	LS003119
Phenylmethanesulfonyl fluoride (PMSF)	Sigma Aldrich	P7626-1G
Ponceau S solution	Sigma Aldrich	P7170-1L
Qubit Protein Assay Kit	ThermoFisher Scientific; Molecular Probes	Q33211
Sample Buffer, Laemmli 2× Concentrate	Sigma Aldrich	S3401-10VL
SeeBlue® Pre-stained Protein Standard	ThermoFisher Scientific	LC5625
SuperSignal West Pico Trial Kit	ThermoFisher Scientific	34079
Tesco Instant Dried Skimmed Milk 340G	Tesco	N/A
Trypsin inhibitor (Trypsin inhibitor from chicken egg white)	Sigma Aldrich	T9253-250MG

8.2. APPENDIX B

Primers used in Chapters 2, 3 and 4. Red indicates 5’ T7 linker and blue indicates 3’ T7 linker.

Primer name	**Primer sequence (5’->3’)**	**Amplicon size**	**Origin**	**Procedure**
Chapter 2				
Ao	CATGCAAATGTCCAATCAGC	N/A	Author	3’ Outer RACE

Ai	GGTATCGAGCGTGCTGCAAA	N/A	Author	3’ Inner RACE
B	TGGGAAGATGCAACAGTCACAG	N/A	Author	3’ Inner RACE
C	GCGGCAATAGTAGCAGTAACC	N/A	Author	3’ Inner RACE
D_F (forward)	GGCAATAACTCTCGCTCGTTC	N/A	Author	+ D_R; Phusion PCR
D_R (reverse)	CACAATAAGCGGGCTAAGTGA	N/A	Author	+ D_F; Phusion PCR
Eo	CGGAACTGGTATTTCCGATGA	N/A	Author	5’ Outer RACE
Ei	CATATTATTGCCGTTACTTTGGA	N/A	Author	5’ Inner RACE
3’ Outer RACE kit primer	GCGAGCACAGAATTAATACGACT	N/A	FirstChoice® RLM-RACE Kit	3’ Outer RACE
3’ Inner RACE kit primer	CGCGGATCCGAATTAATACGACTC ACTATAGG	N/A	FirstChoice® RLM-RACE Kit	3’ Inner RACE
5’ Outer RACE kit primer	GCTGATGGCGATGAATGAACACTG	N/A	FirstChoice® RLM-RACE Kit	5’ Outer RACE
5’ Inner RACE kit primer	CGCGGATCCGAACACTGCGTTTGC TGGCTTTGATG	N/A	FirstChoice® RLM-RACE Kit	3’ Inner RACE
Chapter 3				
Smed-H2B_F	GGCCGCGGTCTGTTAAGAAGATTT CAAAGG	335 bp	Dr. J. Solana	RNA probe, dsRNA, cloning
Smed-H2B_R	GCCCCGGCCTCCTGTGTATTTTGTA ACAGC		Dr. J. Solana	RNA probe, dsRNA, cloning
Smed-porcupine-1_F	GGCCGCGGCCAGATGCAATCAAGT TTTTCGG	634 bp	Dr. F. Jaber-Hijazi	RNA probe
Smed-porcupine-1_R	GCCCCGGCAACCGATATTGGGCTT TTCTGG		(Gurley et al., 2008)	RNA probe
Smed-LPT_F	GGCCGCGGCAATGGTGACGACCC AGAAATC	561 bp	Dr. F. Jaber-Hijazi	RNA probe, dsRNA, cloning
Smed-LPT_R	GCCCCGGCGGCTCAATATCAGTGG CGGTTC		Dr. F. Jaber-Hijazi	RNA probe, dsRNA, cloning
Smed-LPT_F2	CAACAATTACACTCTCTGTGACC	473 bp	Designed by author	RNA probe (for pooled detection), cloning

Smed-LPT_R2	CAAGATCATCATCTGCAATCTCC		Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-1_F	CGGATCGACTTGTTCAGAAACC	422 bp	Designed by author	RNA probe, dsRNA, cloning
Smed-trr-1_R	CCGATATTGTTGCCATTTAC		Designed by author	RNA probe, dsRNA, cloning
Smed-trr-1_F2	GTTGTCTGTGTTTCATTGTCC	613 bp	Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-1_R2	AGTGCTGTTGATTTGTAGACC		Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-2_F	GGACGTATGTGAGACGAATG	473 bp	Designed by author	RNA probe, dsRNA, cloning
Smed-trr-2_R	GATCCGACTTCCACAATCCAGTGA		Designed by author	RNA probe, dsRNA, cloning
Smed-trr-2_F2	AACAATTACCGACATCTTCC	653 bp	Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-2_R2	TTGATCAGGGTAAACATTCC		Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-2_F3	TGTGTGTTATGTAGGAATGTGG	551 bp	Designed by author	RNA probe (for pooled detection), cloning
Smed-trr-2_R3	GTTGAATGATAACCGACTGG		Designed by author	RNA probe (for pooled detection), cloning
5' T7 Universal primer	GAGAATTCTAATACGACTCACTATAGGGCCGCGG	-	-	RNA sense probe
3' T7 Universal primer	AGGGATCCTAATACGACTCACTATAGGCCCCGGC	-	-	RNA probe
M13 Forward	GTAAAACGACGGCCAG	-	-	Orientation of plasmid insert;

				RNA probe, dsRNA
M13 Reverse	CAGGAAACAGCTATGAC	-	-	Orientation of plasmid insert; RNA probe, dsRNA
Chapter 4				
Smed-GAD_F	GGCCGCGGACGATTGCCGTAGCTGGAGA	534 bp	Author	RNA probe, cloning
Smed-GAD_R	GCCCCGGCATGAAGCCATATGCCAAAGC		Author	RNA probe, cloning
Smed-TH_F	GGCCGCGGTGCACTAGAAGGTCTTACAAGG	418 bp	Author	RNA probe, cloning
Smed-TH_R	GCCCCGGCGGACCATATTCACACTCCTCTT		Author	RNA probe, cloning
Smed-TPH_F	GGCCGCGGCGTTCAAGAGGTCGACATGA	576 bp	Author	RNA probe, cloning
Smed-TPH_R	GCCCCGGCTACACGGAATGCAAGACCAG		Author	RNA probe, cloning
Smed-chat_F	GGCCGCGGCATCCCTTGCACTAAAATTTCC	846 bp	Author	RNA probe, cloning
Smed-chat_R	GCCCCGGCATAAGGTTGTTGGTGTGATTGC		Author	RNA probe, cloning
Smed-pitx_F	GGCCGCGGGTCATTCTCCATCGGCTCAT	795 bp	(März et al., 2013)	RNA probe
Smed-pitx_R	GCCCCGGCTGACAACATTGGCTGTCGAT		(März et al., 2013)	RNA probe
Smed-laminin_F	GGCCGCGGTGATGCGTGGTATCCACAGATGTGC	527 bp	Dr. F. Jaber-Hijazi	RNA probe
Smed-laminin_R	GCCCCGGCTCGCTGGCAAAGTGCATCTGTGATC		Dr. F. Jaber-Hijazi	RNA probe
Smed-CAVII-1_F	GGCCGCGGCTCTCGAAAGACATGAAACC	568 bp	Author	RNA probe
Smed-CAVII-1_R	GCCCCGGCTTGGTCTGAAATTGTTCTCC		Author	RNA probe
Smed-SP6_9_F	GGCCGCGGTTCAATAAATAACGTTGAGAGCAA	797 bp	Author	RNA probe, cloning
Smed-SP6_9_R	GCCCCGGCTTCACAATTGTTTGTTAACGACTC		Author	RNA probe, cloning

Smed-AGAT-1_F	GGCCGCGGGAAATGATTGAGTCCACCATGA	504 bp	(Eisenhoffer et al., 2008)	RNA probe, cloning
Smed-AGAT-1_R	GCCCCGGCTGCAATATCTGGATAAGGAGCA		(Eisenhoffer et al., 2008)	RNA probe, cloning
Smed-NB.21.11E_F	GGCCGCGGGTGATTGCGTTCGCGTATATT	574 bp	Dr. J. Solana	RNA probe, cloning
Smed-NB.21.11E_R	GCCCCGGCATTTATCCAGCGCGTCATATTC		Dr. J. Solana	RNA probe, cloning

8.3 APPENDIX C

Databases used for phylogenetic analyses in Chapter 2.

Species	**Database**
Schmidtea mediterranea	Kao et al. (2013), Labbé et al. (2012) and Önal et al. (2012) – combined into one searchable database by Dr. Damian Kao
Schistosoma mansoni	http://www.genedb.org/Homepage (Logan-Klumpler et al., 2012)
Clonorchis sinensis	http://www.ncbi.nlm.nih.gov/
Drosophila melanogaster	http://www.ncbi.nlm.nih.gov/
Capitella teleta	http://genome.jgi.doe.gov/Capca1/Capca1.home.html (Grigoriev et al., 2011)
Caenorhabditis elegans	http://www.ncbi.nlm.nih.gov/
Homo sapiens	http://www.ncbi.nlm.nih.gov/
Strongylocentrotus purpuratus	http://www.uniprot.org/ (Consortium, 2015)
Mnemiopsis leidyi	http://research.nhgri.nih.gov/mnemiopsis/
Saccoglossus kowalevskii	http://www.metazome.net/search.php
Tribolium castaneum	http://metazoa.ensembl.org/Tribolium_castaneum/Info/Index
Helobdella robusta	http://metazoa.ensembl.org/Helobdella_robusta/Info/Index
Branchiostoma floridae	http://genome.jgi.doe.gov/Brafl1/Brafl1.home.html (Nordberg et al., 2014; Putnam et al., 2008)

Brugia malayi	http://www.wormbase.org/tools/blast_blat (Harris et al., 2014)
Nematostella vectensis	http://cnidarians.bu.edu/stellabase/blast/blast_java.cgi
Lingula anatina	http://marinegenomics.oist.jp/lingula/ncbiblast/search?project_id=47 (Luo et al., 2015)
Daphnia pulex	http://genome.jgi.doe.gov/Dappu1/Dappu1.home.html (Colbourne et al., 2011)
Danaus plexippus	http://www.butterflygenome.org/
Hydra vulgaris	http://www.kegg.jp/kegg-bin/show_organism?org=hmg (Chapman et al., 2010)
Crassostrea gigas	http://www.uniprot.org/ (Consortium, 2015)
Planaria torva	http://planmine.mpi-cbg.de/planmine/begin.do
Dendrocoelum lacteum	http://planmine.mpi-cbg.de/planmine/begin.do

8.4. APPENDIX D

Sequences of Transcript A, Transcript B (Labbé et al., 2012) and Transcript C (Kao et al., 2013) corresponding to *Smed-LPT*, *Smed-trr-1* and *Smed-trr-2* respectively.

>Transcript A (*Smed-LPT*)

ATAATTTTCATATATAATGTCTGATGAGAGTAAGGATTCTGATTTATTTGATTCAATGTTATCACTTAGTTCTAAACGCCATAAAA
TTGACATTGTTAATGAAGAAATTTCAGATCCTTCTAATACATTAAGATGCAAGGTTGAAGACGAATATTTCGTTAATAGGATAATT
CAGAATAATAATCCAACAAATGAAAATGAATTTGATGATTTTTTCCAAAGTAACGGCAATAATATGAGGTGCTCGTTGTGTGCTCT
CGGTCCACAAAATAAATTTGGTCAAGGAGAAATGCTGAAATTCTATTCATCGGAAATACCAGTTCCGAGCTATATTTGGTCAAAAG
ATGTGAATAGTAACAAAAGGTCTGATCGAATCAAATTGAAAGATTGTCAGGAGTATTGCTGTTTGGATACAAATAAACGGTTGAAA
GAAAACAAAATAATTGGAACCGACGGTAAACCGTGTCTCAATCGAATCGTGAATGCATTTTTTGACAATGAAGGCCGGCCTTTTGG
ACTGATTGATGAATTCGCTTGGGTGGGGTTGCCAGCAATGGTGACGACCCAGAAATCTAGTACTGGAGCTGGTAGTAACAACAGTT
CCTTGCTGTGGAAAGCCATTAACAGCTGCGATGAGACGGATGAAAATTTGAAATTTGCTCATTTGCTATCGTACGAGTTGTCCGCA
AATGACCAATGCAATGACAGCTGTTTCTACGTGCATCATTCCTGTGTTACTTGGTCGGAGGGGGCGTCTCAGTTTAGTGATCAACT
AGATGGCAATTTTGACAATGTCGAATCAACAGTACGCAATGGATTAAATCAAATGTGCAGTTTCTGCCTACGGTTCGGGGCGACGA
TTTCTTGTCGCAGCAAGACCAATTGCAAGCACAAATTCCATTATCCATGTGCCGCAGCAGCCGGCTGCTTCCAGGATTTCGTCAGT
TTCCAACTCTGGTGTCCAAGTCACACTCAGGAGGCGGATTTGTCAAATGAGCATTGCATTTGCTCGCTCTGCGAACGGAAGGGAGG
CGTCAATGACCTGATCTCGTGTACGAGTTGTGGCCAGCATTACCACAGCAACTGCCTCGAACCGCCACTGATATTGAGCCCGACGG
TGCGTGCCGGTTGGCAATGTGCCGAATGTAAAACGTGCCTCGTCTGCAGTGAAAGCAAGGACGAGAAGAAGATGCTCGTGTGCGAC
GTGTGCGACAAGGGATTCCACACGTATTGTCTGCAGCCGCCAATGGCCGCCATACCGAAGAATGGATGGAAATGCGACTCGTGTCG
CGTGTGCAATGACTGCGGATACTCAATGAAACCGGGAATCGCTCTGACCGGCAACTCTAATCAGGGCTCTAATGTGCGATGGCACA
ACAATTACACTCTCTGTGACCGATGCGCACAGACCAAGCGGCGTGGCCATTCCATCTGTGCTGTCTGTGAGCGGGCTTGGCGATGC
TCTCTACCTGGCACCAAGAATCAATCCTCCAGCAAATGGCCTGGCCGCAAATGCTCCACGTGCCACAAAATGGTTCACATGGACTG
TGACCCGTTGCTGGCATCCATGCCGTCACCAGTCTATTCCCTGAGTTCATCCAACAGTGGGGGTAACGGCACTGCCGATGAGATAT
CCACGGCGAGCTGTGCATTAGTTGCTCCTAGCAACGGCGGAGCAAGTCATCATAACAGCAATAGCAACACCAATAGCAACACCAAT
AGCAACACCAATAGCAACACCAATTACCAATGTCCTACCTGTCATAATAGGGGATCAGCAACGTCCAGTGAAGTGGCCTCCACTCA
CACCAGTCCTATGCCGGGCAGCTACTGCGATGAGGAGATTGCAGATGACCTTGATGAGGTGGATGACGAGGAGACTGCATCGCTGT
CGGGATTCGACCCGTTGCCGCCGAGGATGCTCGGACTGGACTCGCAATGCTCTACAATCATGAACCGCAGCATCGAGCAGGGCTAT
CGCAGTAACAGTATTGACGATTCCAGCACAATATCAGGCTCGACAACTCCACAGACGTATTGCCAGAGCAACAGCAGCAGCAAATT
CCAGTTGCCAATGAGCAGAAACAATAAAGATAAGCCCTATGAATCAAATGCGGCACCCAGTAAGAAACGTTTGCAGCCAGTAAACA
AAGGGAAGTCAAAGACTGCAGGTGGGAAATCATCATCCTTGAAACGGAACAGCACACTCAACAGCACCAAGAGATCTCGGAGGGTT
AAATCAAGGAAAATTAATGTGTGTGAATCTAAAAATAACGAGGAAAAAGACGAGCATCCGAGTACTGTAGTTCTGTGTCGAGCAGA
TGATCTGTTTGTGCTGGAGCAAGATATCTGTGTGTCGTGTGGCTCATTCGGGATCGACACGCCGTTGCTGTGCTGCGTGCAATGCG
GCCAATGCTATCATTCATATTGCGCTGAGATTAGCAAGCTCACAAAAGTCATCATAGAAAAAGGCTGGAGGTGCCTAGATTGCACG
GTCTGTGAAGGTTGCGGTCAGGCAACTAATGAGGACTTGTTGCTATTGTGCGATGATTGCGACATCAGCTACCACACGTTTTGCCT
CGATCCCCCATTGAATGAAGTGCCGAAAGGCAGCTGGAAATGCCAGAATTGTGTCGTGTGCCAGTTCTGTGGGAAACGGGACCCGG
GGCCCAACGGGACCTGGCAAAATAACTACACAACTTGCGCGCCGTGCGCTTCCAAGGCCCAATGTCCTGTCTGTTATAATAATTAC
AATGATGGGGACTTGATGCTAATTTGCGATGTGTGTGAGAGATCGTCACATGCCACTTGTGATCATTTGAATACTGAAGCGGAGTT
AGAGTTCATCTTTGAAGGCGGAGGATATCATTGCATTCTGTGCCGGGACAAAGGGATCGATATCGGTGCTGGTCACAAGCAGCTGC
TGCATTTACGTGCTACTCATGGCGATTCAGTTGCTCAAACATTGGTTGCTGATAAATCTGTATTTAGCAATCTAATTGGCAACTCA

TTGGAATACGGAAGCAAGAACTCGTTGCTATCCAGTGTTGGATCATCTGTAGACGGGCCTATGTCACCCGATCCTTCCAAGAACTT
CTTCATGGACAATGTCGTTCTCAGTGAGAGTGGCCTCAACACCATTAGAAACGCCATGCTCAAAAGCCAGCCACGTCGGGCTCAAA
ATGCCATGAAACGAGTACAGAAATCGCATTCCCAAGATACGATTGGAGATGTCGACAGCAAACAGATCAAGTTGGAATCCACCGAG
CCCCTCTTGAATGCCGACGGCTCTGTTGTTGCTAGTGGCAACCATCTTGACGAGGACTCAATGCAGAGCATCGAAGGAGAAATCAA
CACCACAGACATGAAGAGTCCAAGCGTCCAATCAGATGACGGATCTGAAATACAGAGTTTGCCAGCTCATTCTGCCGCAAACAGCG
TCAATAATTATCTGCCCTCCTCGAAAACCATTGGGAAAGCAGTTATGCCGCGGAACAGAAGGTTCTCAAATTTGGGCATCGGCGGG
TTCAGGGCCCGATCGACAAGGGTGTCACAAGTGAAGAAGTCGGCATTGGCTGCTGTAAATAACCAGAATTTCGATCAAAACAACCC
GCAATACACAATAGAAAATTTCACAGCGAAGCAGGGCCGACGGAGACCCGTGAAGAAGAAATCTGAAATTGATGAGAACTTTCCCG
ATTATTTAATGCAGGCTTTTCTTGGTAGCAGCCTTCTGTCGCACACAACTGAGACGAGCAAGAAATCCGGCCACCTCAAATTGTCA
CCGAGCAGAATGGCTGGCGGCAATTACTCGGTTGATGCTCATTCATTCAGCAGCACCGGCTCCAATCTGAATGACGGTA

Transcript B (*Smed-trr-1*)

CCGATCTAAGGTCTGAGCCGTAAGAGATCATTAGTTCATTGTAGCTTATTATATATATATATATATACACATATATATWTCAGGTA
TATATCTGTATAGAATTTACAAAACTATCACAAAATTACTCTTTTAAATAATGTTGTCATTAATGTCATCATTATCGTTGTCAAAA
TATTAAGTATTATTAAATTGTAAAAGCATCGTTGGCGTTATCATCATATGTTCAATTCATCCACCGTCGACAATTCCTGGCTCCGC
ACAAACACGGGATTTTATCATCTTTACCGTCATCTAAATCAAATTTGTAATCATAGGTTAATTCTTCTCCTTTCTCGATTTTTCTG
TTTGATATAATGACAATGTGACTCTCTTTTTCAAAATCCAACAGTTCAGTAAAACAATTGGGATCACAACAATGATTAATATATCT
GGCTAAATCACCGGCTCTGGTAGCATCGATCACTGTTGAAGGATCGAGGCGAAACATGTAGATTCCTTGCTTTTTACTTTCGTATT
GTTTCTCACGAATGTTGGCTACCTCGTTCCGAATCAACTCCCCAACGTATTCTATGATCATTGTACCTTTTTCAAGATCTTTAGCT
GCATAAAGTCCGAATCCCTGAATTTTGGATCGCTTCAAAACCACATTGAACTTGGTTTCGAATTTCAGTTTCCGATATTGTTGCCA
TTTACTACTGAGAGTTGTTTTGTTGTTGAATAGAATGTTGTCTGGGATTTTCCCTGATGTGGCCGTTGCTGCTGCAGCAGCCAGCA
AATGCCTCAGGAAGTTAGGAGATTGATGGATAGYGGGTTTGTGAATGTGCCGATAAAGTTTTACGAACGGCTCGGTCCGAATCGAT
CCGCTCATATTAATCGGGAGTGGCATTTCGACAATTAATTGCATTTTACCGAATTTGAACGCGTAATCATTTAAACTCTCTAAGCC
CGGCAACGATTCAATAGCTCTTAATACGTGTGGTTCCGTTAATCCATACAAATCTTCAGCTTTGATATGTTCCGGAAAAAATTTCA
CTACTTTGTGTGATTTCCTTAGATTTGAGATCTTCGTGAGAATTTTTTTCCAGGTTTCTGAACAAGTCGATCCGAAATAGCACTCA
TTCTCCATGTTTTCATCTGTTGCAGTGACTTTAAATAAAGGCATTTTTTGATCTCTGTCCATCAGGATTTCACATTCGTAAACAGA
TCGAGTATTGATTACTCGAGACGACCAATGAATCCGATGGGACTTGAATCCGGTCGGATAGACGTGATTGAGATTGTGAAAGTTTC
CTGAATGTATCTGGTGTGGAAGCAGCTGCCCGATTGATAGAACAACGAGAGAACCGATCCTGAGAAAGAAATTCTGATCCTCATTA
TTGATTGTTCCAGCAACTTGGGAATATTCATCTCGAGATATGTAGACTCGCCGATCAACATTCAACGTTGTTAAATGAGGACCCAA
ATCTTTTGTTTGATGCAACTGGCAATACATTGCTCGATCAGCAAAAAACAAGCACCCTTTTTCTTGAGCACAAGGCACGTGATATT
TCAATGTACATCGGGGATAATAGCAAACGATTCCGGCACCGGGTTTCCCACACGAGCTGCAAATTGATTTCAATGAGGCTTCGAAT
GACTCGGGAACATTATTCAGTGCTCCGCATTGCATTTCGAATACATTGTTACACCATAAAGCGCAGTTGATGTGTAGCCATTTGTC
AACATCGAAACACAGCAACCGACCCGGCCCATTCTCCGGTAGATCTCCGGTTCGATGACACAGCGAACAAGTTCGTTTATCAACGA
ATCCAGTTTTCCGCAAAACGGATTTCGATTTCATTGACAATTTCTTGAATTGCGATTCTGATATGATAGATGKGCTCGTCGGTTCC
GGAATCGGGTTCGAATGGAACACAGAATATTTGCTGCCTTTGTGGATTTTCTGCGGGGACGATTTAACTGAAGGACCGATTGCGGC
CTTTTTCTTGAGAAGTTTGTCATTGATAAGAGCACCTAATCGTTTGGAGATTTGTTTGGATTTCGATATGGATCCTTGAACTATTA
CCGAAGCGTCACAAATATCATCTCCAACATAAATGCACTGATCCTTGATGTCAGAACACTTAGTAGTGGTGTTGTTGTTGTTRYTG
TKASTGTSAKYRTCAWSWNNTSAAKWSTYSWRTWSKCMWGYYKNCAWSMCWSAWNCACAGAWNCWCNGATMTYNSWTMYNNGATCM
SATMACATAWYRTNATTGKWATACATAAATTTCTGATCCGGTATTAATTTGTCACAATTTGAACATTTTAAACAAGGCTTATTAAG
AAGTGCTGTTGATTTGTAGACCTTTTCTGAATAAAGAGGAAGCTCAGTTGATTTATCAATTGATATGAGATTATCGAACGGTTTCG
TTGTAATACTATTGTTAATGTTAGTAGTGGACTCCATCTTGTAAACTGATGTATTATATCGTTCACAAATTTCCCTGGGAATCGAA
TTGCTGACCATACTGTCTTTGTCGATAGCGATTTCCCCACCGCGAGTCGTTTCTTTAAATTCGACTGAATTGTAATCGATTCCCAA
CACATCGGTAATATTTCTCACGGTTTTCTCAACATCATCCGGCGGATTGATAGTGAATGTCATCAGTAATTTCTCTCTCGGCTTAT
TGTTACTAGATCTGGTTTGATCATTGACTGGTTTACTGACAATCTCACCGGTTTCCGAATCGATTAGGGTGTTTTCGGGAATCGGT
TCAGCCGCTGTGGAAACTGATGAATATTTATGGGATATTTGAACCCGTGAAATTCCGGTGGGRCGCAACAATGATTGTGAATCGGG
TTTTCGTGTCGACGAGGATGATGACGAAAACGACGTCAACGATGATAAATCCGCATGATCGAATTTGAGATAATCTTTATTCGGAC
AATGAAACACAGACAACTGCGGAGATTGATTGCGATTTGGTGGATGAAGCATTTCCAGATTGGGACCAGCAGATAAATTGACAGTC
GGCGATGAAATAGGATCACAATTGAATACACTGAGATAATCGAGTTCATTGTTACTCGGAGATGGCGCTATCAAACAATCAATATC
GTCTGATTCGTGCGCCGACAATAGCCCACATTCCGGAATACTGACATTGATATTCTGGAGAATGAAACTGTTTTCCGAGCATCGAT
TAACACCGAGATTGATCGGATTGATTTGTGGCATTGCGAATGGCATTCGTTTGTATTTGACGCAGTTGTCCTGCAGCTGTACGATG
TCTGAGCATTGAGTTGATCGTTGTTTGCAGCCAAGCGGTGCAAGAACGGTTAGATTGTTTTTTATTCGAGGCTCATGGACAAGTTT
GCGGAAATTCGGCTGACCCAAAACCTTAACCATATTTTGTTGACAGCGTTGCAGTGGATTGTTCAATTTGTCATGATATATAGGCT
CATTCGATAATTTTTGCTTTTTGCTTGTCGATTTCGCGACAGATGAATTGCGACGTTTGTTATTCGTCGGTTTCGCAGACGGCTCT
AAGTTTGAATGTAAAAGTTCGTCATTGAGATTCGGTTTCTGGTGGATTTGGCAGTTGACAGGATAGAATTCAGATTTCGGAAACGA
TTGAGCAAGGATGTTCGATTTGACATCATTTGCGATTGGTTTAATGTCATTCGGATTCGAATATTGCGGATGGGTAAAATTCACTG
TAGAAATGGTTTTGTTCGGTCGAAACATTTCTGAATGACGTTGAAATGGAATGTTGTTTTGAATCGTCAATTGAATTAGTGGGTTA
TCAGAAATCTGTGATAAAATCGGCGACGACATTGAGCTTGACATTTCTAAACTATCGACAGTTTCCGTAACAGAACGAGAAACATT
TTCTCTAGATTTAATCACACTGAGAAACACATCTTCGTGAGGATCATTACTGAGAGCTTTAATTAACACCTCTTTGATAACTTCTG
TAATTGTGTCTCGTGATTCATTCACATTTGTCTGTACAACTGGAACATTAGGTTGAATAATTTCTGAATCTGTAGAAAGAGGTAAA
AAATTTACATCTTTTTTATCTGGAGATATTTCTTTTTTTAATTGATAAATCAAGTTTTTTCTCACTTATTGGAGAAGTTGATATGAT
ATTTGAAGAATGAGCTTTAATTGCATTTAAATGAGGTGGAATTGTCGATGTAAGCACAGGATTATCTGGCACTTGCGCTGGCAGAA
CCTCTTGGACAAGAATAGTTAACTGATTCGGTGCAGCTGTAATATCACTTGAATTGTCATTGCTATTTAAACATTGTGATGTCTTC
ATTGTTAAATTTCTTTTATATTGCTTCTTATTTTTATTAATATTTGAAGCCAGAAATTTAAATTCATCACTTTGATTATTTGTTGC
ATAGCAATTATGGGCAATGGAACTCGGTAGATTTTGAGCATTTTTTAAATCGTTTGACGTGCTGTTTTGATATTGTCTGTAATTTA
CTGATGAATTGCTATTATTAGGCACAATGGTGTAATTAGTAGTTTTATTTGAGCCGATTACAGTACATGGAAATTGAGCCGAGGAA
ATAATATTCGGTGATGCCAAATTCGATATAAAATATGGTTGTTCACTTTGATTTTGTTCCATGTGTTAAAAAAATTTAATTTCAAT
AACAATTTACGTTTAATGTTAATCATATAAGATTTATATAGA

Transcript C (*Smed-trr-2*)

TTAGATTAACTTATATTTTACCTATTTTATTTCTTTCGGAACAATATTTTATTTGTTAACATATCGATGACATGATATAAAATGAC
AACAGAATATTTTGATGCAACAAATTCTTTGTGTTCTAATTCATTTGAGTTGATCAAAAGAAGAAGGTTAATGTCAGATGATTTTA
AGAATAGTGATTTAGATGGTGAAAATTTAATGGATTATTCTGGTGATAATTTTCAATCAGAAATATCTGCAAAAGCATCTTCCAGT

TCAATTCACGAATCTATCACTAACATGTCTAATTCGCATTCGACTCCCGAAGATACATTTTCCGAAATCACTTTATCTTCACCAAC
ATCTAAGCATTGTAAACCAGATTCCGATATTAAAGATTATGAACCATCTATTGATCAGTATTTGTCAAATTCTAATAATGACGTTG
TTGGCAATCAGAATTGTAACGATATCGTTAAACAATTACCGACATCTTCCATTAATCTTAAAAGTATTGATGATACAATCGAACTA
GTTCTTTCAAAAATTAGAGAAGAAAATAACTTTTCAAAGAATTTTGATTCTTCTGATTCGTGTTATAAAATTAAATCTGATAATTA
TAATTCCAGTGATTTATCTAATTGTTCTTCCAATATTAATGTACAAAATAATACTCTGGTTTCTATAGCACGTATCACTTCAGCGA
AAACTGGATCTCATGGTTCAGATAATATGTCATATTCAAAAGGAATTATTTCTAGCGCATCTTGTCTTAATTCATCTATTCGTAAT
TATGGTATGTCTAATTATTCAAAATCGTATTCTCAATTTTATAATTATAATTATGGAATTTCACCTCAGATGGCAGCAAATTTTTC
TAATCAGATCACTGCTCCTAATTATTTAGAATATAACAAAATGGATTATACTCGCTCAAATTATCTCAAAGAAGCTGATAAAAATA
AATTTTTCTTTCCTTATCAATGTCAAAACCCTACTTATAGACAGTACCTTCCAAATTCGTATAATAATCAACCATTTTATAAACCA
GATTCTCCTAATATACAAATTCACAATCAGAATTATCAGATGTATCAAAAGCCTGTTAAAAGGAATGTTTACCCTGATCAAATTTA
TGTTGATGATGATACTCGGCTTTGTAAAGATAATTCCAAAAACCCAAATTATAATATTTATTTCAAAACTCCTCAACCAAATGCAC
ATGGTTTATTGGAACAAGGGAATTCTGTTCGTAAAAAACGCACAAGTTCAGATTCTACAAAATTAACACCTAGTAAAAAGCGGAAA
TTAGCCAATCAATTTTCCATAACAAATAATTACGAAGAGTTAATAAGAAAATTAAAAATTTCTAATTATAATTCCTTATTAATGAA
TCCAACAGTGTCTCAGTCTGTTATTGTTCTTTCACCAAAAGGATTTATAGAAAATTCTAAAGCTGTCAATCTTAAAAAGTCTGCCA
GGTATAAAAAATATAAATCAACTTTGTTTAAACTTAAGAAGCCAAAAATCTTTTCGGATTCCACACAAGAAGATCGTTATTTATCA
CATTATGTGGATGAAAGTAATGCTGGGAGTTGTATATTGTCACAGAAAAAGAATTTAGATAATGATGAATTGTTAGACATCATTTG
TGGATCTCCGGGAAGTAACAAAGTCACGGAGAAATCAGAAAACTTTTGTTTGCATAATCTTGCAACAGACAACGTCGATTATGATC
CTGCCAGTCCGGAATTGACAAGTTCTTGTTCTAAAAAGAAAAATGATCAATCTAATGAAATCAATTTTTCCGAGCCTAAAATTGAT
GCGATTAATTATGAGAGAAATCATACAATTACTCCAATAGTTTCTTATACAGAAAGTGAAACCAAATTAATAACTGAATTACATCC
GACTAAATCTCAAGAATTGAACAGAAATTACATTCAATTATCTGGCAGCTTTCTGAAACGTACGAAAATCAATGGAAAATCATTTT
CGAATATTCACATCGAAGAAGCCGTCAGTATAGCGTCGTTGTCAGCGTCTAAAAATCAAAACACGTGTTATTACTGCTACGCATTT
CTTAATGGGATATCTTTCATTAAAAAAAACAGCAAATTATTTGGAATGGATTCTACAGAAGATTTATTATTCTGCAGCTCGAGTTG
TTATTCAAATTTCAAGTCATACCATTCGGTTGAATGTAATTTAGAGAAAAAGTCCACTAGCAATAATAACGTCCAGTGCACTAAAG
TTTTGATTAAAATTACTCCCTATAAACCGGCTAAAAATTATAAAAAAGTTTGTGATTTTGAACAATATGTTCGTCGACGATTTCCC
AGTGAAGGCCAAGGAGACAGTTCGGTAAAACCGGCTGTTGTCGTTACTGATTCAATCAAGAAAAGATTGCTCGGTTTAAATTATTC
CTATTTCAAAACCGAGCCATTGAAAACAATTTATAAAGACATTGAAAAAAATGATCAAAATCTATTGCTGAAAAAGTATTCCTGTT
CTTATCGGAAACCGGACTTAGTTGATGTTCGGTCGTGTGTGTTATGTAGGAATGTGGGAGATGGTGACGAGGATATATGTGGCAGA
CTTATCAATATCGATGCCGATAATTGGGTCCATCTGAATTGCGCTTTGTGGTGTTTCGAAGTATTCGAAACTGTTAGTGGATCGCT
AATGGAAGTTGATTCGTCTTTACGGAAAGCAATTAAGACCATTTGTACGTGTTGCAATGAAACCGGTGCCGGAATCCAATGTTATC
GCAGCAAATGTACGAATGTTTACCATGTGCCTTGTGCATATAAAGTCGGTTGCATATTCTACACCGATCGTGGTATGTATTGTCCA
AACCATGTCGTTTTGAATCATCCTATGCATTTGCCATCGTTGAGAGTTGAACGACGAGTCAATATCCTGCGAGATGAAAGTGCTCT
TATTGCCGGAATGGTAACAGAAGATCAGGACGTGAAAATCAAATTTCGTACCGGATCTTTGACATTGAATAATATCGGTCAATTGC
TTCCTGAACAACTGATCAGTGGAAATTATAATTCGTGTTCAATAATCTATCCAGTCGGTTATCATTCAACTCGGATTTATTGGTCG
TCAACTCAAATCAACAAGCGATGCAAATATGAATGTCTAATCGACGAATCCGATGACATGCCTTTATTTAAAGTGAAATTTATCGA
CAATGGCTGTGAAAATGTTTTGACTGATAAATCTTGTGATAAATTGTGGAAAACAATACTGAATAAAATTGCAAAATTGCGCCAGG
ATTCCGGACTAACAAGGCTGTTTCCTGTGTATCTGTCAGGTGAACAGTTGTACGGTTTGACTGAACCCCAAATCATTCGAGTTATT
GAATCTCTTCCGGGAGTGGAAATGATTAAAAACTATGCTTTCAAATTCGGTAAAATGCAACTGATTACCGAAATGCCTTTAGCCAT
CAATCCAACCGGTTCTGCTCGAACAGAACCAAAACTGCGGACGTATGTGAGACGAATGAGAGATTATCAGGAAACCAAATTTAACA
AGTGTGGAATTTTGAAAGCAAAACCCAACAACAACTATCAACCAATCATTTATGACAATGGCAATTTAAACAAAACCGCTTTACTC
AACTCAAAAACTCAACAGTACAAACGACTGAAATACGAATGGAGAAACAATGTAATTCTATCGAGATCTAAAATCCAAGGACTGGG
ATTATTTGCTGCTAAAGATTTTGAGAAATCGACAATGATTATTGAATATCTCGGTGAAATAATCCGAAACGAGGTGGGAAATAGTC
GGGAAAAACTCTATCAGGCTCGAAATAACAGTATCTACATGTTTCGGTTGAATGAAGATTTGATCATAGACGCTACAGTAAACGGA
GGATTGGCACGATATATCAACCATTCCTGCAATCCGAATTGTGTCGCGGAGACCTTGTCACTGGATTGTGGAAGTCGGATCATCAT
CACGACGCGTCAACTCATCGAAAGGGGTGAAGAGTTGACTTACGACTATAAATTCGATTTCGAAGACAATAAAGCCGACAAAATCC
CTTGTTTGTGCGGATCAATCAATTGCCGGAAGTGGATGAATTGATCGTCTGTTGAATGAATTTGCTTTCCTCAAATGTATTAATAT
TATTCTTTACTTTCAAGATTCCTTTTAACACAAATTGTAAATCTTGTGCATATATTTTGTGTAATTAAGAATTTGTACATTATTTT
GTTATTAATCTGTAGATGTTACTTTCGGAATTACGTTATGTGATGATTAATAGTTATGTTTTAAAAATTTTTTTCACGATTGTATC
ATAGTTTAATACATAATTTCTATCATTTTTTTATTTGTATCCGAACATAATCACGCTAAATTATAAATATAAATTCGAAGATTATT
TTTACGTTGTTGCATAATTTGGCAATTATTGTTACTTTTTATGGGTCTGTGATTTTCCTTGATTTGTCATTTTTATCAAAATGGAA
GGACTGTTGGTATCCCGATATATCTGTGAGGTAAAAATTTTCTAATTGTATTTCATATTTCTAGGTAAAAGTGACATGTAATTGGG
CTCCCAGAAAACCTAAATGCAATTAAATTTAAAGGTTAAA

8.5. APPENDIX E

The sequence of *Smed-LPT* after 3’ and 5’ RACE:

GGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGGCTTCGCGGATCCGAACACT
GCGTTTGCTGGCTTTGATGAAAAATTAATGCAAATCAAAAGTGAAATATATATAATTTTCATATATAATGTCTGATGAGAGTAAGG
ATTCTGATTTATTTGATTCAATGTTATCACTTAGTTCTAAACGCCATAAAATTGACATTGTTAATGAAGAAATTTCAGATCCTTCT
AATACATTAAGATGCAAGGTTGACGACGAATATTTCGTTAATAGGATAATTCAGAATAATAATCCAACAAATGAAAATGATTTTGA
TGATTTTTTCCAAAGTAACGGCAATAATATGAGGTGCTCGTTGTGTGCTCTAGGTCCACAAAATAAATTTGGTCAAGGAGAAATGC
TGAAATTCTATTCATCGGAAATACCAGTTCCGAGCTATATTTGGTCAAAAGATGTGAATAGTAACAAAAGGTCTGATCGAATCAAA
TTGAAAGATTGTCAGGAGTATTGCTGTTTAGATACAAATAAACGGTTGAAAGAAAACAAAATAATTGGAACCGACGGTAAACCGTG
TCTCAATCGAATAGTGAATGCATTTTTTGACAATGAAGGCCGGCCTTTTGGACTGATTGATGAATTCGCTTGGGTGGGGTTGCCAG
CAATGGTGACGACCCAGAAATCTAGTACTGGAGCTGGTAGTAACAACAGTTCCTTGCTGTGGAAAGCCATTAACAGCTGCGATGAG
ACGGATGAAAATTTGAAATTTGCTCATTTGCTATCGTACGAGTTGTCCGCAAATGACCAATGCAATGACAGCTGTTTCTACGTGCA
TCATTCCTGTGTTACTTGGTCGGAGGGGGCGTCTCAGTTTAGTGATCAACTAGATGGCAATTTTGACAATGTCGAATCAACAGTAC
GCAATGGATTAAATCAAATGTGCAGTTTCTGCCTACGGTTCGGGGCGACGATTTCTTGTCGCAGCAAGACCAATTGCAAGCACAAA
TTCCATTATCCATGTGCCGCAGCAGCCGGCTGCTTCCAGGATTTCGTCAGTTTCCAACTCTGGTGTCCAAGTCACACTCAGGAGGC

GGATTTGTCAAATGAGCATTGCATTTGCTCGCTCTGCGAACGGAAGGGAGGCGTCAATGACCTGATCTCGTGTACGAGTTGTGGCC
AGCATTACCACAGCAACTGCCTCGAACCGCCACTGATATTGAGCCCGACGGTGCGTGCCGGTTGGCAATGTGCCGAATGTAAAACG
TGCCTCGTGTGCAGTGAGAGCAAGGACGAGAAGAAGATGCTCGTGTGCGACGTGTGCGACAAGGGATTCCACACGTATTGTTTGCA
GCCGCCAATGGCCGCCATACCGAAGAATGGATGGAAATGCGACTCGTGTCGCGTGTGCAATGACTGCGGATACTCAATGAAACCGG
GAATCGCTCTGACCGGCAACTCTAATCAGGGCTCGAATGTGCGATGGCACAACAATTACACTCTCTGTGACCGATGCGCACAGACC
AAGCGGCGTGGCCATTCCATCTGTGCTGTCTGTGAGCGGGCTTGGCGATGCTCTCTACCTGGCACCAAGAATCAATCCTCCAGCAA
ATGGCCTGGCCGCAAATGCTCCACGTGCCACAAAATGGTTCACATGGACTGTGACCCGTTGCTGGCATCCATGCCGTCACCAGTCT
ATTCCCTGAGTTCATCCAACAGTGGGGGTAACGGCACTGCCGATGAGATATCCACGGCGAGCTGTGCATTAGTTGCTCCTAGCAAC
GGCGGAGCAAGTCATCATAACAGCACTAATAACAGCAATAGCAACACCAATAGCAACACCAATTACCAATGTCCTACCTGTCATAA
TAGGGGATCAGCAACGTCCAGTGAAGTGGCCTCCACTCACACCAGTCCTATGCCGGGCAGCTACTGCGATGAGGAGATTGCAGATG
ATCTTGATGAGGTGGATGACGAGGAGACTGCATCGCTGTCGGGATTCGACCCGTTGCCGCCAAGGATGCTCGGACTGGACTCGCAA
TGCTCTACAATCATGAACCGCAGCATCGAGCAGGGCTATCGCAGTAACAGTATTGACGATTCCAGCACAATATCAGGCTCGACAAC
TCCACAGACGTATTGCCAGAGCAACAGCAGCAGCAAATTCCAGTTGCCAATGAGCAGAAACAATAAAGATAAGCCCTATGAATCAA
ATGCGGCGCCCAGTAAGAAACGTTTGCAGCCAGTAAACAAAGGGAAGTCAAAGACTGCAGGTGGGAAATCATCATCCTTGAAACGG
AACAGCACACTCAACAGCACCAAGAGATCTCGGAGGGTTAAATCAAGGAAAATTAATGTGTGTGAATCTAAAAATAACGAGGAAAA
AGACGAGCATCCGAGTACTGTAGTTCTGTGTCGAGCTGATGATCTGTTTGTGCTGGAGCAAGATATCTGTGTGTCGTGTGGTTCAT
TCGGGATCGACACGCCGTTGCTGTGCTGCGTGCAATGCGGCCAATGCTATCATTCATATTGCGCTGAGATTAGCAAGCTCACAAAA
GTCATCATAGAAAAAGGCTGGAGGTGCCTAGATTGCACGGTCTGTGAAGGTTGCGGTCAGGCAACTAATGAGGACTTGTTGCTATT
GTGCGATGATTGCGACATCAGCTACCACACGTTTTGCCTCGATCCCCCATTGAATGAAGTGCCGAAAGGCAGCTGGAAATGCCAGA
ATTGTGTCGTGTGCCAGTTCTGTGGGAAACGGGACCCGGGGCCCAACGGGACCTGGCAAAATAACTACACAACTTGCGCGCCGTGC
GCTTCCAAGGCCCAATGTCCCGTCTGTTATAATAATTACAATGATGGGGACTTGATGCTAATTTGCGATGTGTGTGAGAGATCGTC
ACATGCCACTTGTGATCATTTGAATACTGAAGCGGAGTTAGAGTTCATCTTTGAAGGCGGAGGATATCATTGCATTCTGTGCCGGG
ACAAAGGGATCGATATCGGTGCTGGTCACAAGCAGCTGCTGCATTTACGTGCTACTCATGGCGATTCAGTTGCTCAAACATTGGTC
GCTGATAAATCTGTATTTAGCAATCTAATTGGCAACTCATTGGAATACGGAAGCAAGAACTCGTTGCTATCCAGTGTTGGATCATC
TGTAGACGGGCCTACGTCACCGGATCCTTCCAAGAACTTCTTCATGGACAATGTCGTTCTCAGTGAGAGTGGCCTCAACACCATTA
GAAACGCCATGCTCAAAAGCCAGCCACGTCGGGCTCAAAATGCCATGAAACGAGTACAGAAATCGCATTCCCAAGATACGATTGGA
GATGTCGACAGCAAGCAGATCAAGTTGGAATCCACCGAGCCCCTCCTGAATGCCGACGGCTCTGTTGTTGCTAGTGGCAACCATCT
TGACGAGGACTCAATGCAGAGCATCGAAGGAGAAATCAACACCACAGACATGAAGAGTCCAAGCGTCCAATCAGATGACGGATCTG
AAATACAGAGTTTGCCAGCTCATTCTGCCGCAAACAGCGTCAATAATTATCTGCCCTCCTCGAAAACCATTGGGAAAGCAGTTATG
CCGCGGAACAGAAGGTTCTCAAATTTGGGCATCGGCGGGTTCAGGGCCCGATCGACAAGGGTGTCACAAGTGAAGAAGTCGGCATT
GGCTGCTGTAAATAACCAGAATTTCGATCAAAACAACCCGCAATACACAATAGAAAATTTCACAGCGAAGCAGGGCCGACGGAGAC
CCGTGAAGAAGAAATCTGAAATTGATGAGAACTTTCCCGATTATTTAATGCAGGCTTTTCTTGGTAGCAGCCTTCTGTCGCACACA
ACTGAGACGAGCAAGAAATCCGGCCACCTCAAATTGTCACCGAGCAGAATGGCTGGCGGCAATTTCTCGGTTGATGCTCATTCATT
CAGCAGCACCGGCTCCAATCTGAATGACGATGGCGATGACTGCTGTCCTGACAACAGCAACTGCTTTGAAATGGACAACGAAGCCG
AGAATTTCACCGCAGCTCAGGAGGCATCTAATTATATGCTAATTGAGGTGCCGATCTCCTACACAAAAACTGAATATGACGAACCA
AAGACAGAACCGACGACAAATGTGATGTCAACTTGCTCCTCTGTGTCCCCCAAGGTTACCAACATCAGCAATCACGGAATCCTCCA
ATCAGAATTGGGCGCTAGATTTAATCACCAAATGAATATTCTGACCAATCCCGGTCAGAGCCCCAATAAGCCCGCCGTTGGTATCA
AGAAATCGGAAGAAACAGTTCATGGAACGATCAATAGTGATTTATTCTCTCAAGTGTCTAGTAACAACAACAGCGGCAGCAACAAC
AATAGCAACAATAATAACAGTTGTGTCGTATCGTTGGCCTATCAGAACCAAACATTGACATTTGATGAGGATGACGAATATTTATT
GGATGATTATCTGCTTTCTTTTCCCGATTTAGAAATGGTGGAGGAGGGCAATCCCGACACGGGAAACACTGGTGCAGAGATCAATA
ATTCCACCGAGATCCACTCCAATCCCCTGGAACTGTCAAAGCACATATCTGACGAGGCCGGGCTTACCCAGCTGCCAGAGCTTACA
GTAAATGACATAGAGACTCACCTGGGGCCGTCAAATATGGAGAACAGTGAGTTGCCTATGACATCGCGACGTGGCAACGGTGCTCG
TGACAATAGCGGCATTCAGCTCAGCAGCAATGCCAATATCGATTCCAGCTATCTGATGTCGTCATCAACGAATACGGCAATTTCCG
CCAGTCAGCAGCTCCCTGACTGTAATAACCAGTCTCCAATGCGCTCGGAACTGGTGATGCGTCCTATGCACAACACTGCCGGAGGT
ACTGGGGGAGGAGGCCAATTGGTGCAGATGAAATCGCCCACTGCCATGCAGTCCAAGGCCCCTCCAATGAATGTCCAAATCAGCGG
AGATCCTAATGCGCTCGGGATCGATATTTCTGCAAATTTTATCCCAAACTCACAGCTCATTAATCCCAGCATGAGGCGCCAGCTAT
TTCCCAATCAAATTCCCCGTAGACATACAGAAGTTTCTTTTCAGCTCCAGCAAGTCGGCAACATTTCATCAGTCAGTGCTAATCAA
GCTAGTTTTTCCTCCCCAATTAGGAGCTATGATAATCAATCATTGCTGCAGCAAAAGTTACAATTCTCCCAATCAGCTCCTCAATT
CATAATGTGCTCCAATAATCTGCAAAAGCCTTCTCGCCAAATAAATATTGACCAAGGTTCAGGATTAGTTCAAGATATAAATGGCC
ATCAAATTATGCTCAATGAAAATCTTGGTCTTGTTACTATTGACAACCAGCAAAAGCAGCAGTCACTGGAAAGCTTTCAAGGACAG
CCATCCAATCAGCAGACACAGCAACAGCATATGATACTGGGCCGGTTGCATGATCCCAATGTGACTGTGCCAGTCCAGTGTAACCA
AGCTCCCCCGCCGCCGCCGTATCCAAGCAACTCTCAACCTTCGGAAGGTATCGGAAAGCCATCGAATGATAGCAAACAGAAAGTCT
GGATACATTCACCTAATTTCCAGCTGCAGATGAATGAGCAGAGCATTCTAGAGAAGCAATCAATGATTCCAGCCGATTCAATCCAG
TCAGGAGGACCCAATATGTTCAAGTCAGCTCAGCAAAGCCCCCAATCAGGCATCCATCAGAGCCCCCAGAAGCATTTCCCAGTATC
AATGCAGCAGTCGCCCTTCAATCCCGGCAACCAGCAATTCAACAAATCAAATCCCCAGATCATGATTCATGTCGGTCAGATGGAAT
TGCCTGAGAAGTCCAATAAAGATTCGATGCTCTCACCAGATACAAATTTGCCAGTTCCGCGCCCAAAAAATATATCGTTATCAGCA
GCGAATTTGGAGAATTCGGAAGTGAATTTCACTGTCGATGGAAGTAGCGGCTGTGGAGGAACCACGCCAACACCGAGCACTCCAAC
GGCACCACAGTCCAATAAACGGATCAATAATTTGAAATGGCAGGAGGACGAACACCTGGGTGACAAATCAACGATCGCACCGGTCC
TGTTTGCGAATTGTGTCCATCAGAATATCAGATCGCAAATTGCTGATCCAGAGTCTCGATTCAGAGAGATACAGAAGTTGTGGCGG
AAATTGAAAAGCGAAGAGCGACAGGAATGGGTGAACAAATCCCGGCTCAATAAGTCCACCAAGAAGTGCAACGTTTCTTCCGCTTC
CGGATCAGCCAATTTATCTAAAAAGAAAAAGACAACTCCTGCCATGGCAACCGCATTGCAATCAATGCAATTGAGTAGTCCCAATA
AATATGTGCAGCAGCAATCGCCCAGCTCCATTCATCAATCGCCGTCTCCTTTTCAACAGGGAATAATCAATAGTGCTCTCCAGTCG
CCACCAATGCCTCAAAAACACAACCAATTCCCGCAATTTATACAACAGCAACAGCAAATGGGAATCAATTCGGATATGATCAGCAC
CAGCCTTCCAGGAGTGCCAAAAATCAATCAAATCAATTCATTGCAAATTGGCCAAATGTCATATAATAAATATGAATTTATCCCCC
CTCCTCCGTCTCAGCAGGAGCAGCAGCAAGCCATTTATCAGCAAGAATTCAATAGCCAGCAGCAGCAGCAGCAGCCATCGATCCAA
ATGCAAGTCCATCCTCAGCAGCAGCAGCAGCAGCAAGCAATGATGGTCAGTGCTCAGCAGCAGCAGCAGGGCCTGGACTCCTTGCG
GATGCAGCAGCAGCAGCTGCAGCAGTCATTGAACGCCTCCACTCACTCAGTGGCCAATCCTGGCACTCCCTCTATGATCGATAGCA
ATGAGGGTGCTATCAGCTCCCCCAGCAAGCAGCAGTACTGTGAAGGCAGCAACATGCAAATGTCCAATCAGCAGCAGCAGCAGCAG
CAATACCTTCAACCGCTACCGCCACCAACATGGCCACCAGCGGCTGACCTGTCGTCGTCGGCGGCGGCGGCGGCGGCCGCCAT
GCAAGCTCCCCAATCGATCAATGCGTTCCACCGGTTCAACAAGAATCAGACTGGCAATTCCAACCAGATGGTATCGAGCGTGCTGC
AAATGTCAGCACAGCACCAAATGCCCTCCACATTATTATCCCAGTCCGTTCAGAATCCCGCCGACAACAAACTCAACAACAACAAC
TCAAATAGGGAGAAACTCATGTCCACTAACATAATCAATCAATCGAACTCGGAAATGATATTCAGTCCTCCCGGTCAGTCAAGCGC
TTCTCTGCCGTCGTCCCCGTTCCACAAGGAGATGGCTACCAGTGACGAGACAACCACCAAGTTGCGGGACGTGTTGAACAAGCGCA
ACAGCAACAGCAACAGCAACAGCAAGCCAAATTCATTCAGTCAGCAAACGTATGCCAATTCTCCATCAATCAGTGCTGCGGACAGT
AACTTGCTGGCCGCCAGTCCCTCCTCATCTGAGCACGGTTTCAATCACACATCTATCCAGTTGCGCTCATTGGCAGGCGGTAATAA
TAATAATACTGCTGCGGTTTCCATGGCTGCTACCTCCGGTATGGAGTCGCTCCCTGATCACGAGATGTGCATCAATAACAACTCAT
CATCATTCCAGTCCATGAACAGGCTTGGGAAGATGCAACAGTCACAGCAAATCAATCAAACGGCTCTCCAGATGGTCCCAGTTGCG
TTTTCACAGTCGCACGGAATGGAGGATGCAGTCAATGGGAACATGCAATCACAAATGATGCAGACGCGCCAATATCAAATGCCATC
TGGCTGCTACATGGATCCACAGACCCAGCAGAAGTATTCCATTAATCCGTCGTCACAGCAACAGTATCTGATCCTGGAACCGTCCG

ATGTGAATGCGAAACTGATTCCACAACAACAGCAACAACAACAACAACAACAACAACAGCGAATGAGACAAGTATCATCGGGTCAC
CAGCAGAATGCGGCGGTGGCGGCGGCGGCATTAATGGATCAGGGCTGCAGATCGGCATCTCACATGATGCCGATCAATATTGTGCC
TGGCAACAGCTATCTGGGCCAGAGCAAATTTGTCCCCGGCCAACAGAAATATCAGCAGCAGCAACAGGATATGGTGCAAAATCGCA
ATCCTTACAGACAGAATACATTGAATCCACAATCAGCCAATAGTGGCAGCGGCAATAGTAGCAGTAACCAATTGGATATCTCTGGC
ATGTCTGGCATGTCCGGCATGCCCGGCTCCCACTCTAATTCCATCGGGGAGGAGTGTGTGCAAATGTCTGCCGATCCCAATGCCAG
CAACATTATCCACATGGACACGAGAATGTCACCGAATATCATCCAGCACCAGTCGCCCATTTCCCCGTTCACTCACATGTCACAAT
CGCAACTACCACCACACAAGTTCCAGCAGCCCTCTAGGCCGATATCGACTGGTCTCAACAAATCATTTGACCTATCAGGAGGTTGC
CAAATGTACATTGCAGCCCATTCACCACGCGACAGCCACACTCCCCAGAGCTCTGCTCCATCCAACCAGCCAATGCAGAACCCGTC
TGGCCCTGGCACCCCTCAGAGCCCCTGTGCCGCCGGTGCCAACAAGACATCAGGGCCTCCATCACCGTCATCCAAGGTTTCCAACA
TGGCAGCCGCCATCGGTACAAATAACAGGACCAATCAACAACAACAACAACAACAACAACAGTCCCTCAGTCCTAGTACAGTCCCA
AGCATGTCTTCACCTGTCCCGTTGATGCAGTCAAGTCCTCATTTCCAGCATCAGATGACGCCGGATTACGAGTCGGGTGTACATGC
GGCCCCATTTATGCCCGGTAAATCGATCACATGTCCTCTGACCCCGGGTAATAGAGAGGAAATTGGGGGCTTTGAACCGCCAGTTA
TTGTCACTCGTGTTAATGTCGATAGGAAGAATTTCGGCCCATCTGGCGTCGCGGTGGGATCTGGAGATGGCAATGTTGTGAATAAT
CAGGATCAATATCGCCAGCAGATCATTATGACCAGTAATGATGGGATAAACTGTGTTAGTGGCACTGCTATTACAGCTGATATGTG
CTCTTCAGCTTCCTACCAATCCAATGCTACCAATCAGGTGACAATGAGCCATTACAAGCCACAATCTACCAACGCTGGAAATATTA
TACAGCAATCATCAGCTGGCAATAACTCTCGCTCGTTCTCCCGCTCAAACAGTGTCAATCATGACGACCCCAGGGCAATGAACACA
ATTCAATCTAACAGTTTTGCTTACCGCCCACTCAGTAGTGCTAATATCACCAACAGCAACAGTACTATTGTCAACATCAACATCCA
GTCTCAGCCGTCACATTTGGATGACGTCATTGCCAGTGTCGTCTCTGATTCGATAGCTTATATAGACGAGAGTCCTGCTACACAAT
CAATCGCGTTACAGAGACAGAATTCAAATGACTCTTCGTCCTGTCTTGGCGGCGGCGGCGGCTTTGATCATGATAAATCGGCGGTG
GCGGCGGCGGCAGTAGCGTCATCACTAATGCCCGGTAGCGGAGATCCGGCAGCGATTGAATCCCCACTCATTGCCGCGACTCAATA
CCAGACGGCTAATAATAATAACAACAATCGCTCAATCTCCAGGAATCTGAGCAGCAACCAATTGATATTCAATTCAGCAAACGCTT
CCAATCCGCTCAGCATCAACCAGGCCAATCAATTATTGAATACCAAACAGTCGCAGAACAGCTTTTGGTTGGCTACATCATCCAAT
ATAAATCCCGGCGCAAAGCAACAACAGCAAATGTGTATTTCAAATACGACTCAGCTGATTGGCCAGTTTGCTAATAATCAACCGCA
AACCAAAATTATCCAACATGGGCAAAATTTATATCAGATGTATGCTGCCAATAATTCTCCTCAGCAGATCATTTACACTCCCTCAT
CGATTGAGCTCAGTCCTAATATGAGCAGTAGCAGTAGCAGTACTGACCAGCAGCAAATCAATATGCTCACAGTTACCTCGCTGCCT
AATCGGCAGAACAGTCAAACCTACCAATCAGTTGCCGCGTTCAATCAGCCAATCAGCCAATACCAGAACGCAATTGAGCTAGGACA
GTCTAACAAGAATCAACTGGAATCAGAAAAGAATAGTTTGATCGTTCAGTCCGCTGACCAAGTGATGAGGCAATCACCAAGACAAC
AACAACAGCAGCAGCCTAGCAACAGTAGTAGCAACGCGACCATGATCAAATTGAATCAGCATTTCTCTGCCGGAATGCAATCAATG
CAATCAAATAATAAGGCTAGTAGCATCCAGATCCCTCAGATCTCCTCAAATCAATTTATCTATTCTCCCGGCCCGACACTGCACGG
CGCACAGAACCAGCAGGGCGTTGCCATGGTCGCCCAGGCCAACAACCAATTGAATGCAGCGGCGGCCACCGGCAGCAGCTTGATTA
TGGGGCCGCCGGTGTCCAATCTCCGTTCCAACGCAAATGTCGTCACTCAGAGATTCATTTGCTCAATAAACTCCAATTCCACCGCC
AACACGATTCCTGGCCAATCTCCTCAGCAGTATCTGATCAGCAGCACTCCACAGTCCAGTGCCAGTGTCGGTGGTGCTGTCTGTAC
TGCTGGCTCACAAGCGCAGCAGCAGATTTCTTCCTTGATGATTTCATCTAATAATAGCAACAGTAGCAATTCTATGAATACTACTC
AACAGCAGCAGCAGGATTTGTCCATGAATTTTGGGCATTACAATTAGTAGGCCGTATTTTTGCCTAGCAACCGTTGCTATGGTATT
GTGCGCTGCCCTTATGCTGCATAACAACAATAATAATAATAATGTCATACATATTGTACATATATATTATCTCAATTGCTCCAGCT
GATTCCCGATTGCCCTTAGATACAGTTACACTATTAATTATAATTCCCACCGCTAATGTTGCTATGTGGCGATGGCGCATTCACTT
AGCCCGCTTATTGTGATTATTTATCCATTTAATTATCATAACCATAATTATATCTCTTTCTGTCATTTTATTCAATTTGTGGTTAG
TCGCACAATAAGCGGGCTAAGTGAA

8.6. APPENDIX F

Protein sequences used for the phylogenetic analyses in Chapter 2.

>MLL3 (*Homo sapiens*)
MSSEEDKSVEQPQPPPPPPEEPGAPAPSPAAADKRPRGRPRKDGASPFQRARKKPRSRGKTAVEDEDSMDGLETTETETIVETEIK
EQSAEEDAEAEVDNSKQLIPTLQRSVSEESANSLVSVGVEAKISEQLCAFCYCGEKSSLGQGDLKQFRITPGFILPWRNQPSNKKD
IDDNSNGTYEKMQNSAPRKQRGQRKERSPQQNIVSCVSVSTQTASDDQAGKLWDELSLVGLPDAIDIQALFDSTGTCWAHHRCVEW
SLGVCQMEEPLLVNVDKAVVSGSTERCAFCKHLGATIKCCEEKCTQMYHYPCAAGAGTFQDFSHIFLLCPEHIDQAPERSKEDANC
AVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKRAGWQCPECKVCQNCKQSGEDSKMLVCDTCDKGYHTFCLQPVMKSVPTNGW
KCKNCRICIECGTRSSSQWHHNCLICDNCYQQQDNLCPFCGKCYHPELQKDMLHCNMCKRWVHLECDKPTDHELDTQLKEEYICMY
CKHLGAEMDRLQPGEEVEIAELTTDYNNEMEVEGPEDQMVFSEQAANKDVNGQESTPGIVPDAVQVHTEEQQKSHPSESLDTDSLL
IAVSSQHTVNTELEKQISNEVDSEDLKMSSEVKHICGEDQIEDKMEVTENIEVVTHQITVQQEQLQLLEEPETVVSREESRPPKLV
MESVTLPLETLVSPHEESISLCPEEQLVIERLQGEKEQKENSELSTGLMDSEMTPTIEGCVKDVSYQGGKSIKLSSETESSFSSSA
DISKADVSSSPTPSSDLPSHDMLHNYPSALSSSAGNIMPTTYISVTPKIGMGKPAITKRKFSPGRPRSKQGAWSTHNTVSPPSWSP
DISEGREIFKPRQLPGSAIWSIKVGRGSGFPGKRRPRGAGLSGRGGRGRSKLKSGIGAVVLPGVSTADISSNKDDEENSMHNTVVL
FSSSDKFTLNQDMCVVCGSFGQGAEGRLLACSQCGQCYHPYCVSIKITKVVLSKGWRCLECTVCEACGKATDPGRLLLCDDCDISY
HTYCLDPPLQTVPKGGWKCKWCVWCRHCGATSAGLRCEWQNNYTQCAPCASLSSCPVCYRNYREEDLILQCRQCDRWMHAVCQNLN
TEEEVENVADIGFDCSMCRPYMPASNVPSSDCCESSLVAQIVTKVKELDPPKTYTQDGVCLTESGMTQLQSLTVTVPRRKRSKPKL
KLKIINQNSVAVLQTPPDIQSEHSRDGEMDDSREGELMDCDGKSESSPEREAVDDETKGVEGTDGVKKRKRKPYRPGIGGFMVRQR
SRTGQGKTKRSVIRKDSSGSISEQLPCRDDGWSEQLPDTLVDESVSVTESTEKIKKRYRKRKNKLEETFPAYLQEAFFGKDLLDTS
RQSKISLDNLSEDGAQLLYKTNMNTGFLDPSLDPLLSSSSAPTKSGTHGPADDPLADISEVLNTDDDILGIISDDLAKSVDHSDIG
PVTDDPSSLPQPNVNQSSRPLSEEQLDGILSPELDKMVTDGAILGKLYKIPELGGKDVEDLFTAVLSPANTQPTPLPQPPPPTQLL
PIHNQDAFSRMPLMNGLIGSSPHLPHNSLPPGSGLGTFSAIAQSSYPDARDKNSAFNPMASDPNNSWTSSAPTVEGENDTMSNAQR
STLKWEKEEALGEMATVAPVLYTNINFPNLKEEFPDWTTRVKQIAKLWRKASSQERAPYVQKARDNRAALRINKVQMSNDSMKRQQ
QQDSIDPSSRIDSELFKDPLKQRESEHEQEWKFRQQMRQKSKQQAKIEATQKLEQVKNEQQQQQQQQFGSQHLLVQSGSDTPSSGI
QSPLTPQPGNGNMSPAQSFHKELFTKQPPSTPTSTSSDDVFVKPQAPPPPAPSRIPIQDSLSQAQTSQPPSPQVFSPGSSNSRPP
SPMDPYAKMVGTPRPPPVGHSFSRRNSAAPVENCTPLSSVSRPLQMNETTANRPSPVRDLCSSSTTNNDPYAKPPDTPRPVMTDQF
PKSLGLSRSPVVSEQTAKGPIAAGTSDHFTKPSPRADVFQRQRIPDSYARPLLTPAPLDSGPGPFKTPMQPPPSSQDPYGSVSQAS
RRLSVDPYERPALTPRPIDNFSHNQSNDPYSQPPLTPHPAVNESFAHPSRAFSQPGTISRPTSQDPYSQPPGTPRPVVDSYSQSSG

TARSNTDPYSQPPGTPRPTTVDPYSQQPQTPRPSTQTDLFVTPVTNQRHSDPYAHPPGTPRPGISVPYSQPPATPRPRISEGFTRS
SMTRPVLMPNQDPFLQAAQNRGPALPGPLVRPPDTCSQTPRPPGPGLSDTFSRVSPSAARDPYDQSPMTPRSQSDSFGTSQTAHDV
ADQPRPGSEGSFCASSNSPMHSQGQQFSGVSQLPGPVPTSGVTDTQNTVNMAQADTEKLRQRQKLREIILQQQQQKKIAGRQEKGS
QDSPAVPHPGPLQHWQPENVNQAFTRPPPPYPGNIRSPVAPPLGPRYAVFPKDQRGPYPPDVASMGMRPHGFRFGFPGGSHGTMPS
QERFLVPPQQIQGSGVSPQLRRSVSVDMPRPLNNSQMNNPVGLPQHFSPQSLPVQQHNILGQAYIELRHRAPDGRQRLPFSAPPGS
VVEASSNLRHGNFIPRPDFPGPRHTDPMRRPPQGLPNQLPVHPDLEQVPPSQQEQGHSVHSSSMVMRTLNHPLGGEFSEAPLSTSV
PSETTSDNLQITTQPSDGLEEKLDSDDPSVKELDVKDLEGVEVKDLDDEDLENLNLDTEDGKVVELDTLDNLETNDPNLDDLLRSG
EFDIIAYTDPELDMGDKKSMFNEELDLPIDDKLDNQCVSVEPKKKEQENKTLVLSDKHSPQKKSTVTNEVKTEVLSPNSKVESKCE
TEKNDENKDNVDTPCSQASAHSDLNDGEKTSLHPCDPDLFEKRTNRETAGPSANVIQASTQLPAQDVINSCGITGSTPVLSSLLAN
EKSDNSDIRPSGSPPPPTLPASPSNHVSSLPPFIAPPGRVLDNAMNSNVTVVSRVNHVFSQGVQVNPGLIPGQSTVNHSLGTGKPA
TQTGPQTSQSGTSSMSGPQQLMIPQTLAQQNRERPLLLEEQPLLLQDLLDQERQEQQQQRQMQAMIRQRSEPFFPNIDFDAITDPI
MKAKMVALKGINKVMAQNNLGMPPMVMSRFPFMGQVVTGTQNSEGQNLGPQAIPQDGSITHQISRPNPPNFGPGFVNDSQRKQYEE
WLQETQQLLQMQQKYLEEQIGAHRKSKKALSAKQRTAKKAGREFPEEDAEQLKHVTEQQSMVQKQLEQIRKQQKEHAELIEDYRIK
QQQQCAMAPPTMMPSVQPQPPLIPGATPPTMSQPTFPMVPQQLQHQQHTTVISGHTSPVRMPSLPGWQPNSAPAHLPLNPPRIQPP
IAQLPIKTCTPAPGTVSNANPQSGPPPRVEFDDNNPFSESFQERERKERLREQQERQRIQLMQEVDRQRALQQRMEMEQHGMVGSE
ISSSRTSVSQIPFYSSDLPCDFMQPLGPLQQSPQHQQQMGQVLQQQNIQQGSINSPSTQTFMQTNERRQVGPPSFVPDSPSIPVGS
PNFSSVKQGHGNLSGTSFQQSPVRPSFTPALPAAPPVANSSLPCGQDSTITHGHSYPGSTQSLIQLYSDIIPEEKGKKKRTRKKKR
DDDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSELPQQADQESVEPVGPSTPNMAAGQLCTELENKLPNSDFSQATPNQQT
YANSEVDKLSMETPAKTEEIKLEKAETESCPGQEEPKLEEQNGSKVEGNAVACPVSSAQSPPHSAGAPAAKGDSGNELLKHLLKNK
KSSSLLNQKPEGSICSEDDCTKDNKLVEKQNPAEGLQTLGAQMQGGFGCGNQLPKTDGGSETKKQRSKRTQRTGEKAAPRSKKRKK
DEEEKQAMYSSTDTFTHLKQQNNLSNPPTPPASLPPTPPPMACQKMANGFATTEELAGKAGVLVSHEVTKTLGPKPFQLPFRPQDD
LLARALAQGPKTVDVPASLPTPPHNNQEELRIQDHCGDRDTPDSFVPSSSPESVVGVEVSRYPDLSLVKEEPPEPVPSPIIPILPS
TAGKSSESRRNDIKTEPGTLYFASPFGPSPNGPRSGLISVAITLHPTAAENISSVVAAFSDLLHVRIPNSYEVSSAPDVPSMGLVS
SHRINPGLEYRQHLLLRGPPPGSANPPRLVSSYRLKQPNVPFPPTSNGLSGYKDSSHGIAESAALRPQWCCHCKVVILGSGVRKSF
KDLTLLNKDSRESTKRVEKDIVFCSNNCFILYSSTAQAKNSENKESIPSLPQSPMRETPSKAFHQYSNNISTLDVHCLPQLPEKAS
PPASPPIAFPPAFEAAQVEAKPDELKVTVKLKPRLRAVHGGFEDCRPLNKKWRGMKWKKWSIHIVIPKGTFKPPCEDEIDEFLKKL
GTSLKPDPVPKDYRKCCFCHEEGDGLTDGPARLLNLDLDLWVHLNCALWSTEVYETQAGALINVELALRRGLQMKCVFCHKTGATS
GCHRFRCTNIYHFTCAIKAQCMFFKDKTMLCPMHKPKGIHEQELSYFAVFRRVYVQRDEVRQIASIVQRGERDHTFRVGSLIFHTI
GQLLPQQMQAFHSPKALFPVGYEASRLYWSTRYANRRCRYLCSIEEKDGRPVFVIRIVEQGHEDLVLSDISPKGVWDKILEPVACV
RKKSEMLQLFPAYLKGEDLFGLTVSAVARIAESLPGVEACENYTFRYGRNPLMELPLAVNPTGCARSEPKMSAHVKRFVLRPHTLN
STSTSKSFQSTVTGELNAPYSKQFVHSKSSQYRKMKTEWKSNVYLARSRIQGLGLYAARDIEKHTMVIEYIGTIIRNEVANRKEKL
YESQNRGVYMFRMDNDHVIDATLTGGPARYINHSCAPNCVAEVVTFERGHKIIISSSRRIQKGEELCYDYKFDFEDDQHKIPCHCG
AVNCRKWMN

>MLL3 (*Strongylocentrotus purpuratus*)
NHRKKAIIIADRKLTKRCVLDRRFWGSINAQSFVNAFQWGQRSMRSAGLRIQLQSQNHAWRSQRRRRKRLGHKPESEVSMDVESIP
QRRGRGRPKRETFNVNERRKGLRVEGGSEEEPTDLAITMDTTLSNLEVEPSSNKLNLDNESPEEEEEEEPEDDDPGDMSHMIDHSE
SQLVERTCSDQSLVKSLDSYSAATDSGVTSPELSHRGVSPGHTGGYGGSLMDENQYDDDVMDCNNGSSDPSCSLCCCGERSLLGQG
DLLRFDSTPGFNPFRKPLPRSRRNYEEDAPYNPNESRSQKHLTWRRARGPPKAHKHGLGDRSKSPRRSNPDSIKDPSLDGLDELSH
VGHADEPDLQTLFELAGHTTAHHCCAAWSDGVCQNENFQLLNVDKAVFSGITQRCSYCHRFGATIFCVEQGCNRVYHYPCAASSGS
FQGIKSLILLCPEHLDLAEERVEMEELECAICDLPGNLSESLFCTSCGQHYHGSCLDPPVSIDPVVRAGWQCPNCKICQTCRQPGD
DNKMLVCDTCDKGYHTFCLKPAMITIPKNGWKCKTCRVCTDCGARTPGNGPSSRWHHNYTVCDSCYQQRNKGYCCPICGKAYRHHT
THKVMVQCHLCNRYVHADCDDRTVISKYQQSKAAGQPTPYKCPDCRHRPNRGLELERRRSASPFEDGRRSPFASSSRTPPHLPIPD
DTSQDLSLSASSVQEDSLSLSSIDTDYTMASDDRRPELQQEFSASLSGSSGLSMTIASVTRHHGGGKPLDRLSQLERMGRRRFSGR
PRGRGSSSYAAVGRRRRQGGGGNVGGRRGPKPKIKIAAIQPASPPPEDELELIDPVVVEKPSAKEKDEEDTSMHSTVVLFSVDDKF
TATQDMCLSCGSFGLGSEGRLLTCSQCGQCYHPYCVSIKITKVVLSKGWRCLDCTVCEGCGKSSDEARLLLCDDCDISYHTYCLDP
PLQTVPKGGWKCKWCVCCTHCGSVTPGENADWMNNYTQCGPCASMTHCAYCYRSYRDNELLCQCSHCQRWEHALCNSLYTEDETER
AMDKGFICTLCRPFLPEPEPQLPEPSTEEPPQRIPKLAIKEPELHNKEFSVEGVRLTETGFKQISSMAMARTHKKSKKVRSSLKMG
SLAILRSLADPHSDISKIADQEYDPEGIERVDPEKWRGFSHGEKNDSSVGEGKHHHHHHHHHHDKKKSKHRSSLRPGIGGFIAKPR
GFHSHKENRIAAMSRSMSEGASDTMDSIRDQSELLQRTHSVGEEGHYEKSKKRKRAHKRSQLENQFPSYLQEAFFGKMLLNSSKDQ
PTMNTRLDISDEDSKKHSSSHHQSKGTEIIKVSSAEHKSSLSAGSSRQTHHNQADVHRESAATVDGAAGVLGSGEDPLNFLPQDEL
LGMLIGDGLGKNGENLDIDSIDAEATHSSTTGQNGHTTQASSSSDHLNELDSLLSYNLHIVPEDLAHIDSKEVEDLFSGVLSPEGH
HHAGSTNTPTSGVDGNATNTATPSTVSQQQHQHHHHHHQQQQLQKQPLPHLPLTQQLVQAASVGHSVASPAFSPDFAEPPSPWPGG
GGGGETDGETLSYNQRNIKKWEKDEPLGNQATISPVLYANINYPDLKKDFPDWPSRAKQIAKLWRKAGSGNKKPFLQLARENRAAQ
RIKEAQKIHQQKRQKFDGDLMSDGQDGSGIRPSQQQQLQQGQQPSTPIPQAQRSPKSMPGTPIPGTPMSPATEMPHSSLMNANSQA
LSGHSGLPVHMPLSNQSSGAVNMDSMPPSPMVAQPSNIHLQSGRSGTPSMIPGRSPHHSEEISHGVMGGVNTHQQHLASLIAGQGA
PPTSSSDHLPVSNNQSIIDRLQSPGMPHGSASQVPPGIGNMSPAHAMSVGSPAHVMSVGSPAHAMPVDMPIHSPQQAPFMKNPVFP
DGASNQSLSPAASAAPADRVGMNDGNKHAQKQYIRQLLVRQQEKKRMRREQQQMQNQNWSGESHQRPPPPYPANAPSDQISPATNN
EQMNSFPGTPGMPRPEAPPSFVNQRPQEQFNMTLEQTDEAWRQGQFATGLQRRPSDGGNRIPGPGQFIELRRFEQGARMPVSGPTL
SPRPRNIALLQEKLHQRALQQRTALVGSNPLDPYDHMVRQQAHSRTPTPRQVMTHPNFHGRSPMMSPIVDHPAQQPHLPFQQPTPP
FQGQNTNAIQNHQPRIHNAMPAQQPGTLQSDLHLPPGQDPAMAPQQEDHDAQLEDLLSGELDLIDFAGTDLDTNKHSAENQQLFKD
FGLSDEALDEMTTSSNTPKDSNSSKIENGIPSAAPENNEVQSTSTAGASLEKESDKVPEDKPAANEPKQTDSSGLPADPNKKEETS
SSGRKDAPKTGRDISATVSSNTSEESSSKGLMNHDDKSATQSQGTNISSDKGSKEKTVVHTDITNTSASVPTVPLSTSASVLQTST
SVPVSASPSEALTTSTTFLGKATIKIEATENSQSICQRSKVLTAQSGQLTTTIIGSQLVEKPQSHSQGLQDKASPGTSSVETKTTL
SSSVSAPRGGVPTIDQIAAILPGPASNIQSNSTSQLPVASVASQPTTSDGTQLQTSQSSGSAHTNSLQKRQRLLHEQPLLLQDLLE
KEKREQQKLQQQSGNVEAAISGKVDPKSKTQGSDPFLDNLDLERLQAEAKMIAKETEDFTIKPPGPHAQLLLAMNRQYHVQMQQRH
LTEQQPGANVPLIPPASPGIRQPRTPQGVDFGSIDSVKDPIRSNSTDLFKQGQTVPPQGMQGMSHMNAHGQEMGPFSPGHPQMPNF
SPGPRFSAHENIQTHPSPGPVSPMSPMTQNHFGMRPRGPMVVQNSGIVGRPPMPGSGMHMQGGLPGDQMPMQMHGMQRLPGPGQRM
GMPVMNRMPGPPQVSNFRFPGRFMSPNMQRMPSPFFNVGMDGMSHAMMNFDPYDTTEEKAPEDTSTPEYEEWIFNKDRKLQLRLKF
FENEVGKERKKKKALTSRQRACRKNNKVFPESDRLELEAVQQRHSKQQKFLENVRKQQRQHAILLQEFRMKQQQKLSSNHSPLGHH
NPMMQGMMSPQGMMMQGQGPNPGMHSSMIGGQQHTLIPPGGTSPNIPQQMLINQGMQSPSGSMMQGPQSGMMPQGQMMGHVSQQIP
HPMRSPMSGERPPFAVRSGQQSPMMHGNQIMQPGNAGMMVMPGQTSPMVSNAIMQGSNPMSQASMMPNNRQPSPLMPGSNPQTPMM
PGNKPPTSMIPINSPQTMGMPGGNQTNMLPGGMNMPRGTNPQQIPVQPSQQSPMSHPSHQVSIGASSQQQSPMSAQSPMVPQSPMV
PQSPMVPQSPMVPHSPMVPHSPMVPQSPMVPHSPMVPHSPMVPHSPMVPHSPMVPHSPMVPHSPMVPQSPMVHPGQQSPVVSSGQQ
QSPMMGHQSQQNSPAVQNSEHSPLNAGSHQSPAVTPGVPSPMVPPGAQSPMMPPGGQSPLKQAGQQSPMVLADQQSPMVSQGQSSG
PNSSHPSPMMPPPSPMMTHPGQQSPIASSANQGPTPSVSGQIQMRQAGTPMQSPLQSPISAKTLPQLSPIDKRMPINIGQDNNPFS
ETFQQRERMERLREQHERQRLQLAQELEHVRAMQRYEQQQQHQQQQQHQQLNTSHVFNPSRLPGPHQVQQQGNMGQFSQMPNPGMV
SPISQVPPVHNSSVGIRPPFPSPQAPLQSPVQSPFPREPLQPGMMSSQSSTPTTEGFPMPPSYPGMQTKVQKPKKKRAKKKTSDLF
NPNPTSVHQSPVGVLDESSVMAAIGHSQEKGGMVMSVQNILQLTSPGSATHQALQSIMANDPTASVDLAKLQQSQAMLMIGDINAK

AAKDGILKNNHNQLLKQLLQGPSICASGIVKSVPNKTPVQNQLIKVTGETSNAESIQTQQHYDSSVDMHDEEGPIRLTSEQQKQLE
IIEQLPYVKKDGKPAENDLQILLEIHRQGKIDYPPPQEPKEKVPAPPVIPASPQPTVKQHKHKEKKKHKEKKSKKNKESRFSSVEE
LMAALKQMPMLQLTEPEVGVNFSLCPLFGSGTTNGGTRLRGAFGHAILDGVPDFYDHLLTGSTSDTKLKLPLGVVNGPSSLKEIGS
KAQANTTTTTTTTSTITNTVQSSEVTVLTGDEVMREMQSHQRTITRTMNHAYSGGGGLAHLVAPSPSLSQVVADIVISSSPECEEN
EAQEPSKMKGLRLFTQDVDQVLAAQKRSGSPTIPIIKPTPRKAEAGSLGNANTVQVTLKLMPEVMHNVEGVLSAFSKLLGIKEPSN
FEVIDEAGQKERRGKWAQSTSPVADVALKISPEKISASPTSLTTSVKPTKITSKSSVLAGSLDDPKLHFCRHCDVLVVGEGVRKSA
TPTDKNVVFCSGACYLQHTISLRSQVARTEVNAGRVVNHRSSDPQSVQLDVLSPKASTSLPTSPCTVPKTPPQSKSTPISPRSLPA
QPVVSSPMAAQPAHTSAEGTTATPSSNQEVSPAKEEKMVTDADEMKVIISESKTGESGREELRLTLKTGTRRRPALGDELPLVKKW
KNIKWRRWHGSFSGGKKYRPPSDQDIGQLIKKLGTSMRTSETLRDARCCVLCNGFGDGDTNGSARLLNLNVDTWVHLNCALWSDEV
YETLNGSLMNVESALKRGRTCTCCYCSKLGATISCNKQRCCNVYHFSCAMQDHCMFFKDKTMMCPLHAPSKPTESTLQSHVVFRRV
YISRDEHKQVANIMTMADRKYTVRVGSLIFKSLGQLLPHQMHLFHSLTSIYPVGYESIRLYWSMRYTNKRCSYSCTIQDQGGRPEF
SIRIMEQGHEDIIFRADTPRNVWYKVLEPIAKLRRDADMLKVFPTFLTGEDLFGLTEAAIQRITESLPGVDLIQHYTFRFGRSPLI
DLPLAVNPTGCARAEPRRQTHLKRPHTLTSNKTSRGFHTTLAGEVASPYNKQFVHSKSSQYKRLKTEWKTNVYLARSQIQGLGLYA
AHDIEKHTMVIEYIGTLIRNEVANKWERDYEAANRGVYMFRIDDYTVVDATRSGNPARYINHSCNPNCVAEVVNFDKDQKKIIIIS
SRRLLKGEELTYDYKFEIENDQNKIPCLCKAPNCRKWMN

>SET-16 (*C. elegans*)
MICIECGGTAFNNTELCENCMDVNRSGTMSLDMARQCDMSSSCGSCQRMITADSRQCTTCRKSYHLSCDSSDVSSLNNFVCISCRR
RSQMLPDDVMLSASARPSRGQDAPLSPFSDPSVPQISLQMQPAPLPSSYPRSQDFNGDVYTQIAAMQQQRQQQQQNQVQQMLATQA
QLMDEISDGGGTNDSDRNSSPFYQEASDYDEDFVPASTRGRGRGGGAKKKPGRGGSTGTRRQTNPGPSGFFSATQLQAMSGDFSHL
HQPQQTSSTRGKRGKRGNGTSVPRQPGKGPGRGRGRGRGSAAAIAQLAQLQQVQYGAQGAPMFNGAPQLHPTQPTVAPTLPPSVTY
QNVAPTAPPAPPALQTPLPDPLPSIVPSFQQAPQAVAIREVDELMLEDESSRQSGGSRSDEPEYIRTAIVCRVNDEFLQKACMCLV
CGSIGKGPEGSMVACSNCAQTYHTYCVTLHDKLNSAVVGRGWRCLDCTVCEGCGTGGDEANLLLCDECDVSYHIYCMKPLLDKIPQ
GPWRCQWCSRCRRCNHKAASGNDLTSQGLCFPCASLRKCPRCERNYQLNEKLIRCSQCSKWQHGACEGLYTDEQLEQAAIDRMRCS
ACRPKRVQPSGFSDVDTVWCDYVALDKDAHEILKSKYTPSALKNHMMESVGYRESFDHYDDDYNPLEDSSDQVTPLPIPGQRGRGR
GNPSGRRGMNRIGIGGFYAKLPRHRIQALTEEAAAAAAASSEDSKQAKRPRKSRRSQLEDAYPPQIQEAFFGIKAVEGKVLVDTPVD
EPFLAEFNGWYKRTRSTKAFELCMTASEMLRTDIDENEFLDCVGLPNMETDALDDFDFNLLEDFDVDEENDELENSLLGAPGNTEM
KEDAAEGIEGSYGPTSFNLDQRPSTSSAGGSLGSQDPFHPGMNPANFAQHAPGVRNAIARSGSQTDNSERYQFSARWEEDEPNGLQ
ATTAAVLYSNEKHGYLRQQFPNWVDRVKQIQKLWRTLNHEVRLDYVNRARDNRTRSGNRPRPRRTNVQSTDSPPPPQMGFKVPANP
SSNMMGSSEQSMQPKQVFITHHLPMDVYQVYQDLKRVKFDRERLNGSLEEQLNKARKQKKNLAAKKRQMVKTQTAAPDYDGRNIDL
NESDQQTLATLTEQIKVTQAEIELHRRDLKKHESNLSDFVMKHGILRNEHEVKREDVQLSAIRDQQYAAHADEMLAQQAQNQGLGQ
VHPPVLVAPQAAISLPMLNQQQQHHQRMMQMQMRGGPPGQFMMRAQPPQQPQINMHNHEIAKRMLMQQFRNAHPVMGGIRHDEITD
PIMRDVYDCLEAIVFDVHTVAEREAMQAAGQQPPPPHIMQRMMQQHPGGPGGPRLIMPGQFPQLQGRRIPQQPPPVQLPQVDVDPP
KTAAKRKKGGGNPKKNASSYPAAGEYDAWIEQMRMRFRLCPDIPKKQKEPRLNRPGCEFVSHGLTDMAVRATKKRTPLVGEFGAMR
VKRGTRLFGNEERGEKSIMFQDLNLYQVEPQIRLFALYNNPLPTMEDDVFRDEEDQIIDKPTSALQKLLNKKRGASGAMDHIRSRA
RYMEETNIPPAPERGMFYENVEAEEEVTVELVFNTCDITEESSVDEKKDVCGRLHEQIRDLLAIKDEVAWSMEDTPPESPVSRSPE
PEATVGAVQQTSSTSRAPSEATEREIKREVEEEAAPVAPEIKREIVEPSQSQNSSQCKNCQKALESHSSTSIVTVKAEKLNLAPTG
SVVTTRAAPVPPGVPGAPGAPIAPSEEKTKEELASFCSKKCYYELASATRAPLTEEEVAAAEQHVSEEVYNKLKQVLTENIVKAVS
QGKKPDPQLTSPRDTRCMMDEGRRENVTIVPVASLIGAVEAPKEAPSQRAAEAAEDWKAYTKSIWDSFVNIQQQHAHHILSPKIGV
PHPPYHLDKRICVFCGGIGDGETTRCGRLISLTEFYWVHVNCALWSAEVFENQTGGLTNVDRAVLRAAQTACDHCKRPGASVKCHK
MNCGVNYHVLCAMQNNGFFIKDRTFICKQHEKVSNQAIVGRLDALRRIYVKRDENTMLTRLFELTDGPTLCMRLGAFTFYKIGSIT
PKQLKRFHTKDYIFPNNYRITRLFWSPKSHRERMMFECIIEDRNNQPIFVVKSLEDPTICYKAVSASKAWQPIYEKVHQLREQQQG
DSLKFFGSQICGETLFGLNENAITKITESLPGFDTIFTYQHRHQNSPVLELPLAENLMGCARAEPRNRTIGQHFRTKPQPMGGTHH
HIGSASGLSSQATIIASSSSSANESLQSRQAATSVTSANSGRTRGVRSYYTEEAAARARAFGIPPEMASLSLRMETTQNAQTAFSA
YQKMRRREWKDRVYLARSRIAGLGLYAKVDISMGDFIIEYKGEIIRSEVCEVREIRYVAQNRGVYMFRIDEEWVIDATMAGGPARYI
NHSCDPNCSTQILDAGSGAREKKIIITANRPISANEELTYDYQFELEGTTDKIPCLCGAPNCVKWMN

>LPT (*S. mansoni*)
MGFEGSEYEDDPLKDSTSDVQNPELLYTKVLSSFHPDESKTLQNFSDVTPPLCIFCGMNGNLNLGQGELLRFYTKIEYSEPPTWYS
DLIIKVKLKQEKCTNILVTNRSTHTNRRNQHSKIVVASFTPLTADPEFCKVGPHTSLDGCRRSRSKNCFGYDCKMYLAGMPNAYFD
CNGRPFGLVEELNYVGWPSKPDEDTLKVSNLIVKCTPREGEEGGWIYAHHCCASWSDGVHFNDQMVLEGVEDAAVIAVTRKCSLCL
RYGASISCRISDCDSCFHFLCAAGAGCLQDIESLELLCPVHISEALASNCTSIQCGLCESPGDLTELLFCTGCGSHYHASCLEPPL
QPSPTIRIGWQCAECKTCLICNESKDENKMLVCDVCDKGYHTYCLKPPVSSIPKNGFRCERCRVCSDCGGGRSSTLSGLEGPVAFN
NQLNPNVRWHSNYTLCDRCFHAHKRPNSCCPVCERAWRCSLPVPENFYRNPNSTFLVWPGSRCSQCRRMVHAECDPSSNQATVSPL
SAASEETSGICGTNYVCPVCRARGSTTPSEAASTTASLRASPIQGGGNSSNSGGDIMDDNFTQGGRDLLDDTVRQCIGSGSLVCPG
ANGSTSANTPSSSTSNSTTPNTEPGKQQIHLNNQDYPGSPRVHCKVTSNNKLSVTNSTSGRPAIDQPSINISSGRRRGNTQTSDLN
FTKTNSGTRSNNANTQTSFIGTTKTGALKRTNSSATSNNINKTRRTKTRKITSYESKVSEKDDHPSTVVFCRADDNFMLEQDICIA
CGSIGLDTPLLACSQCGQCYHSFCAEVPKITRTMIEKGWRCLDCTVCEGCGGTSNESLLLLCDDCNISFHTYCLDPPLKEVPKGGW
KCTDCVICTNCGQKDPGLNGKWHANYSVCAPCASLTTCPICNLAYREEELLVRCALCTRWAHANCDQLRTEDELEIATDLGYNCLL
CRELGADIGTGHAQVLAYRQAANGNLGALENLKFGEITENLFADKLPSSLFPYSSTGVASLARSQADDDNNSSSDTRQFFMDGVVL
SECGLNTIKQALLKIQPKRHTNQKRLSDQQSQSVTDTGCKLK

>Trr (*S. mansoni*)
MEFSGMNKTSEETLSYLSSSATAHPSLYKTLRYSPGNEYGEKSLSHSSPEGQVLRLVDAFHCPFNNDLRECESPQIPMVSAMCNRL
VPSDISSGFVDRSEFDHSKVLNNNEKPAAPEVGDTTVLLKPDLNDKAFPCSATKVVNNSPSPASNLIPDAVCIPSPCNGSLSRPES
NSDESCQVDKPETGRDVNGCRPTDVSPTHDQERAKMDLDKLDETIDYVLSVARSDEPCEGMFVLTCRARESTVTSSPNPRVNTPSS
ELINSPTISPAPSPHSNNLNSPVLQTCKTQAVFQSNSSCSMVSPQPMQRYIVDQTHGIIQLSSNPSQCVSAHYGNQGVSFVQLCSD
SSSNPVFQVSQHHIVSQISQPLTTNTHVLFHPSMLHSFQNHPGQQAQFTIPHSTLPTASVQRISTPILSPVRTSVLQNVVNQVAKS
SDTNITSYPVSNHIIESTNPPISTQAVSINPPNICVVQSPQSLLCGTVGTPVSLEASGDVKYSLSVPTVTNDQNCSQTNSSVAFRH
PIIPVNTTSPMLPTSSIGVSSHGLPVTNSHITTCSQSITSTPRKRSHASTSGNNKRRKTVGSPSYPVVPVSNKNITVHEASIRSLH
SGAQEQIDKLKGVDLKSLVSAPTLQPFFSVLPPTGSCDSPFPRGADEAFYHKHGLKYSRGRMSLKFPKCIDRTTSHEVVDSYLTQH
LKVVVPPGPVIYPSKCFKRQSVECHSVCASPEPPSFENNMKTLFSRMNQFTKKYSKDVYDETKVGYDSCKLSGETLIPQPLLQMFS
QNPSNSLERTETDRCTPPLALYHMPNPRLLCQPIHLAPCSGTGMDSGLTKVNGSNISTLKAEQNIPGLDTHKSVMSESSMSVDLNP
CTPFVDSETGEIMVSPQLKSTKASLGPTLNEDDKLHITFTVNPSATVNISKIIHRISEILGVEENSIDCQITRSGAQITLDQKQSQ
SEVVMRNLETHLRQQFTPLSLQFEKDSRDCLAQGIYSSSHIQSNTLENTPHTKLLPSNLPQLSSCKLECDTVNKTDFSPVYSQETC
RYSEEPVSITSILASRNRSAKSCKMCGKPVSPHMMQRKWLDELCSVSGISMNFDNAVDLLFCSSDCLYGFARLLSCYQNKYSETSL
DYSSQTDILNIRNCYNFEQQRQLIMTDFPCSMPILLQNNLSKSKPNLKRQHPQQCVTNKKRSTNTSTVPSKFRKWQGIRWRQYVSD
THHPSDTSSRRNIVDDALSPFISNPAVMRAPHSEDKRLCDLCHQAGDAPEDGAGRLLPLDLDNWIHINCALWCYEVYETIGGSLNN
LDVWLAKAKETACTQCGLLGAGLPCYNPKCSYVYHVPCAIKIRCMFFTDRGMYCPQHQPKEIHPMQLSSLAVSRRVYISRDENSQV

ARVVHEDSDKHVVRIGGVSLHSIGQLLPHQIDSGNFHNRRNIYPTGMCSTRIYWSMRRPRSRAKYVCEILEKDSHPFFRITAIDRD
MENVSVTHETCDGAWQIILSRIEVLCKEHKLVKLFPQLLRGEDLFGLNESHIVRAVESLPGVEGLRDYVFNFGRLELISEMPLAIN
PSGCARSEPKMQTYVKRMHDAGEYSISFNHRTSLPGVRRLSHPSFNGTTCSNGVYNIDTASKQHQSSRSQQYRKLKTEVNSNVILG
RSRIQGLGLFAARDLEQQTMVIEYVGELIRLEIANKREKNYEAHNRGIYMFRLDDDTVIDATVCGGLARYINHSCQPNCLAEFVNF
GGHSHIVIITNRRIKKGEELCYDYNFDLEDRSDKIPCLCRAPNCRKWMN

>LPT (*Clonorchis sinensis*)
MYFHFPCAAGAGCLQDVETLELLCLLHLDEAETYGFHATQCALCECLGDISELLFCTGCGSHYHGSCLEPSLQPNPTIRIGWQCAE
CKACLICNESKDENKMLVCDVCDKGFHTYCLRPPVSCIPRNGFKCERCRVCSDCGAGRASTVSGLGVMVEFNNPQLPVIKWHSNYT
LCDRCFHSRKRPTASCPVCERAWRCSLQVPSYISTQPQTSTHVTWPGRRCTKCRRMVHADCDPLQSVATGTASPSSAMSEDTNIAG
GIAYSCPVCRVRGSATPSEAASTTASLRASPLPGCGNSSHSGGDGSEEVFSGSGGLDDPARQGISAQSALGLGAPSTWVNTPSSCT
SNSTTPNTEPGKQQVLLNVPESSHAQGTRGASASPRLNTKGNATASSVSRGSSSSVNRISSESNLGSSTQSTGRKRGAGVPSETSR
SNASVKNTVPANGTVAITKTGALKRTAGASTTTSMTGRSRRTKTRKTSMYESKMPDEKDDHPSTVVLCRADDTFVLEQDMCVACGS
FGLDTVLLACAQCGQCYHPFCADVPKITRTMLEKGWRCLDCTVCEGCGGTTNESLLLLCDDCDISYHTYCLDPPLQEVPKGGWKCS
ECVVCTNCGQRDPGLNGKWHANYSMCAPCASLATCPVCTLAYREGELLIRCALCSRWSHAGCDQLRTEDELELATDMGYNCLLCRE
AGAEMGAGHLQVLAYRQAGNGNMHALENLKFGEGTENLFSDKLSPSAFPYSSVLNRLRQEEEPGSPAEARQFFLDGVVLSETGLST
IRHAMLKFQPKKPSIAGQRRLSSQGTPTPLVLDSGRSNDTPLSELTPLGSHGQGDEEGSLQSALDAEVESVSYPDWKSPSMLSEED
TGSVTTPVTAGTSTSTQESLRGPPGSSKSVAAAGGGKLSGVIGAKSRRLSNLGIGGFRAKPVRINQTKKQQLAAASDPFCQSSTYG
TLATFTNDSSGLGGKRKKGGKKKLELEENYPDYLMEAFYGASLLAANRRPLKKKKPPGSYGLSAQQFSRQSGSGVLAARSGSALKA
RVSLFLFCHLQTGFTEQQKRTSTDHRAAESNGLSEVFDSIEGASAGDELAADEECALDDITDDLELDDELDNDLGGLDELRDFEDD
FDAEPLECEADEDISTTLSMVDGLEFHSSSEIKFSSNLLANQSSILRSETESQNPRLLRTVSIPEASTEPFILASESSARSLVGRP
QSSYPSSGQPTQTEVLRSTSMERMLSTGSASRIGVPQSKTDTPTRLILEPTERTPAHIMPSPGQAGSVSVPQASVVTDHSRQDRPP
SGLDEVFMNMQDYFMNIDDLPDAEDAEPNETTIQQTLATGVAPSPAQGVASSTLQQLRNQTKHSLVLATSAHVAQPIQNPQNIPKQ
PTLQVVRGGESEAISVQPIESISQHSASLHCDQQSAKQFPLQQENVIYDQQRMRSTQEMAHLQPTEHRIQVHGPGMPAASSRVSEL
LDRPNLLSQTQRQQQEARQFLNIRAQEQQFARVGTAPVSSVASGRLAELIEPSHASDQAAPGAQLIPSNCELDTCSPLSSRVSGQT
NSSGRGSATTAGHLPELSVSDIETQLGPTVGFELNDVFKGLNTSETNSGMLTDSGARDVAQVPAFLGEEVATPPEMLGQRMQQQQQ
AQQHRFSASQPTVHLVQASPEQMLRSSQGQHLAEQSREIPQHIQSGLPIATLQRSSPMTNVQQQQQPRVSVQQPQFETQSPRLQLR
CQQPSYVYAHVVPSQGGRMIPQTQKQQQQQQQRLVNSPGLTTSHQSGSQSHVQMLVSGHASATHFIAHPVQARSPHERTPPPPPPP
YPTQALRQTHMWQHPAAQPRILTPTGQGTIMLTTSGQAIATHGSNSAQLIMSPPATPISIQAQHVLPPSCPPTGQMDSAQRVCIQQ
TGVQRVPTHVVMHGQQVVQQQQQQQQHLHAYPSEHVGIHGTRQVSFVTSHRPGAVQPSQPSAGVLRFATITEEAVPRPPTESHLLSM
LNPSQPIMPASPANQQQPTLTGPATGGVVTGPPPTSSGRRINYHKWEEDERLGNQSTIAPVLHANVSHPTLRGQVPEFTARAKEIT
KLWRRLSSEERGSWVSQARNNRTNLRGSHNQHVPSNAPGSSPSTSTPSTPSDTLMRPQSSTSTPAVFSQDLADAPMTPQRILPIQA
TNISPRTVAEESTRPQYVRPGVESVQSLLSQELQISPDTSSRVRANSLDSLATSTQLPTVEVGDNHGSSDPLTGEPPYGGMHQTSP
FGYSQNRRLEECHPSDVIPGSTGSWRSAANTPCGADQISQTIPHDVQQAVGGPHPRVSPSPGRVTPSAHLFAQSPHASVSASTGSP
IISQAAPTSSSVYATPPGLQQQPTPMKVQQPPLSGPPSPSFMKRHSSMPTSAEVQLPRTPQANIYIGQQSTPMATPPATSQALSAP
GTPVPRSSPGHLTSHQHQPQQQPHVGSTTSNVFVVGGGTVQPFYQNSGPSESDVNQNSFPGDASLDGSLRQSQVPAPLSKTMDDPR
SITAHPSASFHSHISMQQQQQQHLPHHPQQPEQAHEQVVMRHGQVAGPGQHGQPVMAIPYSSMEQPTTVYVQQHPQHVGEPNNPTY
VNYPNQATAWQYQTTSVPNRQQPTRQTYYAQQGPPPPNNPQQHQQQSPMLYQQFVGRPSSLSPAISRSPTTPLAQSAQLISPSIQQ
GSGQFYAPTSSPAAHEYGPTSMKVAHNSTPDFQQHMSSYGPSSSAAPTYSFTAVPIHRNEAMFAGTSSTSSETVIGSSPMMTGAPQ
QVSVSPRGNLGPGMEEVGFEPPVILSRVVSAKEKRYPPSALAAPPTGQMLPGRSLPSQPQTPQHQVQQHQQHTQYQSLSNAVPPAE
IRMQPTPGVQNMSGASTCAELDTGSGTLVLVTCVTAVGRGVRSRSRNICIRSLHDTSEFVDSVQTVNVNATRICSVDVSSLTVREW
NETRLDKPVTTAILFHIVTNIDSTQYFGSYEPDELTVRVTSFGGGGQPVLSYAERSPNIRPSPSGMTNLAEFSSPMSSYTQPGGSG
SLQPPIVSQSGTEPAFGVKVDHFSSLFYRINVERDHKNFEFRLINLNLIQHIGSPMRVISGQTCNIHDLVLTTTTLDIDGIRTLNA
NFDKSIFSRHSCGQTKPRIFGVIVRPILEYGQPIVCPLAKWHRRGKLTEDLPQDYRKTICFDCASRCNVVLCSGNGLAFGRKVHRV
CRKQTGTGYMEQLRQAASCFSCYDIRNVAIHVYSKCTTPKVAENFRQPRTSFTFLGAHSVAWKHYKREIQLGSRPQMWMQCLGDTF
ATMKENKLSPVNRMLNNVLSDIQWDKNGARIYRGVIKAEDEDEGQNANVTYSILDTANFKVEPNSGWITANTVFDREARDSYQVTI
VASDQGNPPLSSTALLNLTILETKQPNPLGRADQKQTGFSLLPGSGRQLEADTFIVRENAPPNTYIGDVVASQDQWQGSGKMYFEL
LDDTPNLYSTRFRLLRNGSLYTTVKLDREEREEYQLEVRVNSQIIFGSVISTGKITIRVLDENDNAPNFVHPPGLLSA

>TRR (*Clonorchis sinensis*)
SAFRLPQQRARCGMELSGMSKVSESTLISSSASCQTHLYDRIPRSPTAVVQPSSPIPSAHNSAGTDGLKFTLVDESPCLAAFKSQT
DNNHHSISPEISFATFSNNLERTDDLSVTNQATSGPDLSQFPNDDSSDINRSSNVGKSHSTPLICTSPAAAPGSTFFSPDHQLLNG
GSTPISFVCPADCTTPTAADTSLDVNGLSGSHTTASPNDVDRTRMDLTKLDETIDYVLSAARSDDPCEGLFVLPCRTREHTTLSSP
TLRANTPTSGHLFSPSPSPTPSSPQSCHATASPLNSTNKQLHLYVSSSSVSSIATATSMPRYVVDQSSGLVQLSTGVVNSQCLPAM
CNGQQSLSFVQVFSDPPSSSRIPNHERSTVPQLGSLVTSPQVVFHSGLSHAVQNHSIAQVNHTLLPTAAAVRGCGPLVSPNKSNLL
SPATGSRSTKLSTIPCAPFSSIDPLLSPNRAVVISPPMCVVQNPRLASPTRNVPVFSTSSNDMPNQCPQLPGERRHYVHSQVPVSF
RQPIVPAQTFTPSCISAANTVCLQPTVPDQNIPLVHPGPTTAESLRHITQNHSAASRKRSNTSGSGLSKRRKTTAAPPTSLANGST
NLIKGSLIEPNSPYRANIVAHEYFEKLKLGDLKSLVSAPVLRPYPSVLPPVGGSDIVTPFLKVDDEQSVLCRKGLKYIHGSMSVHI
PKCSLSEASDRKVDSYILRHLPNATFTGPLVCGSKNPTHAMSPTSPVCPSPEPLTFENHMLALFSTFNRSPKEECKSTLVVDRSVS
QNTSNMPSFASRLLPMLTEVTPNSGKPDPERSTPPLPLFHMPNPRLVYDHARSGSTDRDGVEFNTTAQSETSRSNSTFIYNSPSRE
PNDVKPPVITAPGTGPLIDSETGEIVAGHRFRPPAESMHSDKLHITFTINPSMTGGIPKIVQRIAELLGVEQDSVSYQVTRSGAQI
TLDQKQSQSEGLMRNLESQLREHFTPLSLHVETASVRPTDSNIHSDLQTVVEREELSNHPGTARPELKQNVDTSNLLKLDGPCITN
SEWARYSEEPVSVVSLLANKNGVSKPCQNCDTLVSPYSGHRKTLDDIAALTGLTTRTDTNGDFIFCSKDCMLAFAHSISVRRLSLE
SPIACPTSATTTGSVSDSRPQTLVCDLFSTVPVVLQSSLPKLGKSSRRHLTGVHIVASGHKRRQVSAVHGTAKHRRWRDSRWRIFS
PEFVPTTTPAPSDKLITSGPITFSRDDGPCLRPPPSADPRRCTLCGTKGDASENGLGRLLPMNIDKWLHVNCALWCYEVYETVGGS
LNDVDIWIKKAMETNCTHCGHLGAGLPCYNPRCTFSYHVSCAIAIGCMFFTDRGMYCPQHQPRETHPMQLPSLLVNRRVYITRDEN
AEVGGVIHEEDRTKIVRIGSVCLHNVGQLLPHQIESGYFHTRRYIYPVGFHTTRIYWSMRRPHRRARYVCEIDEADGRPSFRITAL
DDGLEDVTVVSDSCTGAWQVILSRIEGMRRENHMVQLFPKHIKGEELYGLSEPHIVRAVESLPGMDGLRDYVFNFGRMELIAEMPL
AINPSGCARTEPKMQTYVKRSFLSSHKLTPTTSSPVGFRRSVHPSLSISTPSDYSKQYQSSRSQQYRRLKGEASSNVILGRSRIQG
LGLFAARNLEPQTMVIEYIGELIRLELANKREKDYEAHNRGIYMFRLNDDTVIDATVCGGLARYINHSCQPNCFAEFLNFGDHSHI
VIITNRLIEKGEELCYDYNFDLEDGGSKIPCLCRSTNCRKWMN

>TRR (*Clonorchis sinensis*)
MTTVYSATCVSQNCAIPTDVCPLSLSLDVPVLAEETSKYPVASQSLTFDDYDEDASNVFRYSPASVRTLAKESSPGSCELLSMPSE
KSNSRGCSGDGSTGWLVPSEVTEMEHSASESSLSPCFQRNDYNNEMTSPVSVRLEDSVPSTFSQNSSAHPDVMNAYCSVVSPQHTT
LSQSTFSSSMSSIKPTSSAFTPPYSSTPIHDRPFPGHSVKGCSYSHSFRTVIPKPLPDPSLELPTQPNAQNGPLLIDSDFRSHSAF
PPRSITDASSSQFVPSTSLPLHSPIFPWVTQRPQQSSQRSEACSIAVTLPTTMFTSCPLNSPVATFSTFRERSTTQSVSNSLPFSR
RVHWQLDAPLKPQCDSSVTCEHLARSSALSDGSPHLDSNQPFSEHVRQTPTVIINSVESTRYFSPRGMRKSTVTSHPRTTISRSSH
SREMISNGEHILPSTCANSSLPLSDSHSKQRPIFDSSIPTYALQETADHTKRSMQFASSSRNAKKRRLVDRAVTSISVGNSSIPSF

RTFDSRFWSIGPHQYIHDPVISFPATLNAPVPTPIHDRRCAKLEDGAPVTSNDCSVTPPTGIRNNKYSTFPMFPSRPQNFPTTEPQ
MGDFLAGRNHDRTVFHKLIQENLFFICLRKASFRSSQSCISSCPPTPPPDALLADKLLECLGDVRDCGRTTPPLNLIHKCIVSSKG
HGSTKNTDITKSVVAPELFIAESLFCAYCALQLVTNDHVLNLKAPDGDGRILFCSFFCLTAYHNSGALKEVKINGVSHSSPVLGVT
SLLSDLRPSLVILLQQHANTKWPKKTGKRSNYQSKFSRRRPSDSVNYSRSVWNRCSSTRYWHGIRWKKFSRGASCTRLPTPDAVVQ
NSPSIPIIPIRNPPFGFTPNSTLPMKQKCIVDRRVCMLCHKTGDAPNDISGRLLHYDYNKWLHLNCILWCYDTYETVSGSLVQVSR
ALEKAERTLCAHCGSTGAGLPCFYSDCQAIYHVPCAHSIGCSFHTDRGMYCPLHREAANKLCQLNSLSVERRVFISRDESSQVAGV
ILQVNQQNEPSVRLHGEDSFVKLRVGGLILHSVGQLLPEHLSGTSFHSAEFIYPVGYSSTRIYWSYRHVRQRCLYHCRIEDSLSKN
GDPISSFNLYPKFIVQVDERDEPLEIFEDTSCDGVWQQILTRINQARRLSGLSHLRIIQNQLRGELLFGLAEPHIVRAIESLPGVD
RLSTYIFKFGRLQLFKEMPLAINPTGCARSEPKLRTHVRRRPTGDASQHGQLRNAAIANPMSTPVRSGLSSPVYSTAASFLSVSKS
QQYRRLRLEWRYNVILARSQIQGLGLYAARNLSKHAFIIEYLGELIRNEVGNRRERLYELQNRGIYMFRVDEDSIVDATMCGGLAR
YINHSCEPNCTAEVLNCENGSHIIIIASRDIEKGEELTYDYKFDIEEERGDRIPCLCGAPNCRRWMN

>TRR (*Capitella teleta*)
MSRASDQSPATIISSSSPECAQLADVRPHNPLLPLTGLRLISEDSQDSSLLRSDCQSPEIDVIKPVPIKVQPKDEPDLQQPKMEPL
VAVEPMRENQLSVTLTMSSRAAEDIGGVLNAIADLLKIAVPPSFEVSRSPSPDLFSFARHAAARHMEEKVDLQTLIKSRMKFCRHC
DVAVMTGVRKKKSDIPYLGKEEFVNKDDEMMFCSEDCYEQFATAHRHARESQMAREELNAQIEALRLKQLQLLAAQTQSPEPMAEK
LRLRRASSLTVEPKYPPITERKWAGLKYKKWDATMMKVGHVAPPPTEQELLQLMTKLGQTIRCVPDIVDKRTCVLCHVTGDVETDG
PGRLLHLSAEGSGDKWVHLNCALWSYEVYETMNGSLVNVDQACKRGASLQCVVCSQSGATVGCFKLRCTNAYHVRCAQAEGVIFFQ
DKTILCPQHTPKLPENVLPSLAVFRRVYVSREEDKQIASFVHQEEKNSVIRVGSLIVYNVGQLLPTQMHSGIFNTRDYIYPVGFHS
CRFYWSSKRVHRRCRYICCIDENDNRPEFSVTVKERGHEDVTFKGDTCSGVWQSLLENLSRMRLDLDLVKMFPLYLSGEDLFGLTD
SNVVRMLESLPGIELITNYTFRYGRSPLFELPLAINPTGSARSEPKLRTPFKRPHTLHGSSNTRSIPATVTLVPGDSTMPYMKQFV
HSKSQQYRRLKTEWRINVYLRRSNVQGLGLFASRDLERHTMVIEYIGELIRSEVAEARERVYDSQNRGVYMFRIDDNTVVDATMTG
GPARYINHSCAPNCVAEVVPFEKDSKIIIIARRRIARGEELTYDYKFDFEDDQHKIPCLCGAPNCRKWMN*

>LPT (*Capitella teleta*)
MPIERTVSEDSDTSIGMDKTCSLCYRGERSLLGQGDLSRYEPTPGFNAFKQQVRRRGSLDMDEKLGDRSPKHLTWRRNRGPPKNPR
QSYASAGGSSCNVDELNLVGLPDDVDVAHVFEKSGKRHVYAHHCCAAWSEGVLQADDCSLQYVDRAVLTGLSQKCTFCKRYGATIQ
CRYPKCSQRFHYPCASGGGCFQDIKTLSLLCPDHIDQAESIAGPEALCIVCDVPGCISDQLFCTSCGQHYHGNCLDPPVQVNPVVR
SGWQCPECKICQTCRQPGDDNKMLVCDTCDKGYHIFCLRPVMTTIPKNGWKCKVNFFGVFL*

>MLL3 (*Hydra magnipapillata*)
MELDDGMELDDGMELDDGMELDDGMELDDLSWMELFKTKKDSEQKNCNISHKNDSNESEISSSKATDGLVDNEGSITSNR
EHYGSDDEDDIFNRNQNKPFMKRSNRAENELKKINERLLKKVSVMLSDDEEAFLGRNQDEYISDCSSHVCSDSEEEILRVSNKKLC
SDHKKDENEFGNNKKTDIVHQNEQEILGRGTSLRVNETYIRDQQIAENHNSEEFKHHLYHNEKGTVQSSSFPSTDKYTIEQPGLTN
GGSLSRLQNNHIPLFTNNQPTSYLSNNLYSSQNSYPHPSVYQLGVMGYPNEGNDKHFESRQHSALKNESQPDWGSFSGSNKNNYQK
AGVLHQSQTKYSGFSSPSQNYNAESVSTSETLKTSFSHFHSKSSLPQSVSSVSSASNPSFSNSPDSSYFSPTITSSANNSINSDKA
TVSFSTTTSSNIYSTCHPNSMEISGNSMFQQGSISGSMQFPSSGIVNPHYSNPNGPYIPQSTNPLFPESGMAPFTQSSSLPPLSSS
ITSNPSLIQSSPIPGSVPQSSSPFKINSSPQKSNIVSLHPNSQRYHQIFFYQHQLLIMQFQQYQFQLQLQFQQLNQQQIGQQQQFL
LQQQFHQQLMLLQQQLVQQQNQLYRMAANGIQFTEHQLNAMYQQQMLTQQSYSQWIVALSPEQQVLFQQRMQACMQQVQHQMAQIL
QQSSRKEEEQKKKPKPRRSLNKKKSAALLAADSVQTSQPSTKFPVPELSEEMKIRTGQNLVVSKEESSHSEPNQITSKPDSILPTN
SVSSDSKPNQQNTCKPKTEEKTINPNIISPQYEKIHESFIRHQKQFRGEIPPGPSPLSGYSRCDTQPGQDFQRTQQNDISKPVTTV
LSGEISPGSMHQKAEQLLNPRNTAIPPKPSFFPTHPSKLFKQATEKCEKSPSQESLVSNAILPMGETPSFGSRPGINQLSSEPSRI
RNPFLNTKTGESSPSNNSAFSVSNLLGTPKKQSLVSTNSQPLFEPAKKAIITSEHLSVSQVIKQNDSDPGTNSPNDLACDNMYHMT
DGETEKYEGMNTQLKNLIKLHNKKKDKILEEKKRLRELERIKKKEEKEKIKVIREQRKAEKYQQRLLKQKALQEAKAQQLVIKKHN
KKIIVPLEIPDMPSGSKTPTIVSPLSSFTPKTPPMVVTPKIPLCEPDVHLVCPLVQPYGFNSSNTFKITGKFGYALIENVIDVYNE
KPYTTSPIATSSFLKPISQCSVVVNDILNDVSLFPSPEPNNESSLMHSDIQKDKLLGSIVKEESESVRTISSTQNIMNFKDEDVDS
EIEDTISRCNLCYGLLHNVVFYVDKNMENVTECMYDKFELMSSEQLAFCSNECVEAYQRKVDYTFENDDQELIDAVKVSSMVTTDE
LHVFLPLPSDEIRGDVCASGMSNKLVSKSNCRWKRWRVSSDNEKKTKIKSVKEDINQLMKKYSVALKPSESIRDRRQCCFCLVVGD
GEANQSSRLLNCNVDSWVHLNCALWSSEVYESLNGGLHNVDKALARSKQMRCSYCSKIGATLNCSYTFGAQQRLHCEKSFHFHCAL
QSHAALYKDKTILCSEHKSLGIRKEELDSYAVQRKVYIERNITEQIARMLSLNNSFERKEYTFRVGNLIFHELGHVPMQLLDNFSS
KDFVYPVGFETTSIFWDNHDIQKRRYYTCSISECMKSPRFKVKAHSNGSEGSLYTSSTAKGAINEIIANIQKLREKENCLKVFPEF
ICGEDFFGLSEPTIQRMIESLPGIDTISNYNFRFGRTPILIAELPLALNPTGCIRTEKCRKHHTKPLAMRTAAEHTHTSDAIITDD
SWLVYSQRNVHIMNTKTGQYRRMRQEWRNSVVLGRSRIQGLGLFTKKDIEENTFIIEYIGSLIRNEVANRLEKHYERQNRGIYMFR
IDNDTVVDATMAGGPARYINHSCDEFVNGF*

>MLL3 (*Mnemiopsis leidyi*)
CKDCQVCEQCSNPSDESRLLLCEDCDISFHTYCLTPPLEQVPTGPWRCSWCVKCVKCTKTTPGVHATWYQDYSLCGPCHSQTTCLV
CNQEYVAGEFCVKCFSCEHLMHGRCENIASEETCPKKFVCSQCEEMVKKEGETPPKKLQSTTFSATKQALDSVIKHDARIVAYKDK
DEKEFDAEREEKEKSRMLGYDYIKALNKQRHEHQLLQKSKKFSAQDKDPKMNELIQEGLTPPKDFSSSSAYQQIVRELTGNMPGNK
SAVKRGRGRSVNKASDLMSPTKDPNAPKNRGRPRGSKNKTPSQRTRGKGIPHPEMLQQQLSSDDSINFSDFTGMGLPVRPPPNIKS
PMKAPLKRSNSNKGRGKLIPQIMDPSQMMKQEISPAKMNPELAMNFFGDELMPAPGPDMMSPPQLMASQTMGPPMPMAPPVLSSPV
NHPVNAMLPPQPGKPPYNVQIKPRNKLPPSPLLDPLQNKNSSVPPSPNNMNVLPSSKPQTPIKKEPEPFDLSKVNIDEINVCPESP
FAPDHPDTLPEQEKIQAMEDQEMLRQSQRERRQRLYERHPTANRRQDGMITGNDPMMGGWQRFRTPEETSNSTVTVGPIPSKAGMP
NQMPTIRDRPMTPAGPMTPGGPMTPGGPMTPGGPMTPGGPMTPGGPMTPGGPMTPGPIRSPMNPRAMATMNVQMMQSQMMGGPGYN
MGRHFDHTTGQWVPHAVPGAHTPLRHLPPQAIIAQQNMQAQFNHMNQFGMQPTPEPVQPPPKTKKRTKKKKEKAPPVELPTPMDPQ
VNLENFPPHLRFGGPQTPAPPSSEFQGAGPGHPGFDYTEHFRQNIQMGPGMQNPMSKNFKRGNSPLLKSPQTSVDMERPKSLNINN
SSPSATPKSHPNTPNKGAASKNSSPMIKNTNDDFSNDSHDSQQQMHQKHAIRIALEQRKQQQSHYMSQQMMMPSNLHFPMNVPQFV
QFRLQQDWLKKFNQLKMTGGGIPQPPQHPNSMTNQQIQALYNRLMENIFNMTAGRMLGPGHPGMMEPMGHVIEEEKKPKTKKRKKK
KAETPVKKDDKSKMLKIGPNLVMQTAMGPGPLGANSPQTAQDNFTGPHGINQNSLGFQQGMGNQMPQNIPGQGGGPPLGIINPKHL
NMRSPGPAIMAGPHVSHIPGQIHQQQRLMAPHYAIALSRNSDPRHIRPTLTTQSPHFHDNYDIEPPRTPKPTPKSKSKSKTKESLP
SSVLKPSDTPMQADSPRYQFETPRYQHQIKIEDNQDFSDVKSPQSQRSQTPSSAGTPTMKDPFDNPPLTPSTYNSLESAKRRKRRT
KAEMDEHRAERERVGYNGPPIRAANKHIPPVQQPDIGKMVDIVVRGTGTSPHCKLVGSYGRAIIENQRDFYRDLVAQIKESRQKQA
AADIAPGLILNDSKPQLTSYDKEAVKLHKKLSVQEQHMEKLRANIPVHHAPTVLTPQQEKMLCQRFGVTPAEAQRMYYYQQHQHMM
ALREQQIKAYQMQEELEQAKSGNNDSDKSKVLEHIRLLQESSREKNSISRGNSPAIAMEKIKQEENVSVSVKQQSDTIQLIFQRPM
NEKTESKEQVEHEKPLLKPHKPMLELTKPTPPAHIKQEYMNKEVGSDNSMSPRELPFAITGPDFTPSIKPCNPPTDEKTTRVINLH
KAFQMLSDMLGMKLKKDNLASLNELRVQNIIPGSLPHNLVCYNCIIVLHPNKLKHIAPGRPYYGIKDEHGATFCSDSCLAQYYCKK
FSSLRPEDHVTPAPPPLRQPISSPLPSSPGQRVLTSPGKRRRQSESWKYWRLEDVVGMDNHRQNLADDNKRLFEMFKKWHEEHTTQ
TDAVDQRVCLLCGSVGDGSPNGSSRLLNYRKNTWIHLNCALWSNEVYETRDGGLVNIPQICKRAQDTTCCCCNRIGASIVCFKAKC
NAAYHFLCARHSKCAFYTNKTVYCAAHKFQGAICDDFTVRRRVVIERDVQQLLANEGFKNPGDKIRIGPMIFHHPGHILNNTSKFH

SRDIIYPFDYSATRIFWSTEDVGKRCSYDCSIDVKNEAPLFKMSYQKQGKTVDIENVCVDALWKEILNSIAELRSKTPDILSIHVK
FHKGTFLHGLTERNVLRIIEGLPGSDKLEGYTFCYGRIETVTLEHLVPFNPSQSARTEKYTKHTKRVGGHRQISSSIQGSQAPSLW
NREFGMFHDMLPYQKSFNRASQYRRLKTEWKLNVLLRRSGIQGLGLYARRFIDRNTMVIEYVGQLIRTSLCDQREIYYNSKNIGCY
MFKIDDVVAVDATLTGGPARYINHSCEPNCVAEIVTFEREKKIIIISNRKIRQGEELTYDYKFDFEDDSEKIPCNCGSVICRKWMN

>LPT, isoform A (*Drosophila melanogaster*)
MSTNVVVSMEEKAVADDLPYFPEKFPGKVCCLCNLGEKSALGQGEILQLKAPAPAEIKDKYPTDRLREDGSRLLYGPKQAASSSGD
CHYELDKIGQPDVVHPAEFLDEGFVYVHRMCIMWSLRKSQISDADGAYFATHFAEFLEQKCNFCGRYGASINCKMNCRQVHHWPCA
AAAGCLLILESFTVFCTEHLSQVPVICSDNNVECLSCSSLGDLSKLIMCSTCGDHFHSTCIGLANLPDTRSGWNCARCTKCQICRQ
QDSNDTKYVKCEQCQKTYHASCLRPVISAIPKYGWKCNRCRVCTDCGSRTPGGGSSSRWHSHYTICDSCYQQRNKGFSCPICQKAY
RAASHKEMVKCSWCNKFVHSTCDEEADLTAYHKKKEQNPDYDYVCPNCKSNSSGPGSSQQTIDSIVLSAMDSSSEQLSLKEIELDP
LEGKPTMDPSSDELHKLPTGKKKVCLTSVRGRSGKFVLHRMGVMSQINKKRSTRGKGRQLALPTISSDRCLSRSMETDLTSDKKLL
LCSARDKFIQAQDICVMCGSLGIESDSVMITCAQCGQCYHPYCAGVKPSRGILQKGWRCLDCTVCEGCGKKNDEARLLLCDECDIS
YHIYCVNPPLETVPTGNWKCSFCTLCQKCGRNPTEKSEFGDSNMLECPSCTSQSSCPVCKVSYSNGEMIIQCEHCELWAHFHCDSV
NAQLTIDHYDNNVYKCFKCRCSTRSTNSLADAKFGNEDSLAITPQGSKPFVHITASSGPSFDSKSGAERNHPTLSGDMPESAPEAL
HWIDGVCLSESGLGMIKSLSTEIKRKRKMRQALGNGKDGQLEAAAEEALEKYKDGMVWDGTENAIPEGFTISINDEGVNILRKKRQ
RNLQKLGIGGFSVRNRGLKKDSEETTAVDQINSLMTMDKKKKIIRKKQKNKLIEAYPVYLQEAFFGKPLLEGGELVMADTDSSDEI
DASMKVYFTRPEGKSPANQDTLAVNKSPAKTVKKVKAELKTENKTVFIQQMPMGIQNKNGTTPVSNVFDPLQAEISNTVFANRHPL
DALASPNVPELDMANSVSTVPSAEVIEKPMRESPAVLVDSFMVPNQLPQEQKFHNQLYLGGNTLVNPAQQLQFRQQAPQNANKQNM
QTIMSQVVLQTEKRPEEEQKVVEPVTETLNAGTQKTAEKMRKDEDLGLMATISAVLYANTEHPNLKELFPIWNDRCKQILKRWRSL
CNEKKAPFLQKAKDNRSALRQRREQNKIPMPPKPQKQEEMGRVWKQPPKLKEDQPNMFVTYNGNAYDMGNYVGGSAQAASNNPNPH
HVMPNVNDNLVIKATMQRTQVHTTAASTLKMTTVDNRLELNTAVYGTEPIGDKKLRNLLQKNSTEPMLMGANSDLFLGNDVMRLGM
MQNTPITQNNPMLSISGKSQPSEGRSLEQDVKLGEPQEATSSAVELENVGFSDLLGGLGEGDDDDLLKSLTSEMGDDFNILEYADP
ELDVNVLNSLDFDDNEKCST

>Trr, isoform D (*Drosophila melanogaster*)
MNIPKVTTSLGAAEKAKPERVASVAAAAFNAVSLQKRSGDDTATPAEDPTRKKAKTELLLGTGTAAPSLPAKASSTAPQQLLYQRS
GQQAKAQVKAASEPQDVETADGVWDARDQQIIVCNFGSGTEMGAIKAEDADKQSEYRISTPRNSQSNPLLHRNTAFTSFTKKEGAS
SSASSSSSTASVISIEPSGSGQDHAENSGKSEDLDYVLMPASGADSSTSVGNSTGTGTPAGTPIGATTSTIILNANNGTAGVSGAG
TTTILTQKSGHTNYNIFNTTATGSQTPTTTLLNRVNLHPKMKTQLMVNAKKLSEVTQTTAKVSIGNKTISVPLLKPLMSASGAATA
GGATIVESKQLLQPGGQVTTVMSAAQQSGGQQVHPHVHSHAHHNFTKLIKRGPKNSGTIVSFSGLQIKPANTKIVATKVVSKKMLQ
LQQHQQQIQQQQQLQQLQVTSGGGLAPPTGSIVTITTTNPSQTYAMVQDSATVGPAAHSEDDAPAPRKITAYSENLQKILNKSKSQ
ESTGGPEEEFTNINSVVIKPLDKNTLNCPPSFNIFKQQQHSQAAQSQSISAVGSGAGTPVTFTMASGNASDLATTSTVSVSAGTICI
NSPMMGTRPIISIQNKNISLVLSKTTMAQQKPKMITTTTLSSQAALQMHHALIQDSSADKAGSSANSGSATSGASMQLKLTTANTP
TKLSVSLAPDVVKLEEVGSESKAKLLVKQEAVVKDSTGTPTSEERAEEIGTPEKRLNANATMTAINQVQNQSANQIQMATSTSTAS
NPSTPNPTVNATPMNNQRSAAEDNALLKQLLQNNSSSHSLNQISITSAHVGSASASAPLSARKVINVRAPSMGKVRSLEDQLARPV
IPPVPTATQAAGSSSSSGSVATSTTTTTVASGGSSQQVATASATALPVSAVAITTPGVGGEAKLEQKSDQPAAIMQNQSQNQAPPP
PPPPQQQQQQQLHQPQQLQPSPHQVKQTVQIVSKETSFISGPVAAKTLVTEATSKPAELLPPPPYEMATAPISNVTISISTKQAAP
KELQMKPKAVAMSLPMEQGDESLPEQAEPPLHSEQGATAAGVAPHSGGPLVSAQWTNNHLEGGVATTKIPFKPGEPQKRKLPMHPQ
LDEKQIQQQAEIPISTSLPTTPTGQGTPDKVQLISAIATYVKKSGVPNEAQPIQNQSQGQVQMQAQMQATMQGHLSGQMSGQISGH
AAGQIPAQMHLQVQHQLHMAVHPQQQQQQLHQNQPQNATIPLPVTGQGAVPIPVPTMESKAGDQRKRRKREVQKPRRTNLNAGQAG
GALKDLTGPLPAGAMVQLAGMPPGTQYIQGAASGTGHVITSTGQGVTLGGVGASTGASSSPMLKKRVRKFSKVEEDHDAFTEKLLT
HIRQMQPLQVLEPHLNRNFHFLIGSNETSGGGSPASMSSAASAGSSSAGGGKLKGGSRGWPLSRHLEGLEDCDGTVLGRYGRVNLP
GIPSLYDSERFGGSRGLVGGSARTRSPSPAESPGAEKMLPMSSIQNDFYDQEFSTHMERNPRERLVRHIGAVKDCNLETVDLVESE
GVAAWATLPRLTRYPGLILLNGNSRCHGRMSPVALPEDPLTMRFPVSPLLRSCGEELRKTQQMELGMGPLGNNNNNNYQQKNQNVI
LALPASASENIAGVLRDLANLLHLAPALTCKIIEDKIGNKLEDQFMNQDDEKHVDFKRPLSQVSHGHLRKILNGRRKLCRSCGNVV
HATGLRVPRHSVPALEEQLPRLAQLMDMLPRKSVPPPFVYFCDRACFARFKWNGKDGQAEAASLLLQPAGGSAVKSSNGDSPGSFC
ASSTAPAEMVVKQEPEDEDEKTPSVPGNPTNIPAQRKCIVKCFSADCFTTDSAPSGLELDGTAGAGTGAGPVNNTVWETETSGLQL
EDTRQCVFCNQRGDGQADGPSRLLNFDVDKWVHLNCALWSNGVYETVSGALMNFQTALQAGLSQACSACHQPGATIKCFKSRCNSL
YHLPCAIREECVFYKNKSVHCSVHGHAHAGITMGAGAGATTGAGLGGSVADNELSSLVVHRRVFVDRDENRQVATVMHYSELSNLL
RVGNMTFLNVGQLLPHQLEAFHTPHYIYPIGYKVSRYYWCVRRPNRRCRYICSIAEAGCKPEFRIQVQDAGDKEPEREFRGSSPSA
VWQQILQPITRLRKVHKWLQLFPQHISGEDLFGLTEPAIVRILESLPGIETLTDYRFKYGRNPLLEFPLAINPSGAARTEPKQRQL
LVWRKPHTQRTAGSCSTQRMANSAAIAGEVACPYSKQFVHSKSSQYKKMKQEWRNNVYLARSKIQGLGLYAARDIEKHTMIIEYIG
EVIRTEVSEIREKQYESKNRGIYMFRLDEDRVVDATLSGGLARYINHSCNPNCVTEIVEVDRDVRIIIFAKRKIYRGEELSYDYKF
DIEDESHKIPCACGAPNCRKWMN

>MLL1 (*Homo sapiens*)
MAHSCRWRFPARPGTTGGGGGGGRRGLGGAPRQRVPALLLPPGPPVGGGGPGAPPSPPAVAAAAAAAGSSGAGVPGGAAAASAASS
SSASSSSSSSSSASSGPALLRVGPGFDAALQVSAAIGTNLRRFRAVFGESGGGGGSGEDEQFLGFGSDEEVRVRSPTRSPSVKTSP
RKPRGRPRSGSDRNSAILSDPSVFSPLNKSETKSGDKIKKKDSKSIEKKRGRPPTFPGVKIKITHGKDISELPKGNKEDSLKKIKR
TPSATFQQATKIKKLRAGKLSPLKSKFKTGKLQIGRKGVQIVRRRGRPPSTERIKTPSGLLINSELEKPQKVRKDKEGTPPLTKED
KTVVRQSPRRIKPVRIIPSSKRTDATIAKQLLQRAKKGAQKKIEKEAAQLQGRKVKTQVKNIRQFIMPVVSAISSRIIKTPRRFIE
DEDYDPPIKIARLESTPNSRFSAPSCGSSEKSSAASQHSSQMSSDSSRSSSPSVDTSTDSQASEEIQVLPEERSDTPEVHPPLPIS
QSPENESNDRRSRRYSVSERSFGSRTTKKLSTLQSAPQQQTSSSPPPPLLTPPPPLQPASSISDHTPWLMPPTIPLASPFLPASTA
PMQGKRKSILREPTFRWTSLKHSRSEPQYFSSAKYAKEGLIRKPIFDNFRPPPLTPEDVGFASGFSASGTAASARLFSPLHSGTRF
DMHKRSPLLRAPRFTPSEAHSRIFESVTLPSNRTSAGTSSSGVSNRKRKRKVFSPIRSEPRSPSHSMRTRSGRLSSSELSPLTPPS
SVSSSLSISVSPLATSALNPTFTFPSHSLTQSGESAEKNQRPRKQTSAPAEPFSSSSPTPLFPWFTPGSQTERGRNKDKAPEELSK
DRDADKSVEKDKSRERDRREREKENKRESRKEKRKKGSEIQSSSALYPVGRVSKEKVVGEDVATSSSAKKATGRKKSSSHDSGTDIT
SVTLGDTTAVKTKILIKKGRGNLEKTNLDLGPTAPSLEKEKTLCLSTPSSSTVKHSTSSIGSMLAQADKLPMTDKRVASLLKKAKA
QLCKIEKSKSLKQTDQPKAQGQESDSSETSVRGPRIKHVCRRAAVALGRKRAVFPDDMPTLSALPWEEREKILSSMGNDDKSSIAG
SEDAEPLAPPIKPIKPVTRNKAPQEPPVKKGRRSRRCGQCPGCQVPEDCGVCTNCLDKPKFGGRNIKKQCCKMRKCQNLQWMPSKA
YLQKQAKAVKKKEKKSKTSEKKDSKESSVVKNVVDSSQKPTPSAREDPAPKKSSSEPPPRKPVEEKSEEGNVSAPGPESKQATTPA
SRKSSKQVSQPALVIPPQPPTTGPPRKEVPKTTPSEPKKKQPPPPESGPEQSKQKKVAPRPSIPVKQKPKEKEKPPPVNKQENAGT
LNILSTLSNGNSSKQKIPADGVHRIRVDFKEDCEAENVWEMGGLGILTSVPITPRVVCFLCASSGHVEFVYCQVCCEPFHKFCLEE
NERPLEDQLENWCCRRCKFCHVCGRQHQATKQLLECNKCRNSYHPECLGPNYPTKPTKKKKVWICTKCVRCKSCGSTTPGKGWDAQ
WSHDFSLCHDCAKLFAKGNFCPLCDKCYDDDDYESKMMQCGKCDRWVHSKCENLSDEMYEILSNLPESVAYTCVNCTERHPAEWRL
ALEKELQISLKQVLTALLNSRTTSHLLRYRQAAKPPDLNPETEESIPSRSSPEGPDPPVLTEVSKQDDQQPLDLEGVKRKMDQGNY
TSVLEFSDDIVKIIQAAINSDGGQPEIKKANSMVKSFFIRQMERVFPWFSVKKSRFWEPNKVSSNSGMLPNAVLPPSLDHNYAQWQ
EREENSHTEQPPLMKKIIPAPKPKGPGEPDSPTPLHPPTPPILSTDRSREDSPELNPPPGIEDNRQCALCLTYGDDSANDAGRLLY

IGQNEWTHVNCALWSAEVFEDDDGSLKNVHMAVIRGKQLRCEFCQKPGATVGCCLTSCTSNYHFMCSRAKNCVFLDDKKVYCQRHR
DLIKGEVVPENGFEVFRRVFVDFEGISLRRKFLNGLEPENIHMMIGSMTIDCLGILNDLSDCEDKLFPIGYQCSRVYWSTTDARKR
CVYTCKIVECRPPVVEPDINSTVEHDENRTIAHSPTSFTESSSKESQNTAEIISPPSPDRPPHSQTSGSCYYHVISKVPRIRTPSY
SPTQRSPGCRPLPSAGSPTPTTHEIVTVGDPLLSSGLRSIGSRRHSTSSLSPQRSKLRIMSPMRTGNTYSRNNVSSVSTTGTATDL
ESSAKVVDHVLGPLNSSTSLGQNTSTSSNLQRTVVTVGNKNSHLDGSSSSEMKQSSASDLVSKSSSLKGEKTKVLSSKSSEGSAHN
VAYPGIPKLAPQVHNTTSRELNVSKIGSFAEPSSVSFSSKEALSFPHLHLRGQRNDRDQHTDSTQSANSSPDEDTEVKTLKLSGMS
NRSSIINEHMGSSSRDRRQKGKKSCKETFKEKHSSKSFLEPGQVTTGEEGNLKPEFMDEVLTPEYMGQRPCNNVSSDKIGDKGLSM
PGVPKAPPMQVEGSAKELQAPRKRTVKVTLTPLKMENESQSKNALKESSPASPLQIESTSPTEPISASENPGDGPVAQPSPNNTSC
QDSQSNNYQNLPVQDRNLMLPDGPKPQEDGSFKRRYPRRSARARSNMFFGLTPLYGVRSYGEEDIPFYSSSTGKKRGKRSAEGQVD
GADDLSTSDEDDLYYYNFTRTVISSGGEERLASHNLFREEEQCDLPKISQLDGVDDGTESDTSVTATTRKSSQIPKRNGKENGTEN
LKIDRPEDAGEKEHVTKSSVGHKNEPKMDNCHSVSRVKTQGQDSLEAQLSSLESSRRVHTSTPSDKNLLDTYNTELLKSDSDNNNS
DDCGNILPSDIMDFVLKNTPSMQALGESPESSSSELLNLGEGLGLDSNREKDMGLFEVFSQQLPTTEPVDSSVSSSISAEEQFELP
LELPSDLSVLTTRSPTVPSQNPSRLAVISDSGEKRVTITEKSVASSESDPALLSPGVDPTPEGHMTPDHFIQGHMDADHISSPPCG
SVEQGHGNNQDLTRNSSTPGLQVPVSPTVPIQNQKYVPNSTDSPGPSQISNAAVQTTPPHLKPATEKLIVVNQNMQPLYVLQTLPN
GVTQKIQLTSSVSSTPSVMETNTSVLGPMGGGLTLTTGLNPSLPTSQSLFPSASKGLLPMSHHQHLHSFPAATQSSFPPNISNPPS
GLLIGVQPPPDPQLLVSESSQRTDLSTTVATPSSGLKKRPISRLQTRKNKKLAPSSTPSNIAPSDVVSNMTLINFTPSQLPNHPSL
LDLGSLNTSSHRTVPNIIKRSKSSIMYFEPAPLLPQSVGGTAATAAGTSTISQDTSHLTSGSVSGLASSSSVLNVVSMQTTTTPTS
SASVPGHVTLTNPRLLGTPDIGSISNLLIKASQQSLGIQDQPVALPPSSGMFPQLGTSQTPSTAAITAASSICVLPSTQTTGITAA
SPSGEADEHYQLQHVNQLLASKTGIHSSQRDLDSASGPQVSNFTQTVDAPNSMGLEQNKALSSAVQASPTSPGGSPSSPSSGQRSA
SPSVPGPTKPKPKTKRFQLPLDKGNGKKHKVSHLRTSSSEAHIPDQETTSLTSGTGTPGAEAEQQDTASVEQSSQKECGQPAGQVA
VLPEVQVTQNPANEQESAEPKTVEEEESNFSSPLMLWLQQEQKRKESITEKKPKKGLVFEISSDDGFQICAESIEDAWKSLTDKVQ
EARSNARLKQLSFAGVNGLRMLGILHDAVVFLIEQLSGAKHCRNYKFRFHKPEEANEPPLNPHGSARAEVHLRKSAFDMFNFLASK
HRQPPEYNPNDEEEEEVQLKSARRATSMDLPMPMRFRHLKKTSKEAVGVYRSPIHGRGLFCKRNIDAGEMVIEYAGNVIRSIQTDK
REKYYDSKGIGCYMFRIDDSEVVDATMHGNAARFINHSCEPNCYSRVINIDGQKHIVIFAMRKIYRGEELTYDYKFPIEDASNKLP
CNCGAKKCRKFLN

>MLL2 (*Homo sapiens*)
MAAAAGGGSCPGPGSARGRFPGRPRGAGGGGGRGGRGNGAERVRVALRRGGGATGPGGAEPGEDTALLRLLGLRRGLRRLRRLWAG
PRVQRGRGRGRGRGWGPSRGCVPEEESSDGESDEEEFQGFHSDEDVAPSSLRSALRSQRGRAPRGRGRKHKTTPLPPPRLADVAPT
PPKTPARKRGEEGTERMVQALTELLRRAQAPQAPRSRACEPSTPRRSRGRPPGRPAGPCRRKQQAVVVAEAAVTIPKPEPPPPVVP
VKHQTGSWKCKEGPGPGPGTPRRGGQSSRGGRGGRGRGRGGGLPFVIKFVSRAKKVKMGQLSLGLESGQGQGQHEESWQDVPQRRV
GSGQGGSPCWKKQEQKLDDEEEEKKEEEKDKEGEEKEERAVAEEMMPAAEKEEAKLPPPPLTPPAPSPPPPLPPPSTSPPPPLCP
PPPPPVSPPPLPSPPPPPAQEEQEESPPPVVPATCSRKRGRPPLTPSQRAEREAARAGPEGTSPPTPTPSTATGGPPEDSPTVAPK
STTFLKNIRQFIMPVVSARSSRVIKTPRRFMDEDPPKPPKVEVSPVLRPPITTSPPVPQEPAPVPSPPRAPTPPSTPVPLPEKRRS
ILREPTFRWTSLTRELPPPPPAPPPPPAPSPPPAPATSSRRPLLLRAPQFTPSEAHLKIYESVLTPPPLGAPEAPEPEPPPADDSP
AEPEPRAVGRTNHLSLPRFAPVVTTPVKAEVSPHGAPALSNGPQTQAQLLQPLQALQTQLLPQALPPPQPQLQPPPSPQQMPPLEK
ARIAGVGSLPLSGVEEKMFSLLKRAKVQLFKIDQQQQQKVAASMPLSPGGQMEEVAGAVKQISDRGPVRSEDESVEAKRERPSGPE
SPVQGPRIKHVCRHAAVALGQARAMVPEDVPRLSALPLRDRQDLATEDTSSASETESVPSRSRRGKVEAAGPGGESEPTGSGGTLA
HTPRRSLPSHHGKKMRMARCGHCRGCLRVQDCGSCVNCLDKPKFGGPNTKKQCCVYRKCDKIEARKMERLAKKGRTIVKTLLPWDS
DESPEASPGPPGPRRGAGAGGPREEVVAHPGPEEQDSLLQRKSARRCVKQRPSYDIFEDSDDSEPGGPPAPRRRTPRENELPLPEP
EEQSRPRKPTLQPVLQLKARRRLDKDALAPGPFASFPNGWTGKQKSPDGVHRVRVDFKEDCDLENVWLMGGLSVLTSVPGGPPMVC
LLCASKGLHELVFCQVCCDPFHPFCLEEAERPLPQHHDTWCCRRCKFCHVCGRKGRGSKHLLECERCRHAYHPACLGPSYPTRATR
KRRHWICSACVRCKSCGATPGKNWDVEWSGDYSLCPRCTQLYEKGNYCPICTRCYEDNDYESKMMQCAQCDHWVHAKCEGLSDEDY
EILSGLPDSVLYTCGPCAGAAQPRWREALSGALQGGLRQVLQGLLSSKVVGPLLLCTQCGPDGKQLHPGPCGLQAVSQRFEDGHYK
SVHSFMEDMVGILMRHSEEGETPDRRAGGQMKGLLLKLLESAFGWFDAHDPKYWRRSTRLPNGVLPNAVLPPSLDHVYAQWRQQEP
ETPESGQPPGDPSAAFQGKDPAAFSHLEDPRQCALCLKYGDADSKEAGRLLYIGQNEWTHVNCAIWSAEVFEENDGSLKNVHAAVA
RGRQMRCELCLKPGATVGCCLSSCLSNFHFMCARASYCIFQDDKKVFCQKHTDLLDGKEIVNPDGFDVLRRVYVDFEGINFKRKFL
TGLEPDAINVLIGSIRIDSLGTLSDLSDCEGRLFPIGYQCSRLYWSTVDARRRCWYRCRILEYRPWGPREEPAHLEAAEENQTIVH
SPAPSSEPPGGEDPPLDTDVLVPGAPERHSPIQNLDPPLRPDSGSAPPPAPRSFSGARIKVPNYSPSRRPLGGVSFGPLPSPGSPS
SLTHHIPTVGDPDFPAPPRRSRRPSPLAPRPPPSRWASPPLKTSPQLRVPPPTSVVTALTPTSGELAPPGPAPSPPPPEDLGPDFE
DMEVVSGLSAADLDFAASLLGTEPFQEEIVAAGAMGSSHGGPGDSSEEESSPTSRYIHFPVTVVSAPGLAPSATPGAPRIEQLDGV
DDGTDSEAEAVQQPRGQGTPPSGPGVVRAGVLGAAGDRARPPEDLPSEIVDFVLKNLGGPGDGGAGPREESLPPAPPLANGSQPSQ
GLTASPADPTRTFAWLPGAPGVRVLSLGPAPEPPKPATSKIILVNKLGQVFVKMAGEGEPVPPPVKQPPLPPTISPTAPTSWTLPP
GPLLGVLPVVGVVRPAPPPPPPPLTLVLSSGPASPPRQAIRVKRVSTFSGRSPPAPPPYKAPRLDEDGEASEDTPQVPGLGSGGFS
RVRMKTPTVRGVLDLDRPGEPAGEESPGPLQERSPLLPLPEDGPPQVPDGPPDLLLESQWHHYSGEASSSEEEPPSPDDKENQAPK
RTGPHLRFEISSEDGFSVEAESLEGAWRTLIEKVQEARGHARLRHLSFSGMSGARLLGIHHDAVIFLAEQLPGAQRCQHYKFRYHQ
QGEGQEEPPLNPHGAARAEVYLRKCTFDMFNFLASQHRVLPEGATCDEEEDEVQLRSTRRATSLELPMAMRFRHLKKTSKEAVGVY
RSAIHGRGLFCKRNIDAGEMVIEYSGIVIRSVLTDKREKFYDGKGIGCYMFRMDDFDVVDATMHGNAARFINHSCEPNCFSRVIHV
EGQKHIVIFALRRILRGEELTYDYKFPIEDASNKLPCNCGAKRCRRFLN

>MLL1 (*Danio rerio*)
MAHSCRWRFPARPGGSSSSGTGRKAGRIRVNASLLISAGTNPNANGLGPGFDAALQVSAAIGSNLQKFRDVLGESSGSSSGEEEFG
GFTTVSDNRRLHSPGRTSIGSITPDKKPRGRPPRTPAVQRVGTDTETAPLPVATSPTEKLKRQPGRPPGTREKKRGRPPASVSQRT
WQHSGHALPEEGREVPQECSSSPVHSKEGVEENKEKRQTPLGSGHHQGSEAKLHKVSRESKVTKLKRLREVKLSPLKSKLKAIVRK
TVTVPGKQRRKRGRPPSAERLKAEAAAAAAAAAQAANASMAQETSTTAPRTAKKKAFRVRRSQDLGARTPHELRASHTADEHTHSDS
QDSPTAADPLTPTKVGRPLGLRQSPRHIKPVRVVPPSKRTDATIAKQLLQRAKKGAQKKKLLEKDSVGTQGKAGLEAGKHRRRTQL
TNIRQFIMPVVSTVSLRIIKTPKRFIEDEGSFSTPPPHMKIARLESALTAPAPQPATPSTPALVSTAPSTSGTTATPGSGSAVESL
PPPPPPVSTGSTTAIAASLLNSSCNNSTSNGRFSSSAASCGSSAVSQHSSQLSSGESSRSTSPSLDDSSCDSQASEGTQALSEEAD
HSPASQGETEASLHHASHPPSPTSEPEPDHIVLEHSRRGRRGQSHRRGAVVARGRGNLIIGRKQAIISPATGVSQAGSQQASSTAS
SSSSPPPPPLLSPPQPPQTASSNAAEHHSHSPWMMSHSIAPFLPTSSILSSSHDKRRSILREPTFRWTSLSCAENKYFSSAKYAKE
GLIRKPTFDNFRPPPLTAEDVGLMPPVTGGGGVTSGGFPAPGGAAGTGTRLFSPLHHHPHHNHHQHSSSRFETPLQKRTPLLRPPF
FTPSPAHSRIFESVTLPSSSGSSPGSLSPLQVSPTSSKKKKGSRFSRGQPRSPSHSMITRSSQSGVPTGKSSEQSIISSSVPITVT
GNSSPLPGVAVSPLAASALTQASFSGFPSGSIGLTSHGVSDGRRAAGGLSVSGNSASSSQLFPLFTPSPQASGGGTGKAGKERGIS
ATRDTGTKEKDREMEKSRERERKENKRDGRRDWDKRGKSLPSEASPSSISSFFGLEAIEESLTQKRTPGRKKSVTVDCAEASPSDSA
AVQAVGSLSSKGRLTKKGRPPEKSIESEGVEREKDKEKLSALTQAGQMGKPPTTTSIDSILDRAEKQPVTDRRVVRLLKKAKAQLN
KIEKRELQPGDQPKLPGQESDSSETSVRGPRIKHVCRRAAVALGRNRAVFPDDMPTLSALPWEEREKILSSMGNDDKSSVAGSEEA
EPPTPPIKPVTRQKTVHEAPPRKGRRSRRCGQCPGCQVPNDCGVCTNCLDKPKFGGRNIKKQCCKVRKCQNLQWMPSKFLQKQAKG
KKDRRRNKLSEKKELHHKSQCSEASPKSVPPPKDEPPRKKSETPPPAQGDDKQKQTQPSSPSSPASSPKDPLLSSPPDDHKHSLTS
LSSACRKERKQQPSSSPTFLHAAPSSPPAQSQHSLQQPCQMPAKKEGLTKSQSHTEPKKKSQQQSQPSSATDTAPDAKLKKQTTRC

VQPLKPKPKEKEKQLPKPDSSTLNSQSTPSTGGTAKQKAPYDGVHRIRVDFKEDYNIENVWEMGGLSILTSVPITPRVVCFLCASS
GNVEFVFCQVCCEPFHLFCLGEAERPHDEQWENWCCRRCRFCHVCGRKYQKTKQLLECDKCRNSYHPECLGPNHPTRPTKKKRVWV
CTKCVRCKSCGATKPGKAWDAQWSHDFSLCHDCAKRLTKGNLCPLCNKGYDDDDCDSKMMKCKKCDRWVHAKCESLTDDMCELMSS
LPENVVYTCTNCTESHPAEWRTVLEKEIQRSMRQVLTALFNSRTSTHLLRYRQAVMKPPELNPETEESLPSRRSPEGPDPPVLTEV
SPPNDSPLDLESVEKKMDSGCYKSVLEFSDDIVKIIQTAFNSDGGQLESRKANSMLKSFFIRQMERIFPWYKVKESKFWETSKASS
NSGLLPNVVLPPSLDHNYAQCQEREEIAKAGQSVHMKKIIPAPHPKAPGEPNSLMAPTPPPPPPMLIHDHSLEDSPVIPPPPGVGD
NRQCALCLNYGDEKTNDCGRLLYIGHNEWAHVNCALWSAEVYEDVDGALKNVHMAVSRGKQLQCKNCHKPGATVSCCMTSCTNNYH
FMCARQQQCAFLEDKKVYCQHHKDLVKGEVVPESSFEVTRRVLVDFEGIRLRRKFVNGLEPDNIHMVIGSMTIDCLGMLTELSDCE
RKLFPVGYQCSRVYWSTLDARKRCVYKCRILVCRPPLSETLNKNIAAQEENHTVVHSPPPVSAVDTFLPGPIDSTKPSNVPSTPKP
RVYFRNRHPSFPPCHRSPSTRPLPSPDGFNNTGHEIVTVGDPLLSSSLRSIGSRRHSTSSISAQQPRQKVSSPPQGGTVYSQTGNS
SASFMSSTSKEPLTKDTDKGRVSSGETSFSREPNSINIGAQRRLSFGFTERVDGSKEATKKHSDGESLKSSQPASVSQVSPPLGTA
VLTGHQRASGGIKNEKGKQATKDNDLPAGATFMSSHPLAMLPKDKANPNKEGNMTSMAALKDTVKTGSPQRIYNKSGSRKSHDYAS
GPAAVVAMKPLWSSGAKLGEEDIKRGFQASAGITGSHGTSSTKEKHSKVKMNVSRDVSKERKETPQNRNAVLNSNSKSSNVKTQGQ
VPPPHNISNKATALSSNTGSGTVEVNKFDQKEVEKPLKSKERFSFEKKHTSAMDAIQPKAGSERSIRPPQVHPKSSKEVPLVGKKH
TERLSLMSQKMDPNRTKAVSISPNTQTYTSVTPSNQGPQRRSSRAMVFSPSASSESSESDSHIHPDDSEEHLMDHQCADDGEDNNL
EDEGSVDKHHEEDSDGSAGSAKRRYPRRSARARSNMFFGLTPFYGVRSYGEEDIPFYRSGEISMKKRTGSSKRSAEGQVDGADDMS
TSSSADSGEDEEGGIGSNKDTYYYNFTRTIINPSSGLPSIAGIDQCLGRGSQIHRFLRDQAKEHEDDSDEVSTATKNLELQQIGQL
DGVDDGSESDISISTSSTTTATTSSTQKGSTKRKGRESRTEKSNVDSGKEAVNTTSNSRDSRKNQKDNCLPLGSVKTQGQDPLETQ
LSLTTDLLKSDSDNNNSDDCGNILPSDIMEFVLNTPSMQALGQQAEAPSAEQFSLDESYGVDVNQRKDMLFEDFTQPLANAESGES
GVSTTIAVEESYGLPLELPSDLSVLTTRSPTVSNQNHGPLISETSERTMLALATEESEAGKSKKKTRTGSTVSSKSPQEGCADSQV
PEGHMTPEHFIPPSVDGDHITSPGVAPVGETGNQDMTRTSSTPVLPSSPTLPLQNQKFIPATTVTSGPAPITSSAVQAAASQLKPG
PEKLIVLNQHLQPLYVLQTVPNGVMNPNAPVLTGLSGGISTSQSIFPAGSKGLVPVSHHPQIHAFTGTTQTGFQPVIPSTTSGLLI
GVTSHDPQIGVTEAGHRHDHAPNVAMVSSASTITPAPSMIPSGHGKKRLISRLQSPKSKKQARPKTQPTLAPSDVGPNMTLINLSP
SQIAAGIPAQTGLMELGTITATPHRKIPNIIKRPKQGVMYLEPTILPQPMPISTTTQPGILGHDSSTHLLPCTVSGLNTSQSVLNV
VSVPSSAPGNFLGGSSVSLSAPGLISSTEITGSLSNLLIKANPHNLSLSEQPMVLHPGTPMMSHLTNPAQTSIASSICVFPPNQSI
TVPVNQQVEKEGTVHLQHAVSRVLADKTLDPNVSSAGQVALAPNPISQELNKGHVVSVLTQSSRTSPISRPQHQHQASKLPAGASS
VAFGKGKHKAKRPRPCPDKSSGKKHKGLHSDTPTVDTSAIQLSYIKGDQELSSPEPMDTGQSNETGSKKRDSTTMTTNSSALKRKT
VDAVDEKPSTAGLPSKGDGTGNKAFSVDTPDQRDSGRDSSLDHKPKKGLIFEICSDDGFQIRCESIEEAWKSLTDKVQEARSNARL
KALSFDGVNGLKMLGVVHDAVVFLLEQLYGARHCRNYRFRFHKPEETDYLPVNPHGSARAEVYHRKSVLDMFNFLASKHRQPPVYN
PQEEDEEEMQQKSARRATSTDLPLPEKFRQLKKASRDAVGVYRSAIHGRGLFCRKNIEPGEMVIEYSGNVIRSVLTDKREKYYDDK
GIGCYMFRIDDYEVVDATIHGNSARFINHSCEPNCYSRVVNVDGQKHIVIFATRKIYKGEELTYDYKFPIEEPGNKLPCNCGAKKC
RKFLN

>MLL2 (*Danio rerio*)
MMAAAGCGSATAAAVVGAGGVGAARGRFPGRPWSSRSRLRSEKRWQLGRCGTDLDEISANGGPRPTCLVLAPNEDQPLSRLLGLEA
SYKRLGEAGYSSSGSEEGSEFKGFEVQNENRSIPVRAGRNPCKRGRVSVKSSKRIRQRLALPVSPLVESTGSHDPQNDSGTVQQLH
SQESQKDTSVDDGGGGGGGDDVDGALEKPTRRKRGSRKRETQTTPRITIKLVAKKTVVTVPTLRRKPGRRPKEVKTEVSESQCAQ
ISSAHVKLKTKTPASESSAGDTKVEVCTRRRGRSADQKLETTLQKEPKKEANDQKLKLRKSTRKCQNPLPSENKEPKPADKSAQEA
SEVNKLVIRRQRTRSSIAPQESGDQSLAESQLEVSTIDESRSPSKKKKKKKKKKNRKESERNKDSSAKDHGHSADDPLKLAFGND
FVGIPGLKLTRIRNPKARSRKKRCKFVWTLTLVKCKGKDSHIQENLSKTQDPQNQEDLGKTQEIESQDSRVQKSPSKADESPSIEA
ELPPKCDKEVDVSEEAEPASAPSSTLETDSLQESQKVENTASTVIPESPENPVAIDKDEEIPEKENEEVLERSEIEAHPEEATQTE
PEDLINKDIVPPLQIKVVSSPGKNNSLKQSFLIQQVTTSPEITEPEKVEEHQHLENSENMLQKTTTSPDHKVRNKLHQSAARKKRA
RHRHWTIYRRKGRTNPSQENSASEIATETEPQEEELLVLPTEVPAPVESKPKRHLDKITPKRGKRSVPPTAEPLPEAAEPQVIEQT
VEAPSTDLLPVTEPQPEIKLEAPSIENEEVLVLEPSVKDHPQVRRPLKRRIKHAKRRRRSLIGRRVKNQKLQKKVLVEPTDGCPPT
TESEGGEPKLVGLFATKYKRKRITSIQSLEARRKKAKFLSKQLENTSSQDQLSTQEESDSFSVKTDQDHDIQPNKAKIVKNIRHFI
MPVVSARSSRVIKTPKRFMDDLGMSDLPRKNSQKKLGLQPKSKKSDPSERGEPGCFQSQDEEDNQSEIDLDLSSTWESKPKSPDDI
DLFSEEKQSGKRRSLLRDPGFKWGVLEPASEDVYSFGEDLDKELETLLSAKDFTVDSFLDPVQKKKSTKFKAQSSKLRLYKKLKLK
QSLARQKKIPKEKISKVVSSPAKAPKKELEEGASLIKSLREMKKEKAKLKIEDLNTPGVVRKVSICVRALSAKLLAQHQAKDIQED
DLPDEFSIHTDISLPNKCKDFDKLHLTESEAPVSEDPEKHDVDVEEKNASQRPRLTGANKRMFNLLKKAKVQLIKIDQQKQLKSSG
LLPEGRITATKRQRRKPKERPVKACPNKTEATPEQEPARGGPRIKHVCRAAAVVLGQPRAMVPDDIPRLSALPLHERSDISPSPAE
KGVESHSDPESPVLQEHRTPKFRRGRSHMFGHRSRRCGECKGCLHEEDCGRCINCLDKPKFGGPNTKRQCCVYKRCDQVEERKALR
LSGKLQKGHGKKRRPSVSVYYSSNEEGEAGEGNGRMFPSSLSDSPSVRKQPRRHVTPRSYCDLLDYDSDFDWIGSSNSASPGRRRT
PGLRNPDFLSFDDFIGDPFDEGVRHRRPGFHKVPTIKRKVEKSPQEQTPPSVLAALANGFAEREQEPAEPTYKIRVDFKEDCSVQS
VWATGGLSVLTSVPLMPSCVCLLCASKGQHEMLFCQVCCEPFHRFCLDPSERPLEENKENWCCRRCKFCRVCGRKNKESKPLLECE
RCQNCYHPACLGPNYPKPNKRKKPWVCMTCIRCRSCGVTPGKSWDTEWNHDKGLCPDCTRLFDQGNYCTMCFKCYEDNDYDSQMMQ
CSTCNHWVHAKCEGLTDDLYEILSSLPESVVYSCQPCLKEQPDGEAVSGAGWRELLDQELRTGLERLLSCMMGSTVTQHLVTCKKC
STLGSSNIGVEELPICDLSAIGKKFEGGHYTSLKSFHEDVVKVLRRYLEEEEVLPEDQKHTALARSYYLTLLEDVYGWFNSQDPKL
WDPRSKHLPAGMLSHAVMPPNYEHVYAQWRERDKALDGNLQLTGRHADVKMGDEEDLPSHLANMQKGHLMDMRALRMNLKGKKSRQ
GKSDLETGWSKDDERQCALCQKYGDAKASDAGRLLYLGQNEWAHVNCSMWSAEVFEEDNGSLLHVHSAVARGRLMRCERCNRTGAT
VGCCLSSCQSNYHFMCARFRNCVFQDDKKVFCYKHRDLISGKMITGQGFEVLRRVYVDFEGISLRRKFLTGLEPESINLMIGSLLV
EKLGKLSELSACDGKLFPVGYECSRWFWSTVNPLRRCKYTFRVTEVRPAVQEKQVEESPDHGENCTIAHSPCPQSECEALELGLSS
TNPSTEANFTSGSVFAKPDLGARPKITSYPQNRRPAGGLSRPLPSPGAAPSKPHHILTISDLEDTRRPRRHSPHTQNSGTRSHMTS
PTLGSPAGPIPSRAGGSHHPKSSQTTSPMFPLGATENLLTSSSARPVGRSASSVRGTGILTPHSTSGLFSQSTKQGGIGSTPCTNR
SLTSPTHLSTRPRPFDFNPIDSLELPHNFIASPQEHLSETGALPLGDALSMPKVNHLCSDRDFPYTSIHLEADLTVAPELRTELEI
EESVVNEGVAMNCSAQIVVEGEDNQEDFWEPDDNKSKMSSQRLSCTVDPCREDWGNSSSDEDMENYFNFTRTVVTCKAQRDSSKKS
PSVRQISQLDGVDDGTESDTSLANSDNHAMKLSNQTQKPHKSLSAVPHEQTASNGLCLTFPSPLSDISSINCCFESVPNSEVLTDQ
DSFGLTQILESSSGVCGPQQDTLINFVESKSSNPTFSTYGNDYQSLAEPETPLILERCDPLSPNAHFTELVPVQDSPCDPQESKEL
YLDPSSGHFVSSVDDATVYQNSHLQDDSLNLNQVNGIQFSLQQNPVPSTELLKPAPLLEPSSSVQVPFVAFPNQTTPTNVSLFPVQ
TSAGKTILPPMESFRTKLNPHPEQRHPAPSVPGGALPYVPANIPQKPESMLSVPSATGVPLGTYGPISGTVSVPLHPTYMQPIGST
TVGVVSSSTSVTSQPSAQCQAMPSAIQPASAPCVINGYNTMPMQRDSAPGRTISINFSTPRSAVEPQQQLVTQALPGHAILTVKEV
GGPNVDPTPHVLLVNRLGQIFVKNPESNTFQLPTPNSPSFSCVTQIASLLQSNALSATLAAAGSLPAVAGATIPTQVSHIVTPLVQ
TSNTITQLLATNNNEAAPPPPADVKKPRRNTKSTGEKVKKPRTKKESSTPRRTKSTSTSGKQSTGQVITSSDKFNQLDSAQAIINQ
AMAGYYDPNKTVSNILPSSSPCNTKPVGQTQTTVSPVLPETLLVKPEKKSSPPVTPEVAPRPKQVRMKRVSCLSDRIATKKSKSDF
LDMERPSALEEPHKSSSVSASRYGVRIKTPTVKGILDLDKLNEEPSSDSESARPGPWDRLSLPRTGESNKPSTWDPANRSGLSDLN
KYSGMLSCSDDELSPSDGEEHVPLSRDQPHLRFEIASDDGFSIEADSIEVAWSAVLEGVQEARAAAGLRLLTMCGMSGARVMGMLH
DAVVYLVEQLQGAKRCHRHTFRFHKQASQDDDLPVNPSGCARSEVYLRKSTFDMFNFLASQHRQLPDTDLYDEEEDEVLLKSTRRA
TSLELPMAMRFRHLERTSKEAVGVYRSAIHGRGLFCKRNIEAGEMVIEYSGIVIRSVLTDKREKYYDGKGIGCYMFRIDDFDVVDA
TMHGNAARFINHSCEPNCYSRVINVEGQKHIVIFALRKIYRGEELTYDYKFPIEDASNKLGCNCGAKRCRRFLN

>MLL3 (*Danio rerio*)
MSSEEKSLEPSEQGPSPPLSSAGATPSGSPAPLDRRPRGRPRKDASSVTVTPKSRRKSRSRGCVQADDEDSMDATDVTPTPDIEIT
PEIDVVEPETDDEFPEPLDLSPRDERPLSPLFQRSVSEDSAGSSASTTRKAKTREKLCAFCYCGEKSLLGQGELKLFGPTPGYVPL
HIRNRRGSSEKDDDYHDDDNDFDDNDNNRSPGHRGSGLKKVSSPDSGLHSTSAPGSTDPAAEEPGRKFWDELGQIGLPDEINVAFL
FDPTGQCCAHLRCAAWSEGVCRGEGQSLLYVDKAIDSGSTECCAYCKRLGASIKCCEESCGRSFHYPCAAAAGTFQDIKTYTLLCP
DYIQQALVRSKDGVKCLVCDSSGDLLDQLFCCTCGQHYHGTCLDVTVTPLKRAGWQCPECKVCLSCKNPGDDSKMLVCDMCDKGYH
TFCLQPIMETIPTNGWRCKNCRTCAQCGMRPNGQWSSNGVMCEGCQQQHESALLCLICGQSQDSGSPADKHTCCTCKRWVHVDCVR
QEGTNPDTQLQDGYTCRACKQVEEETRHSQPVSSQEASLTDPEPAEDVFMSVEEVQEAEGEVVLEPEPSASPLATLPDNFLSSTEP
AQEIQVTVTLTEGPDPPPCSSTAVGVSSFSFQASTTTEDHIQESSLIGALEMGKESEDKLPETTEEEPEPNVSKALPALPTKETSP
MHTLAQEEEPMDVYSSDPYSMKDKRMTEAAVEDPEQKQEHEKPEVPDPEVTSIIPEPESVPLIPESSPTMEMDTTPVSERSFSSFS
SPVEDKCLVLKHSPGLSSEGSIKLTQSPFSTEASPRGSSQNQTPTHGLNAAGFSPNVSNTTFFPLTPKIGMGKPAISKRKFSPGRP
RVKQGAWINRSAMSSPSWSPERSDAWDSPKTRPMHGLPVWTMRLGRGSGFPGRRRPRGANLSGRGGRGRSRLKMESSPVVTPGLSF
LDSPYVVKEEEENSMHNTVVMFSTSDSFTLKQDMCVVCGSFGRGAEGRLLSCSQCGQCYHPFCVNIKITKVVLSKGWRCLECTVCE
ACGQASDPGRLLLCDDCDISYHTYCLDPPLQNVPKGSWKCKWCVLCTHCGATSPGLRCEWQNNYTQCGPCASLTVCPVCTRSYREE
ELILQCRQCDRWVHGSCQGLNSDEDVENAADEGFDCTLCRMHTVPTPGAAVPKLVEISEPPMVAQILTKVKENEFQRTYTQDGVCL
TESGLCQLQSLSVQTTRRKRPKPKLKLKIINQNSVAVLQTPPDAQSELFRDGDLDDSREGDPMDCDGKSDSSPEREHCDDHLKGVD
DSNKKRKRKPYRPGIGGFMVRQRSKAGQGKGKRSLSRKDSTSSLTENLTGRDDGWNEPDTPVDGAPPIISEPLEKVKKRYRKKKTK
LEEAFPSYLQEAFFGKELLDKSKQTKSALETGAVEEASHTDSKSQSSSFLDPLSDPLLSATSTPVSSKAGTLPNVEDPLTELVLHS
DDDLLGMLSDDLVKPGHESGLDLCPFQVDSSPSPFAGLDIGTLPDDQSVPPHPGPARGSRLQEEPLDVILSPELDKMVSDGAILSK
LYKIPELEGKDVEDLFTAVLSPNASQPPQLPHPPNTAPNASAPAMSLQNQGGVGGMFPRMPMMNGMMGPNTHFPPNPMIPVAGVSG
TFSSVHRMPFPENVREKKFSQTSGDVIGPWSSPVPVPGSVPPPEVETDTMSNAQRSTLKWEKEETLGEMATVAPVLYTNMNFPNLQ
DEFPDWATRVKQIAKLWRKASSQERAPYVQKARDNRAALRINKVQLSNDTMKRQQQQQAVEVFDPNVSMDTDLLFKDPLKHKESEQ
EQEWKFRQQMRQKSKQQAKIEATQKLEQVKNEQQQLQKQQQQLQYSGDASNSPHTSSGGNSPLQALTPKDSFAMTQMPGTASSGTS
ADDVFLRPQPPPLSSTSKTPTQDTQYPQGPSSQPQSPQMFSPESSGSRPSSPYDPYGKMVGTPRPPPSGSSTPRRGSVSDMQERCR
PSSAHESFGSPTTTSTDPYAKPPDTPRPSDPFVKPMCTPRLGSLSEQQGRQLMGNVTPGDNFSRPGQRSEAYQRMAHSRMVLSDPY
SRPLLTPVPGSNESGSVPVFKTPMPPSGLQQDSFNGGQQGMRRGSSDTFPHVQHTDTYSNQPLTPHPSMADGINNEGRMMRPGQGS
HFAQPLSMGRHPQRDPYSQAPSTPRQDYPQQMSDPYAQMPGTPRPHDPYAQMPGTPRPNCDPYPLPPPTTRTEPIDKFSQSQPSSR
RQSPSHAIDPYAQMPGTPRPSPGERYSKSPGSLRSSDPFSQPPGTPRPIKNEIYDQQGTPMPILPDPYSQPPGTPRPGAGLEGFSR
LNPRLNFITHQTDPLGLQAQSRTQDPFSRSCGTDSQTPKHPGISDENISIQQTIPNQTPIHDPFEQAPLTPCSQTGERSSQDVIVS
TGNPSDSQNTMQPPLGETEEKIKQRQRLRQLILKQQQEKSAIRKEKGLQEATASAAPGTPRHWSQDDSGPQSDIFGRPPPPYPGTM
RSSPIHAGQRFPGTFLADQCGPMPNEPQCSRQPFPRDLNNVGMRPQGQRFAFPPGAPNQEQFLRPPHQMQGGPLVDNVPPHMRRHL
SVEGVRSMGNNHQIGLPQHFPPRGIPMQQHNIMGQQFIELCHRAPENRVRLPFGPPNVLDYPNHPQRQPHGFTGNMVFTGNQGPRF
IEPLLGQPRQESISHLQTSANIDNLHHNTTQQQTPLTAQVSHSVRTDSVTPAESPLHPDTTSGSTEEIPLPCRDGVEKLETDDSVV
KDLDDVEVKDLVDDDLDNLNLDADDGKDLDLETNDLHLDDLLESGKFDLIAYADPEMNLEDKKDMFNEELDLSDHIDDEHGDTSDL
QKALSEKRNARCEGGSAQSTLEGEQGLNILKPEKTLSNEVLDHSQSEITIKTEIQGLATSVDDHRPASSKSSQNDVSGLFEDSCED
AHCAPTLSSLMVKEKLEDTRSESIISAIQGDPNSQTNLDSNMSMVMLQGQSQEHRIKSGVTMGHRVDSALTQGTLLSNSNMMHSGN
QQVPLQGFGPSQGQHIHLNMPAVGTQQPVLINPQNQQGIFSPGIPVQQGSLMQQSQQNRPLLLEEQPLLLQDLLDQERQEQQQQRD
MQAMIRQRSSDPFFPNIDFDAITDPIMKAKMVALKGINKVMVQNSMGMPPMVMNRMQQVPASHAPDTPAFSQPFFGPDGKLPPQIA
RPTPPNFGPGFVNESQKKQYEEWLQETQQLLQMQQKFLEDQIGAHRKSKKALSAKQRTAKKAGREFPEEDAEQLKNVTEQQSVVQK
QLEQIRKQQKEHAELIEEYRVKQQQQQQRLQTGMMGPMKTMPVMQGQPPGMMPVGPPMAQPMMGSMLPMQHPHFGKPSRMPSMPGW
HPGTTGPISGPRMPPHLPPQMALPNPAQPLQAQPPAPIAPGGKPGQSGTVAGGNGGGASSSAPHVKFDDNNPFSEGFQERERRERL
REQQEKQRVQLMQEVERQRALQRLELEQQSLGLGPDNNGDSLAQMPFYNSDLSQEFMQTPRPQQQQQQLGPLFPQQGNVPCDFVGQ
GSAFMQGGEHRPIPGNGSFGPEMGPNFQPKNPMMPGFSAGRPRPPGFGGTRMPHHGGGEPSSFGMNSTTPLPPNFPGSGQSLIQLY
SNIIPEEKGKKKRSRKKKRDDDADSVKTPSTPHSDITAPLTPCVSDTSSTPTRSVAIMGDQDTSEFSNPLGSMTGLPSSELEISLS
GSTPLRRQTSVSSEAEHGLLHDIKLEKLESSVCQKAEGEAPVGHSAEETGIIKTEEVNEIISHNRTSQSPAQSTKGEPGNELLKHL
LKNKNTPPPPSAPLMHQMSNESLRSEDEGLNDSKGYLRQDSTDSMSGINSRLTDLEPSLPEQCKKKQRNKRTPKNGVDKPPSRWKK
RKKDEDERQAIYTNTDTLMTQLKQQLSMLPLMEPLIGVNFAHFPPYGSGQLNGENRLSGSFGSASLDGVSDYYSQLIYKQNNLSNP
PTPPASLPPTPPPVARQKLLNGFATAEELARKAEVMVSKSLGQKQLTASFKAEDDLLGHTITQGPKTVDVPASLPTPPHNNQEDLR
GQEHCIDRDTPDSFVPSSSPESVVGMEISRYPDLSRVKEEPPSPAVSPVIPILDCKGSEARRWEVKTEPPSNFFASQFGQLSTCLR
TDLVPVSITLNPVAAMNIQGVVSALANMLCVSIPGNYEVTRSPGPEHRSSIDMLSGLKVPHPTSLFHQQPGMRLHRPPLPHQTFSD
GTAMVRRDYGQMHGSSLGSKPQWCSHCRVVVLGNGVRKSTKELPMNKLGHGGAVESLVFCSNNCLVLYSASMQSKLLDGKPNFPPP
SESGDKQNPTKVLHQYNNNISDLDVHCLAQLQPKQSPPPTPIISFPSITSEMLRGDCKPDTLKVTVKLKPRPRAVHNGGDEPRYLS
GKRWKALRWRRWSVQIIVPKGNYHLPVEEEIDALLKKLGTSLRPEPMPKDMRRCCFCHEEGDGLTDGPARLLNLDLDLWVHLNCAL
WSTEVYETQAGALINVELALRRGLSVRCAYCQQTGATSGCHRLRCTNAYHFTCALKAQCTFFKDKTMLCHLHRPRGTSLIGGGLLG
TEHELRCFAVFRRVYVQRDESRQIASIIQRGEREHTFRVGSLVLHAVGQLLPQQMLAFHSMTAIFPVGYEASRIYWSMRHGNRRCH
YVCSINENDGSPEFVIRVVEQGYDDLILTGSSPKGVWDKVLEPVAERRNETGFLRLFPVYLKGEDLFGLTVSAVTRIVESLPGVEA
CERYTFRYGRNPLMQLPLAINPSGCARSEPKACTHIKRFVLRPHTLTSTSKALQSPGNPAEIGVPPYSKHFVHSKSAQYRRMNAEW
KSNVYLARSRIQGLGLYAARDIEKYTMVIEYIGTIIRSEVANRKEKMYEAQNRGVYMFRIDSEHVIDATITGGPARYINHSCAPNC
ITEVVALERGHKIIISSNRRIQRGEELCYDYKFDLEDDQHKIPCHCGAVNCRKWMN

>MLL4 (*Danio rerio*)
MDEQKSNCEENDSETAADDGTPAKESVQVEGELNAPVAAEDISGTTVDTSSSSLETVRACALCSCVERSLHGQGELQCFGPSSDRP
SLESSVSTLPAAGNDDLSSIGFSESTCLASLFDDTGSCWVHHWCAVWSEGVEQTEGEVLINVDKAVVSGIQRPCDYCKRMGATIRC
RAEGCSRFYHFPCSAASGSFQSMKQLALLCPEHIDKAEEIAGEEARCAVCDSVGDLSGLLYCTGCGQHYHDACLEISATPLQRSGW
QCPECKVCQTCRQPGEDSKMLVCDACDKGYHTFCLLPAMDSVPPDSWKCKRCRVCIDCGMRGLKLPGSEQWFESYTVCEGCQRRRT
SVCGVCSKATEPSVSLQHHCAICHRWVHSDCASLSGPPEEKCICVLCKEIPDEVSVEAEVSSMQVDEKTDLEVPVELGTELEVVEI
AEVKESIMEEALRPEGTEGSTEANVHTEVSTEQQPLLATVSDVDAKQKEDETTSETVVAEAVEQKPEDTQDPPIKDSPDSPVSDTK
EEAQSDVAGTNIKRQDESQDKPPQSETELDGTVKSTKAKPPSSPCSQLSQITLPVAEEEEPVAMDLEEMGSQEETINRTPPSPAPH
TEMPTASDTDDRMGQSEDEEEEDEQGVEESLDDSLLKDESQCEEDQARKELQEDMKPELVLDETSNLSHGDESSSGFLGSPAEADS
QMLSMELSLVPAGRTRSDSLLTETEDSLPFDPLKPDGEKVKRRGSPGRSRMKQGRGSGFPGKRRPRGGGSGGGGGRGRGRSRLKAM
ASCIDAFLLSMTADTGISKEEDEEEDDTMQNTVVLFSNTDKFVLQQDMCVVCGSFGQGVEGQLLACAQCAQCYHPYCVNSKITKMM
LRKGWRCLECIVCEVCGKASDPSRLLLCDDCDVSYHTYCLDPPLQTVPKGGWKCKWCVCCMQCGASSPGFHCEWQNNYNHCGPCAS
LVTCPVCHENFMEEELLLQCQHCDRWVHAVCESLYTEDEVEQASDEGFACTACTPYVPRPVVESPMMSAFKIKEPEPQFYRLEGVW
LTETGMSLLRSISMSPLHKRRQRRSRLGTMSIDELKEGDADGDEGKGGAEAMECEPKSEAPGSPDRDGGADRETGPEGGIEGMADC
ETLKGGVEETDDGKKRKRKPYRPGIGGFMVRQRKCHTRMRRSLSASGEGVPEGQAPETKLEEGPHPDIPLDVPEAKTAEGEGEQAK
KRRGRKKSKLEDMFPTYLQEAFFGKTLLDQSKRAFLAPPTPRHTQGLLRPPLPGLGARTAASESVESTARPSGLPMELKQEGEGSK
ATGDGAVVGSGATGQLFNKDQSSGAAHGSDMLKNKDLPQGLERQDSEQFFRKVLGPTEGSSVSGQLQPSIKPVQEGDPNQSSCPQR
NLLPGASLSGTLPSTGMMDSFSGLSQSPFFDLRDRGGLFSPDQDEESPWATPSTPATPSTPPTPTEQEGDGLSYNQRSLQRWEKDE
ELGNLSTISPVLYANMNFPSLKQDYPDWASRCKQIMKIWRKVSAADKVPYLQKAKDNRAAQRINKAQKQAESQVCRPIKTEPGRIK

GERPNLHLQIPPPSGSVSTSSQPSSAESPYTLAPDSASSAFFPDGPLKTPCSADLQADPFSKLSSQSPHLQSQPSTPYSQPSSSPL
PASSSGHCMPGPQGATQGRPSSFGSLEMQPGATRRAQQVDPFFRPQQQIPKPQEALGPPESPRPRPGGPGDAQMFSPPHMTHCGDL
YRGQHGAGRQDHASSPSPSAIASSPAGVGPSRADINVPSPRSSFGRQDLNTGSPASLDSGDGLFKAPMTPRMHQGDAGGLSVSHQS
GASPNHPLESYRQSPSTFSDPYAQPPLTPRPQSGDGSSPLPQRLPAAQQEAYSRLPSSPQSQGSSQSPLTPGAISNDVLSVQSPAT
PRFQSPDPYSRPPSRPQSRDPFASLHKPPRPSSVAPEGNSPFRGSPHPGQQPPCSPSVADPLSGKPPAFSCSPGVGIQMGPQQIVG
QQHMTPQQAQVHLQQPQSQTQQVGADFNSRIALPPGNQDTAVRPSDGPHLPNIPGMQDISDMSAAQDPALMSLSPSELEKHRQRQR
LREFLIRQKMQRNSLRQEKEAPVVSNPAQSWQDGEMPAFQQDKSQRAPPPYPQDRAAVTQAALASRMPLPMGVMDEKLCRPPPPGT
PANMDPSALRQAGPVSQGMYPRPQFPAQWQNQGAPRRFPQPSMIEAVPRQHLTTPINNPLNMQSMQNMPNPLTTVPGTGTEAVQSH
VSGAPPQFIELRHNSQRMPPGPQFMQRAPIPRARMFLPQQEMNAAFAQQNLTSSLDASAVGMRDPRLGLQQGGMDLLMQQQSNAPP
PQSQLQPQQVPTTTPKTQISHTTQQLPQTTSIGSQQTGVNPVDLPESDLVDAFAPKGLSDPVAETGVEDDDLGLDLDPDKGDDDLG
NLDNLEANDPHLDDLLNSDEFDLLAYTDPELDQGDPKDVFSDQLRLVEAEVEGPVSSANVEVKVEVKPKVEMASQDTRTTATQPPL
KTENSSAVNTPKVESSDVISLSSVKLEDKGLVEQPQSGQASVKDDISETVSVLLSGTKASNQPDTPTANLNTVRLGGIPFPLPSQA
DTLSFPATAPHPDLGSDPLGLPDSGHHSTADDLDKVESSLEASELPLLIQDLLEHEKKELQKQQQLSALQGGIGSQIHSIQNQQQP
PQTTNQIMMPQHHHPSQGIVPQAGMVPRPPHMFPQQQQRLISAAMPSPPHAAMGQQPGILRGGQPGRLHPGLAPQQQTTQPPHIGK
QQNFFPDKDLDKFMTDDIIDPIAKAKMIALKGINRVLAQDPMGVPPGMNRQQVSLLAQRLTSAPGNEDSQPPPAPGSSKEGEGSNQ
VQTRPNPPQFGIINDAEQHQYEEWLVHTQQLLQMQLKFLEEQIGAHRKSRKALCAKQRTAKKAGREFAEADAEKLKLVTEEQSKIQ
KQLDQVRKQQKEHTNLIAEYRSKQQQHQQGSSMLAPGPSTPAPHMLSKMPGQMIGQPPGGMPQAAMPPGGMPIRMPQMQPFGAAQP
SLPLGAGPVQPPPGFFPQGPGPQAPDPRVLQERQYKMQMVQKMQQQQQQLMMAQQPMSHQPQQPGIMNQQPGAMGSQPAQVLMGSL
GQPNLQQGLINSQQNQQGLVQTPPGMMGNQALGQTQHNVMGQQMTQTANLMAGQPQGMVGNQQIVQQLQRPQVLMDQHGPVGQHVV
RGPPTQLTPQQQSVMAQRMLMSQQQQNAAKSLAQLQQQQILQLKQQAGPSEPQRALEAPSSGGNQGVAENLQQGVQDHQNPNKDGG
LLSPKTPPQQGGPPTPLQMQQGSSGEQQCVVGQPQQQHAFLGQQQTAQPQTNQNIGVAEQHGFVDNQQAGPHLKVQPPQPQMQIGL
QTQTKIQIKQEDQQMCFMAQQQAGQIAAQQPQGVAVQNSGQTMGNITNQQQQALLTQQMGIPRAQQMMMTQRPGAPAGQIRTSINI
QAFIAQNPQLRHLTPSQQLQQIQAMLAQRQLQQGQMLRLQGQAQGQMRPQVLQGPSVGPRIPVIETQQQQHRFGLAGKHGSSTQQQ
PGVIRQQSGMAPQFHQGVIGAMQQPSQGPLPQYINSAQQQQMIQQQQVHQHQMMRGQVPRPIMGQVRNTAPGHLARPMSPRQLLGQ
GSPGSPLSNQQRLVRMSQTPPGTQHLQQQTHGTTALSPFQASPSHAGSPAGSSVPEAGSPSLNKPEPGNMALSSNLSSPSRPEGIV
GKSSPFSQTGGLIPSPLRSPTAKTQSPNTEALSAPFPLNGPSSQVAQNQAPNSGQISGGHTSQPTQMPSEMCKMTLQNIKQEPREV
QCDTGNIVETHSDGIKREISGESVSGSGIVNSAVGSRNQAGDIGSQIQRTETGQQLLQKLLRTKNLQVAAQRPSDGIHNEINGHIN
NKLAMLEQKLQGTPRNMEDLQSLTKRAPIVKAKRTAKAGERGPNSRKKTKKEEVGKTEAIMKQLKQGLSLLPLMEPSITASLDLFA
PFGSSPANAKTQLKGSFGNAILDNIPDYYSQLLTKNNLSNPPTPPSSLPPTPPPSVQHKLLNGVTQVEELAEGQKETEASEDTIGS
VPEEVKSVDILAALPTPPHNQNEDIRMESDDDSDASDSIVPASSPESQLGDETTRFPSLLDPKEEEEERAISPVIPIIPRSTIPNF
SETKPFESVDGKVVASSSQWDKAKSNEVSVTFMLSSAAAKNLNTVMLAVAQLLHIRMPSSYELAFPQSPGRVGAGPGKVPDSSNPG
LKTDASVSNDAEWLKQFDVSLPGCTLRKQVDILALIKQESQEQEDKPAQHCYMTNVSDLDVRHLPVIPVEVSPPPSPSPPPPSSST
VAEPENTTQPEPKPESEATVQTTSSPSASPSAPVKTEILVKQESHCSPVTVEPESHEPAVLPIAVPPPQEQPSAPASSPPLDSPTI
EISLKSIKHRSRPLSDEEQRPKVKKWKGLRWKRLQIVITIRKGGSKKESSREVSELMERLRITLKPDKLPRDKRKCCFCHEEGDGA
TDGPARLLNIDVDLWVHLNCALWSTEVYETQGGALINVEVALRRGLRTRCAYCQKTGATNSCNRLRCPNVYHFACAIRARCMFFKD
KTMLCTQHKLKGPSEEELSSFAVFRRVYIERDEVKQIASILQRGDRIHLFRVGGLIFHAVGQLLPSQMAVFHSPSAIYPVGYEATR
IYWSTRVPNRRCRYRCRISEQDGRPLFEVRVLEHGYEDLHFRDSTPDGIWAKIVHQVAKLRDEAAMLKLFTEHVRGEEMYGLTVHA
VMRITESLPGVESCQDYTFRYGRHPLMELPLMINPTGCARSEPKILTHCKRPHTLNSTSMSKAYQSTFTGEINTPYSKQFVHSKSS
QYRRLKTEWKTNVYLARSRIQGLGLYAAKDLEKHTMVIEYIGTIIRNEVANRREKIYEEQNRGIYMFRINNEHVIDATLTGGPARY
VNHSCAPNCVAEVVTFDKEDKIIIISSRRIPKGEELTYDYQFDFEDDQHKIPCHCGAWNCRKWMN

>Trx (*Drosophila melanogaster*)
MGRSKFPGKPSKSINRKRISVLQLEDDAANPAEPQQPAPESQQPSGSGSGSSAAREKGNNCDNDEDDNAPGGASISGNTASSSAGS
GNSGNGSSSGSSTGSGSSGSGSTNGGSVNGGTHHKSAANLDKEAVTKDQNGDGDKTRGNVSSAPSGKLSAAASGKALSKSSRTFSA
STSVTSSGRSSGSSPDGNSGASSDGASSGISCGKSTAKSTEASSGKLAKTTGAGTCSSAKSSKASSGTTSEATTSGLSGACLKALF
VATPATSTGLACALVSPGGSSQGGTFPISAALLRARKNSNKKFKNLNLARGEVMLPSTSKLKQLNSPVVDNPSPSPPIASGSTPSV
EGGIGVGGVVSPGEDAALKRVLTEMPNEVARDPSPSSCTAAANGAASGKGSASNGPPAMASSGDGSSPKSGADTGPSTSSTTAKQK
KTVTFRNVLETSDDKSVVKRFYNPDIRIPIVSIMKKDSLNRPLNYSRGGECIVRPSILSKILNKNSNIDKLNSLKFRSAGASSSSS
NQESGSSSNVFGLSRAFGAPMDEDDEGGVTFRRNDSPEDQNNAEDDEMDDDDDEEAEEDDENEDDNDEAVSEKSAETEKSAGADE
RDPDEKQLVMDSHFVLPKRSTRSSRIIKPNKRLLEEGAISTKKPLSLGDSKGKNVFGTSSSSAGSTASTFSASTNLKLGKETFFNF
GTLKPNSSAAGNFVLRQPRLQFQADNQQATFTAPKACPTSPSAIPKPANSLATSSFGSLASTNSSTVTPTPSACSICSAVVSSKEV
TQARKYGVVACDVCRKFFSKMTKKSISANSSTANTSSGSQQYLQCKGNEGSPCSIHSAKSQLKNFKKFYKDRCTACWLKKCMISFQ
LPAAHRSRLSAILPPGMRGEAAAREEKSAELLSPTGSLRFTSTASSSSPSVVASTSVKWKSSGDSTSALTSIKPNPLAENNVTFGS
TPLLRPAILENPLFLKISNAADQKLAAAEAISPSLTKKNSKQEKEKVKESEQSEKLLSPTQAGTKKSGAAEAQVEEVQPQKEEAPQ
TSTTTQPSASNGASHGVPQAELAGETNATGDTLKRQRIDLKGPRVKHVCRSASIVLGQPLATFGEDQQPEDAADMQQEIAAPVPSA
IMEPSPEKPTHIVTDENDNCASCKTSPVGDESKPSKSSGSAQAEVKKATALGKEGTASAAGGSSAKVTTRNAAVASNLIVAASKKQ
RNGDIATSSSVTQSSNQTQGRKTKEHRQQRTLISIDFWENYDPAEVCQTGFGLIVTETVAQRALCFLCGSTGLDPLIFCACCCEPY
HQYCVQDEYNLKHGSFEDTTLMGSLLETTVNASTGPSSSLNQLTQRLNWLCPRCTVCYTCNMSSGSKVKCQKCQKNYHSTCLGTSK
RLLGADRPLICVNCLKCKSCSTTKVSKFVGNLPMCTGCFKLRKKGNFCPICQRCYDDNDFDLKMMECGDCGQWVHSKCEGLSDEQY
NLLSTLPESIEFICKKCARRNESSKIKAEEWRQAVMEEFKASLYSVLKLLSKSRQACALLKLSPRKKLRCTCGASSNQGKLQPKAL
QFSSGSDNGLGSDGESQNSDDVYEFKDQQQQQQQRNANMNKPRVKSLPCSCQQHISHSQSFSLVDIKQKIAGNSYVSLAEFNYDMS
QVIQQSNCDELDIAYKELLSEQFPWFQNETKACTDALEEDMFESCSGGNYEDLQDTGGVSASVYNEHSTSQAESRSGVLDIPLEEV
DDFGSCGIKMRLDTRMCLFCRKSGEGLSGEEARLLYCGHDCWVHTNCAMWSAEVFEEIDGSLQNVHSAVARGRMIKCTVCGNRGAT
VGCNVRSCGEHYHYPCARSIDCAFLTDKSMYCPAHAKNGNALKANGSPSVTYESNFEVSRPVYVELDRKRKKLIEPARVQFHIGSL
EVRQLGAIVPRFSDSYEAVVPINFLCSRLYWSSKEPWKIVEYTVRTTIQNSSSTLTALDVGRNYTVDHTNPNSKEVQLGMAQIARW
HTSLARSEFLENGGTDWSGEFPNPNSCVPPDENTEEEPQQQADLLPPELKDAIFEDLPHELLDGISMLDIFLYDDKTDLFAISEQS
KDGTQAMTSNQAQNQNQQAGGANSVSICDEDTRNSNTSLGNGWPASNPVEDAMLSAARNSSQVQMLKTLAWPKLDGNSAMATAIKR
RKLSKNLAEGVFLTLSSQQRNKKEMATVAGVSRRQSISETSVEGVATTSGSVRSKSFTWSAAKRYFEKSEGREEAAKMRIMQMDGV
DDSITEFRIISGDGNLSTAQFSGQVKCDRCQCTYRNYDAFQRHLPSCSPTMSSNETESDVSGQGMTNNATQISAESLNELQKQLLA
NAGGLNYLQSATSFPQVQSLGSLGQFGLQQLQLQPQSLGSGFFLSQPNPATQANTDDLQIYANSLQSLAANLGGGFTLAQPTV
TAPAQPQLIAVSTNPDGTQQFIQIPQTMQATTTPTATYQTLQATNTDKKIMLPLTAAGKPLKTVATKAAQQAAVKQRQLKSGHQVK
PIQAKLQPHPQQHQQQQQTQVQQPITVMGQNLLQPQLLFQSSTQTQAPQIILPQAQPQNIISFVTGDGSQGQPLQYISIPTAGEYK
PQPQPTATPTFLTTAPGAGATYLQTDASGNLVLTTTPSNSGLQMLTAQSLQAQPQVIGTLIQPQTIQLGGGADGNQPGSNQQPLIL
GGTGGGSSGLEFATTSPQVILATQPMYYGLETIVQNTVMSSQQFVSTAMPGMLSQNASFSATTTQVFQASKIEPIVDLPAGYVVLN
NTGDASSAGTFLNAASVLQQQTQDDTTTQILQNANFQFQSVPTSSGASTSMDYTSPVMVTAKIPPVTQIKRTNAQAKAAGISGVGK
VPPQPQVVNKVLPTSIVTQQSQVQVKNSNLKQSQVKGKAASGTGTTCGAPPSIASKPLQKKTNMIRPIHKLEVKPKVMKPTPKVQN
QNHSLLQQQQQQQPQLQQQIPAVVVNQVPKVTISQQRIPAQTQQQQLQQAQMIHIPQQQQPLQQQQVQVQPSMPIITLAEAPVVQS
QFVMEPQALEQQELANRVQHFSTSSSSSSSNCSLPTNVVNPMQQQAPSTTSSSTTRPTNRVLPMQQRQEPAPLSNECPVVSSPTPP
KPVEQPIIHQMTSASVSKCYAQKSTLPSPVYEAELKVSSVLESIVPDVTMDAILEEQPVTESIYTEGLYEKNSPGESKTEQLLLQQ

QQREQLNQQLVNNGYLLDKHTFQVEPMDTDVYREEDLEEEEDEDDDFSLKMATSACNDHEMSDSEEPAVKDKISKILDNLTNDDCA
DSIATATTMEVDASAGYQQMVEDVLATTAAQSAPTEEFEGALETAAVEAAATYINEMADAHVLDLKQLQNGVELELRRRKEEQRTV
SQEQEQSKAAIVPTAAAPEPPQPIQEPKKMTGPHLLYEIQSEDGFTYKSSSITEIWEKVFEAVQVARRAHGLTPLPEGPLADMGGI
QMIGLKTNALKYLIEQLPGVEKCSKYTPKYHKRNGNVSTAANGAHGGNLGGSSASAALSVSGGDSHGLLDYGSDQDELEENAYDCA
RCEPYSNRSEYDMFSWLASRHRKQPIQVFVQPSDNELVPRRGTGSNLPMAMKYRTLKETYKDYVGVFRSHIHGRGLYCTKDIEAGE
MVIEYAGELIRSTLTDKRERYYDSRGIGCYMFKIDDNLVVDATMRGNAARFINHCCEPNCYSKVVDILGHKHIIIFALRRIVQGEE
LTYDYKFPFEDEKIPCSCGSKRCRKYLN

>LPT (*S. mediterranea*)
MSDESKDSDLFDSMLSLSSKRHKIDIVNEEISDPSNTLRCKVDDEYFVNRIIQNNNPTNENDFDDFFQSNGNNMRCSLCALGPQNK
FGQGEMLKFYSSEIPVPSYIWSKDVNSNKRSDRIKLKDCQEYCCLDTNKRLKENKIIGTDGKPCLNRIVNAFFDNEGRPFGLIDEF
AWVGLPAMVTTQKSSTGAGSNNSSLLWKAINSCDETDENLKFAHLLSYELSANDQCNDSCFYVHHSCVTWSEGASQFSDQLDGNFD
NVESTVRNGLNQMCSFCLRFGATISCRSKTNCKHKFHYPCAAAAGCFQDFVSFQLWCPSHTQEADLSNEHCICSLCERKGGVNDLI
SCTSCGQHYHSNCLEPPLILSPTVRAGWQCAECKTCLVCSESKDEKKMLVCDVCDKGFHTYCLQPPMAAIPKNGWKCDSCRVCNDC
GYSMKPGIALTGNSNQGSNVRWHNNYTLCDRCAQTKRRGHSICAVCERAWRCSLPGTKNQSSSKWPGRKCSTCHKMVHMDCDPLLA
SMPSPVYSLSSSNSGGNGTADEISTASCALVAPSNGGASHHNSTNNSNSNTNSNTNYQCPTCHNRGSATSSEVASTHTSPMPGSYC
DEEIADDLDEVDDEETASLSGFDPLPPRMLGLDSQCSTIMNRSIEQGYRSNSIDDSSTISGSTTPQTYCQSNSSSKFQLPMSRNNK
DKPYESNAAPSKKRLQPVNKGKSKTAGGKSSSLKRNSTLNSTKRSRRVKSRKINVCESKNNEEKDEHPSTVVLCRADDLFVLEQDI
CVSCGSFGIDTPLLCCVQCGQCYHSYCAEISKLTKVIIEKGWRCLDCTVCEGCGQATNEDLLLLCDDCDISYHTFCLDPPLNEVPK
GSWKCQNCVVCQFCGKRDPGPNGTWQNNYTTCAPCASKAQCPVCYNNYNDGDLMLICDVCERSSHATCDHLNTEAELEFIFEGGGY
HCILCRDKGIDIGAGHKQLLHLRATHGDSVAQTLVADKSVFSNLIGNSLEYGSKNSLLSSVGSSVDGPTSPDPSKNFFMDNVVLSE
SGLNTIRNAMLKSQPRRAQNAMKRVQKSHSQDTIGDVDSKQIKLESTEPLLNADGSVVASGNHLDEDSMQSIEGEINTTDMKSPSV
QSDDGSEIQSLPAHSAANSVNNYLPSSKTIGKAVMPRNRRFSNLGIGGFRARSTRVSQVKKSALAAVNNQNFDQNNPQYTIENFTA
KQGRRRPVKKKSEIDENFPDYLMQAFLGSSLLSHTTETSKKSGHLKLSPSRMAGGNFSVDAHSFSSTGSNLNDDGDDCCPDNSNCF
EMDNEAENFTAAQEASNYMLIEVPISYTKTEYDEPKTEPTTNVMSTCSSVSPKVTNISNHGILQSELGARFNHQMNILTNPGQSPN
KPAVGIKKSEETVHGTINSDLFSQVSSNNNSGSNNNSNNNNSCVVSLAYQNQTLTFDEDDEYLLDDYLLSFPDLEMVEEGNPDTGN
TGAEINNSTEIHSNPLELSKHISDEAGLTQLPELTVNDIETHLGPSNMENSELPMTSRRGNGARDNSGIQLSSNANIDSSYLMSSS
TNTAISASQQLPDCNNQSPMRSELVMRPMHNTAGGTGGGGQLVQMKSPTAMQSKAPPMNVQISGDPNALGIDISANFIPNSQLINP
SMRRQLFPNQIPRRHTEVSFQLQQVGNISSVSANQASFSSPIRSYDNQSLLQQKLQFSQSAPQFIMCSNNLQKPSRQINIDQGSGL
VQDINGHQIMLNENLGLVTIDNQQKQQSLESFQGQPSNQQTQQQHMILGRLHDPNVTVPVQCNQAPPPPPYPSNSQPSEGIGKPSN
DSKQKVWIHSPNFQLQMNEQSILEKQSMIPADSIQSGGPNMFKSAQQSPQSGIHQSPQKHFPVSMQQSPFNPGNQQFNKSNPQIMI
HVGQMELPEKSNKDSMLSPDTNLPVPRPKNISLSAANLENSEVNFTVDGSSGCGGTTPTPSTPTAPQSNKRINNLKWQEDEHLGDK
STIAPVLFANCVHQNIRSQIADPESRFREIQKLWRKLKSEERQEWVNKSRLNKSTKKCNVSSASGSANLSKKKKTTPAMATALQSM
QLSSPNKYVQQQSPSSIHQSPSPFQQGIINSALQSPPMPQKHNQFPQFIQQQQQMGINSDMISTSLPGVPKINQINSLQIGQMSYN
KYEFIPPPPSQQEQQQAIYQQEFNSQQQQQQPSIQMQVHPQQQQQQQAMMVSAQQQQQGLDSLRMQQQQLQQSLNASTHSVANPGT
PSMIDSNEGAISSPSKQQYCEGSNMQMSNQQQQQQQYLQPLPPPTWPPAADLSSSAAAAAAAAMQAPQSINAFHRFNKNQTGNSNQ
MVSSVLQMSAQHQMPSTLLSQSVQNPADNKLNNNNSNREKLMSTNIINQSNSEMIFSPPGQSSASLPSSPFHKEMATSDETTTKLR
DVLNKRNSNSNSNSKPNSFSQQTYANSPSISAADSNLLAASPSSSEHGFNHTSIQLRSLAGGNNNNTAAVSMAATSGMESLPDHEM
CINNNSSSFQSMNRLGKMQQSQQINQTALQMVPVAFSQSHGMEDAVNGNMQSQMMQTRQYQMPSGCYMDPQTQQKYSINPSSQQQY
LILEPSDVNAKLIPQQQQQQQQQQQQRMRQVSSGHQQNAAVAAAALMDQGCRSASHMMPINIVPGNSYLGQSKFVPGQQKYQQQQQ
DMVQNRNPYRQNTLNPQSANSGSGNSSSNQLDISGMSGMSGMPGSHSNSIGEECVQMSADPNASNIIHMDTRMSPNIIQHQSPISP
FTHMSQSQLPPHKFQQPSRPISTGLNKSFDLSGGCQMYIAAHSPRDSHTPQSSAPSNQPMQNPSGPGTPQSPCAAGANKTSGPPSP
SSKVSNMAAAIGTNNRTNQQQQQQQQQSLSPSTVPSMSSPVPLMQSSPHFQHQMTPDYESGVHAAPFMPGKSITCPLTPGNREEIG
GFEPPVIVTRVNVDRKNFGPSGVAVGSGDGNVVNNQDQYRQQIIMTSNDGINCVSGTAITADMCSSASYQSNATNQVTMSHYKPQS
TNAGNIIQQSSAGNNSRSFSRSNSVNHDDPRAMNTIQSNSFAYRPLSSANITNSNSTIVNINIQSQPSHLDDVIASVVSDSIAYID
ESPATQSIALQRQNSNDSSSCLGGGGGFDHDKSAVAAAAVASSLMPGSGDPAAIESPLIAATQYQTANNNNNNRSISRNLSSNQLI
FNSANASNPLSINQANQLLNTKQSQNSFWLATSSNINPGAKQQQQMCISNTTQLIGQFANNQPQTKIIQHGQNLYQMYAANNSPQQ
IIYTPSSIELSPNMSSSSSSTDQQQINMLTVTSLPNRQNSQTYQSVAAFNQPISQYQNAIELGQSNKNQLESEKNSLIVQSADQVM
RQSPRQQQQQQPSNSSSNATMIKLNQHFSAGMQSMQSNNKASSIQIPQISSNQFIYSPGPTLHGAQNQQGVAMVAQANNQLNAAAA
TGSSLIMGPPVSNLRSNANVVTQRFICSINSNSTANTIPGQSPQQYLISSTPQSSASVGGAVCTAGSQAQQQISSLMISSNNSNSS
NSMNTTQQQQQDLSMNFGHYN

>TRR-1 (*S. mediterranea*)
MKSKSVLRKTGFVDKRTCSLCHRTGDLPENGPGRLLCFDVDKWLHINCALWCNNVFEMQCGALNNVPESFEASLKSICSSCGKPGA
GIVCYYPRCTLKYHVPCAQEKGCLFFADRAMYCQLHQTKDLGPHLTTLNVDRRVYISRDEYSQVAGTINNEDQNFFLRIGSLVVLS
IGQLLPHQIHSGNFHNLNHVYPTGFKSHRIHWSSRVINTRSVYECEILMDRDQKMPLFKVTATDENMENECYFGSTCSETWKKILT
KISNLRKSHKVVKFFPEHIKAEDLYGLTEPHVLRAIESLPGLESLNDYAFKFGKMQLIVEMPLPINMSGSIRTEPFVKLYRHIHKP
XIHQSPNFLRHLLAAAAATATSGKIPDNILFNNKTTLSSKWQQYRKLKFETKFNVVLKRSKIQGFGLYAAKDLEKGTMIIEYVGEL
IRNEVANIREKQYESKKQGIYMFRLDPSTVIDATRAGDLARYINHCCDPNCFTELLDFEKESHIVIISNRKIEKGEELTYDYKFDL
DDGKDDKIPCLCGARNCRRWMN

>TRR-2 (*S. mediterranea*)
MTTEYFDATNSLCSNSFELIKRRRLMSDDFKNSDLDGENLMDYSGDNFQSEISAKASSSSIHESITNMSNSHSTPEDTFSEITLSS
PTSKHCKPDSDIKDYEPSIDQYLSNSNNDVVGNQNCNDIVKQLPTSSINLKSIDDTIELVLSKIREENNFSKNFDSSDSCYKIKSD
NYNSSDLSNCSSNINVQNNTLVSIARITSAKTGSHGSDNMSYSKGIISSASCLNSSIRNYGMSNYSKSYSQFYNYNYGISPQMAAN
FSNQITAPNYLEYNKMDYTRSNYLKEADKNKFFFPYQCQNPTYRQYLPNSYNNQPFYKPDSPNIQIHNQNYQMYQKPVKRNVYPDQ
IYVDDDTRLCKDNSKNPNYNIYFKTPQPNAHGLLEQGNSVRKKRTSSDSTKLTPSKKRKLANQFSITNNYEELIRKLKISNYNSLL
MNPTVSQSVIVLSPKGFIENSKAVNLKKSARYKKYKSTLFKLKKPKIFSDSTQEDRYLSHYVDESNAGSCILSQKKNLDNDELLDI
ICGSPGSNKVTEKSENFCLHNLATDNVDYDPASPELTSSCSKKKNDQSNEINFSEPKIDAINYERNHTITPIVSYTESETKLITEL
HPTKSQELNRNYIQLSGSFLKRTKINGKSFSNIHIEEAVSIASLSASKNQNTCYYCYAFLNGISFIKKNSKLFGMDSTEDLLFCSS
SCYSNFKSYHSVECNLEKKSTSNNNVQCTKVLIKITPYKPAKNYKKVCDFEQYVRRRFPSEGQGDSSVKPAVVVTDSIKKRLLGLN
YSYFKTEPLKTIYKDIEKNDQNLLLKKYSCSYRKPDLVDVRSCVLCRNVGDGDEDICGRLINIDADNWVHLNCALWCFEVFETVSG
SLMEVDSSLRKAIKTICTCCNETGAGIQCYRSKCTNVYHVPCAYKVGCIFYTDRGMYCPNHVVLNHPMHLPSLRVERRVNILRDES
ALIAGMVTEDQDVKIKFRTGSLTLNNIGQLLPEQLISGNYNSCSIIYPVGYHSTRIYWSSTQINKRCKYECLIDESDDMPLFKVKF
IDNGCENVLTDKSCDKLWKTILNKIAKLRQDSGLTRLFPVYLSGEQLYGLTEPQIIRVIESLPGVEMIKNYAFKFGKMQLITEMPL
AINPTGSARTEPKLRTYVRRMRDYQETKFNKCGILKAKPNNNYQPIIYDNGNLNKTALLNSKTQQYKRLKYEWRNNVILSRSKIQG
LGLFAAKDFEKSTMIIEYLGEIIRNEVGNSREKLYQARNNSIYMFRLNEDLIIDATVNGGLARYINHSCNPNCVAETLSLDCGSRI
IITTRQLIERGEELTYDYKFDFEDNKADKIPCLCGSINCRKWMN

>LPT (*Crassostrea gigas*)
MSSWQKGGFNLTEKGYYTLIVDYGNTGVANMKSSLEEFRFNTVQIQKVGSGCLLKSFVKV
KLNVACSVYIMTSYFRPKAREPEEVLVFPKEIEENNSISLICAAYVGSPRGYIHIWRVFQ
NSNKSKLIYKSNFINNETENCTKYINVTTTYTVTREDNGAVFRCSSQNNLTQGPGLSKYS
QNISVLWNMASFFNAAILLGLLGLSSALLEPIPFPTELQECYEYRSDNVTASAESAIYIQ
NTCYRSFLTDQMTDGKVWSGENLTQEGINYIDSLFRRIMAEADEVEKYKKLGGRQKRQTS
TRRFRQEVRSPGAFQPYANCVQRLQNETVEPASAGRNTYQTLAAFHSGRTLGRAHNGPAF
LPWHRIYLLLLETECDAAIPYWDSGLDHDMVDPTTSILWSDQYFGNGDGEVMSGPFQDMR
TLLGTPIIRNYGTGDSSLFTKEGLRAVLSRRRFQDISEPLPRGSIYSLEGHHNGPHVWVG
GHISALNSAPWDPVFYMHHAYVDAVWARFRELQIQNGFNPETDYPRRPRQGHRRNDVINF
GPYFELVTNLEAMANRFADLVTYEPFPVCENNCNNSPHLYCDQLRLVCISRIRTAVQTSV
AGIVAMGASRGVSSNIQSQSLARAVARGPLPVGRKFSDSPFIDIRNRPDNIGTARAAPQI
RRASSQVRTRARRVQRAVSQNAAFQENHLQSVSSLERSFTNTFVIDGVVDVKRWVYIPVR
IVYSRSKNLKGIDPTLFGTVNQSQEKCQTAHSGASKVFVSSNGLNYYGMYTEYAIIDNRQ
PVYSTAMVVGVKNPEYGEGETLFTAYDSCGRPCRPVCQTSVNGQRNYKGCNSALFTKAGV
RAVLSRRRFQDISEPLPRGNIYSLERHHNGPHVWVGGHLSALNSAVWDPVFFMHHAYIDA
VWARFRQLQIQSGINPETNYPSTSQRGHRPNDVISFGTLFERVTNIQAMANRFANLVTYE
AFPVCENNCNNSPDLYCDQNRRVCISRARANSPATASGFVAMGAARGVSLDMQSQALDRA
MARGPLPVGEKFSDSPFHDIRNRPDNLGTARVATQIRRAHSRVRRHSRRRRMQRRRVRRS
ATLNATLQHNHLQSVSALERSFTNTFIIDGVVDINRWVYIPVHVIYSRSQNVQGIDPTLF
GTVDQIEDTCQTAHSGASKVFVTSNGLNYYGMYTEYAIIDNRQPVYTTTMTVGVKNPEYG
TGETLFTAYDSCGRPCRPLCLTFVNGQRNNKGCSGAFRVSTSSPQMYSITYSDALNSTLM
TYSQIDTHADFSAPALTFLCDNKPETVKLENTVEEAGTQLSVDTIESSVAADINSDNSQS
STSVPPDITQADTYSLDDPYTSDPPDLNPEEDTNLPTKVIKKEGEYLTLPTYEAAELSDS
SGPPDITSAAYEKTEEVEDIEIKDYSYVDSPNSPAPLLTSLEEQGEFTDFPTELPEDPQN
IFPDTDPPPDLLHQEDLTPTEGYLGSSFPESSLSESLLAVAPGDILPVVSLPQINPNFFQ
QNLSPTLAQNIKRGPGRPRKDGQEPAPRRRSSSKPLAKSIVTKAQRILHNSAGISRKFFP
SAEQMDSGSQSSLSQSELTMTIDTGEDTTQYEMEEEMLIGVSESLDETSLESQLPVRVCS
FCNCGERSLLGQGDIFRFEPTQGFNFRKTLAKNDKKPADFDERNQEKEGTTKPLTWRRNR
GPIKGVGERERERKSHSPRRTGNEEDQNILLGDELTFLGFPEDVEAMQVFEPTGHVWAHNR
CAAWSEGVTAGVDGSLLSVDQAVFNGLSQKCSYCRRYGATITCIYPECSKKYHYPCAAAG
ACFQDKKKLAILCPDHSDQAETIAGEEAFCVLCCQADKIGKQLFCTSCGHHYHGGCLHPS
VALSPEVRAGWQCPDCKVCQMCRQPGEDSKMLVCDTCDKGYHTFCLKPVMTAIPKNGWKC
KNCRVCGDCGSRTPGSGPSSRWHLNYSVCDSCYQQRNKGLSCPLCGKAYRQFTQKAMIQC
GTCKKHVHAECDDAIDNLMLDRVRNEEQVDYMCSVCRNRDPEFSSPPHIHPHGEGCGIGS
KGGKLNAMSRKKIGVSSSSSSRSRGRPVAPEKKKKPPPSYSEGKRGSKTKMKSNQPGAQA
QISPAQPGEPLKKSQFPTEDDDDGDDHPQTIILSNAQDKFVLDQDVCKSCGSFGRGEEGK
LIVCTQCGQCYHPYCASVKVTKVILSKGWRCLDCTVCEGCGKPHDEGRLLLCDECDISYH
IYCLDPPLDQVPKGTWKCKWCVMCINCGTTTPGFGCNWQNNYTQCGPCRSKIDCPVCRHK
YQDDEMIIQCLQCNRWLHALCDGLRSEDDMERAADYDYQCLFCRPKTGKDGPLPPPPPPP
TPPPIEMEEPPTPPFYREPEPIPQKRYLMDGVYLSETGVQHMQEITIQIPKVKRQRRNNK
RLSVDMLPGQRLATQMSTEGDEEKDGMDEGSELSTPVTAEARGEGDPLLLSPTSQAPSGI
PAEGEKKERKKRTTIGLGVGGFIAKARSRQTNVKRQMSEVSAEMADGQPRPEGEDGEPVP
LLVEGEVKKQRKRQARKKSQLENSFPNYLQEAFFGKDILDKSKAKVKQGHRADSDSDTES
RASTPNLPKDIPQPSLFPELSQSNTKLPSTLPSVGTSQSVLSSGASMVPSFPGTGDDINP
IGDVLPDLELPHDDIFSIFKEGGKFLLSILIAKLPMSEEQPTSSQAGASGSGELPDILSI
ENIDVDLLPEHAEDLPPINGQEVDDIFNGVLPPEEDPSQAAHGGGPQFPMPGSVPRLPPH
MPPHPGQMPPGMPQLPYGPDFRGMGGDQPPWQQVATEEEGGGSTSSRRNMLKWECDEDLG
ESATISAVLYCNLRHPEFKQQYPDWSERVKKIAKAWRELSSDEKQPFLTQDGKKPKTKVP
SMPMGPPLPPGPQGQMPYPPPAAADNLGSPVVLSPSNRVPSNPSMVTTPEGHSPGQMPGS
LPESPLSHPGTPQDPRMMVGHGDPYNPEGETDDFAVMVHPGMRHPQGGPQRPPMARNPSG
GMDPYSQDPYAKPPSTPQTPQKSPSKIQWPGQAQESDPYENPPATPAEGMSSGHPGHPYD
PYAHPPHTPRPGMVPPQGIHRPPFTRQSSVPATQTSHSDDPYAFPPHTPGPRPGGDGQNP
DIVGRQQGADMYPGMGDHLRSPYPPSSVPGSMGQPLPSTAPDIYRMAGQSGMRHPFRPPG
PPTSQSMARPDLYAQPLHPGMEPPREQSAGLQDKNSSGETSQQVRNILHKQAESRMRQGK
GSEPPQPPWENHPQAFTSEEMTRFPVSRGAWPPAPPRVPGPRQIPHPTDMENFGMRPPYS
EGVHPGQMRQPRPHGPMMPEMYSPKHSPQMSPGQLPSLRPGQNPLAGHDPRSQFPNMMEH
RFPVPGQQVPPNQQQVPNISGQPQTYKSYPPETGEQSRINPGAMGEQRKMSDHDQREEEK
SVETKEEKKELVEDKQMNEADIEELLSSDGTFDIIKFVDTDTELNLDENKSIFDDLDDVE
SGSMDKGDSKEDMKMTKEKMDSDLDSSKSGGGIPDFQSKFLEFSQKKNEERKIGAEGSGQ
MSEMDKKDQQSISQIAALLQQQSSEHMPPTLERRDSVKSDSSLPGKEQLPRTPGSGPLTP
SGKQPGVDQGFPGMGSRGPYSAGMHSPLHQMQSIPTQPSPGSTYPPGQGPPTPGLPSPKI
MPSPRSNIPSPRTPSVQSPFNPMNSQQSPFPQTQSPFSPTVSSAPQSPYAANKPQPSFSL
PVGSTNQPGYSTVESGVPVQSPGQRSPRGTITPTNQYMQGTYGQPMMQGPPRATTLTPTP
TSILYGGQSGMAQHGMRNPVPASMNGSRPPFSGASAHPMGHSMETATSQMHSMDPSAVPH
MATSQSSHTPGDMQGHPPSSQSTATMPPYGMPGQRMPGIRPPMGIPTSSVPSSAPGALGG
GRPQLLQDQPLLIQDLLEQEKQEQQRQAQQQAMGMMQRPPPEAMMGQPPRPGMPMGQYRP
RMEGMRHPVDPNWNVQRPYGQPEIPPGQPQFVRMPGQQMPPRMPGPYVGQPPGVVGPMGI
PTPPPPPPQPPMTGELTPELERQQQQYEDWLMKHGNYLEMQVKTLEQQIGKCKRTKKAIN
ARNRQAKKTGREPSANDATELERVTQEQAGLQKQLESYRKQVKQHQMQTQDYRTKKRERY
GQDWAFSQMPPVTGPAIMNVPPGGLQPRIPGQQGSARLTASARQAYDEYMQDRLRQNQQI
PGAVGGPRPKHTVVEDNNPFSEEYQEREQRERVGCLPKPGPEPEKPEMMRQMPPYFDPRT
IPYDQTGRFPAPRMPGPAEPMQRFPGPAATTVGTAEPRPPVQFTPQSETEKRIMEILNNS
AGLAARSQHEAVDKSPGKAKAGDTQGSPATEQKPEGRKPEEGEGAQRVTSVSVPGPSTQP
TYTTAVSQPAYQSESNDPDDQQEEHILPMNVMIHQSIPQSAAIACSVPPVQVSQSQQMEA
ARIQESESMPQTAAKSVEGSVQSDQGHTQERPPISGSSEAQPVSSVNSNVNTVQEPETVS
TNVQHLPPHSYPMTSLPGHPLPPPNTQERMSPRGMVPPAYMGAYGHFQGRMSPRQAHYGG
VPPTLMPVSQNQSPRQSTTPSPGRYSPRQASPSYPGGPSPGRYSPRQGSPGVSRPPSRPP
SQPGSGYAPPRQGPTPPITGYPSSTPPASPNVTYSPVTTFSAVSVDISSRVETSVAHTTI

NSRPNVSTSTVQSVSQNTPTDSVAPPPATESASAPVVEQSNLKLQDTNTTGDHVKFNLPM
AFSAHVLAYGLNRWKDGYSSSNQLQNMYEMLRTPLDYFMKCWRPQSQEYYAQVGNGAADH
AFWGRPEDMHMQRPAYKCTASNGGCSDVEGITVAALAAGSMAFKASDAAYSQRLLASAKS
LYDFANAHKGIYNKGPISDATSYYGSTGYKDELCVAAMELYKATKDAKYLNDAKANFEGN
DVAWALSWDDNHVMCELLLYEETKDNRYKGLVESFVRSYMPGGSVHQTPCGLAWRDQWGS LRYAGIFQRE

>TRR (*Crassostrea gigas*)
MHPPGPGDMHQQRAPSALVRPHNEIPSSQDTKQPRIPQKGGTKRQQRPHQLQQEIPQGMR
PRMMPDQFNRGVRPMSMIDPRLMSTYFSPENIRMLQMRYPDPRMMPPQLLQQMQVFARMH
QAQQQAQQQAVRASHPPSSQPTGVSVAPGGQQQPQGLPSQSQASMIPSQPMSVEGSQPAN
SSAQKQPHHPIPPHPGQPYPHVPLSHHPHEPGKQGLGPHAQEFVTSRSNPMDMNSNSTNG
TELQSQSGVQTALASTSAPHLPNDVMNRPLSSGRPPSAGRPPSAGRPPSVGRPPSANRPP
SVGRPTSTEPPLNRPPSVSRPPSSNSRPPSADPSSRPSSVSRSPSTPQYEAVPPLFGYPV
SSSESTYSGTDKSSVPTTSTLSSQMESAASMQSFSISSSQPSPATPTVTQTPYTGEASHE
GAPEGEQQKTAQKRSAAPPHSSTTSAAPNMPSPFGPPTSMAPIPQGYVQTPFGMLPHHSP
GITQGVPFSGMDGHRMMSPHMRMPGFPTQGFPMPPHMQRLPPPHGGPLSPNSMQAMASLA
QRMQGPPVSGQNVRPGMPPNLPGMGQRGMTPGMPPVSMDGAQAYNYNYRMPFGIGPHQRP
PGGPTPEHPPRGPRAKGPRPPRPMAPLHMPPNMPPNSEHGLISPGQSMPGMMPGQHGNLP
TDPNMQSVHSQGRPPSIGTGPQPSPLDSSSPSPKVPATPPANVDQAHLDEMIQNTQVKHE
PIDPSIKTEMTQEQHNALLKQLLGKNLKVNPEDTEDSVPLISKLTPEQQKQLELIENMPL
VMEHEAKTPEEKERLLALTELKKQAYEQRRKEYEQLRKQKRKQQNAENKKRKRKEKEAEA
ALMVDPSQPAPPAKKRPRKKKPDSLEEDLDAKANHFIQMLGTLPTVPLEEPRVGNFFSVL
NVPGSPNVITGENCLKGNFGNSYLEGVADFYGNTLLHGLHPLGSFVRMPKPRDIESRRQF
IRGGQEESGPDLHTDQQRYHLNSDRLPIPGPPFGLPGLPPFFGRGEKPPVPSLPPSPRIE
PRLDYLKGSDSPCTVISSSSPEHEFGDLDEEFPMLKPIEPASAADDRSSPLVPILQPVAV
KSVPLKLKEEPAIQIKSEPKDKVKAVTDTDLTQKKLTMSDKHENGKLPPNLSALTRPFQN
ANLNQDISVTLTLSANAADDIGGVLSQIAELLKIAVPPSYEVGSRSPSPEQGKINSKHLE
DKVNVQNLIKSKAKFCRLCDSFVYDSGFRKRKCDIPFLAKEDYFSIKDFDDEEVTFCTEK
CYTKFATVYQEINKPPPEDPMRNLTLSKHHSSDFSPRSLGSHTPTTPTKSDSTLTLTPTT
PNSGGALTPVQPSPPIKSRLEDKISKKHRRTASIDAPVPKPLIKKHKNVRYRRWDQAFTD
SLPQEPGVPDKEIQNLWKALGTIIVPKTLPEDTRKCLFCQQQGDGETDGPGRLLNMDIDS
YAHLNCALWSSEVYETLNGALMNVDSAYKRGLKIECTACGIKGATVGCFNNRCPKYYHLG
CAKKVGCMFFQDKTILCPSHAAKQGAENVLESLSVFRRVYINRSEDKQVANMVHGQEEGH
YILRIGSLTLHSVGQLLPHQLQTGKFHTRDCIYPVGYKSSRYYWSMKRLNKRCRYVCSIV
DNEGQPGFKIRIVEPGEEDVVLLDVSPKAVWNKVLEPLDKLRRNADLVKIFPSFTSGEEL
FGLTEQAIVRVLESLPGTDLLTNYNFKFPRSLRIEMPLTINPTGCARTEPKLRTHFRKPH
TLQSANTSRSLPQTVTGVSGDINSPYMKQFVHSKSAQYRKLKTEWRTNVFLGRSRIQGLG
LFAARDLEKHTMVIEYIGYLIRNEVANRTEVVYEEQNRGVYMFRIDNETVVDATMAGGPA
RYINHSCNPNCVAEVVPFDKESKIIIITNRRIPRGEELTYDYKFDFEDEQHKIPCCCGAP GCRKWMN

>LPT (*D. lacteum*)
MFHDFNNVDDGNLGMNVKRLKIKIAGEVMETEGVEDAEVGVQEFRRRGSLRGEACALCGLRGECGLGQGRLKAWTDESGEIQRLVR
LKSRAEESGGGQETLRGWSEESGALRETIKKEETVLEVIREGGMCPSVIKEEAVAHGPVKQISASEDTDNSCEKSVEDHLCLDTAR
ISREKGIPGTDGRKCLSEIVNAFLDSENRPFGLVSALKDVGVFDDDEFYVKSEEGRTVVYAHETCVSWSGGCDNLLLKRNDVKNHL
CNFCLRPGASIKCTASNCSHSFHYPCASASGCFQDHDNLRLWCPLHCLQAEISNERCICTLCERKGEICDLIYCTSCGQHYHSNCL
EPPLILSPSVRAGWQCAECKTCLMCSESKDEKKMLVCDVCDKGFHTYCLQPPMAAIPKNGWKCDSCRVCNDCGFTTTIVSSQSVNV
RWHNNYTLCDRCAQTRRRGHSICSVCDRAWRCSITIKNTNGQSCNKWPGRKCRKCRKMVHIECDRVPGIMPSPVYSLGSSSSGGGE
ENSINNTATGGSHYQCPGCRNRESATPSEIASTRTSPMPGSYCGEEIADDFEEVDDEENTSLSGIDPLPPRMLAFDSHCSSNSLNR
GVSVDAGGYQSNSVDDSSTISGSTTPQTYCQSSGNGTFQLPLSTNRKDKSVSEQYVNSSSNKKSQVTTKSKSAGGKFLAASSSSTI
KRNSSLNNKKCRRVKVRKTNVCEARISEEKDDHPSTVVICRADDTFVLDQDMCVSCGSFGIDAPLLCCVQCGQCYHTYCAEVTKLT
KVIIEKGWRCLDCTVCEGCGQATDESLLLLCDDCDISYHTFCLDPPLKEVPKGSWKCQNCVLCHFCGRKDPGPHGHWQHNYTTCAP
CASKVQCPVCQVNYNEGELMLICDVCERSSHSVCDQLITEAELEFVFESGGYHCVLCREKGVDLGPGHKQLLHLRATHGDAMTQTI
VADKTAFSSLLADQGRNSLLTPVGGSYDEPASPDASKNFFMDNVVLSENGLNTIRNVMLKMQPRRNLNALKRAQKSASLDASADQD
NKQLKMEPMEALQGIEGGNVNNHMDEDSLQSLENDVTNTDLKSPSQQSDDGSEMTTMQAVSSVIGSGGVNPQQINSKNAGANKNTM
PRNRRFSNLGIGGFRARSTRVSQMKKSAVSNVDQNSAQYTIETFGSKQGRRRGQSKKKSEIEENYPDYLMQAFLGSNLFNISTETL
KKTIMTPSGHPTPSTGLLDNRSFGSASSVHNLNEDDMDDLDNDDDSENLNAAQEASNFMLLEAPDPSIIKCESDNTKIEPTLSSLA
SQHLNSSNNMLAVNVNQRQMPLDINQRFMPQKATSISNHSPSLQAQSLVQASNYAKKLDVDDDPMRSINTNNDRRLHHSNQAASLPT
PYQNAGPMTFEDDFITSEDYLLDEYWLSINDLGICEEPNADITNTTSENINPVTINSNSMQLPKHLGDDSGLTQLPELTVTDIETH
LGPSSLENNESIASKRNRETVISSHPDLESSFLLSSASNQGMVVHQNQNDMNTKIDMAAMLMRQQHSSTQRMNSNLPTMIPSVSVQ
ISNDHNSLVDIGGGNLVPNSQLMHPAIRNQVYQTSNVNQSNRRPGQMAFQLQQLGGMPQSLSTSQHNFHSIRSYDSENIPQKIQFQ
QPVTQQYIMCSKKPANRPISLVDQTVDHIVLPQDSIITAIDSPSQKQLDSSHPSQHILISRLTAESGGQQAPPPPPYPCSSGNTTS
GQSNDPGTNSTIAKALWMQSSNFQSPQSSSDQEKIINVSENNQSSFMTPSQHSQQYLQTAQRNIIVSAQSPSHQQQFNQGNKQFSK
IRSPSILEAVPLSDNAEKINKDASTPLDAQSQVSRPKNISFSSQDNSEASVVNPEAITPTANSQSNKRINLIKWQEDEQLGDKSTI
APVLFANTQHPNLRNQYQDIQIRSREIQKLWRKLNSDDRQLWVNKSRMNKSNKKSGGSTSSSNSQPKKKKAAATPTTSGNVQQMGS
PVKFIQQQNIELQSPGSLHQSPSSSFQQHSVGQPILQSPPITQKHVQYSSFMPQQQMNINSDIINMNPQNPSKMCQMTPQQQMQMN
QMQMRYKYEFIVPQSQEQQLLYQPDFNQQQGSMHVQVHQQGLASNSPNIDSLRVQQVSNQQKSNINSQTSSMLDNESTKQQYVDSG
IVSSNQPTYHPPAPPPTWPPVSTPPMMPAPSSMAAYHRFKNPAMGNQYNTTQSSIQQQQSPHNPDKVNKEKLVIAINSTDGVVYSP
SNTPSSVNKDLHSQMTSLDPESNAKLKEILSKCNNQKQNAFNQQIYLNNQPSDNLLACPSPDNNAHNFGQNVLIRSNSIPIATNNE
SCQEHDSNLNNSFPTMTRTSKSQQIMPIAFAQSQLDDSNMQAQQMTVATPRQYQTITGCYLDNSQQKYPLNQQQPQQYLILESTDN
TQSSSKINQTHQIHGHNIGQQQQQQQQQLQS

>TRR-2 (*D.lacteum*)
KRLLGVNYRYFDNNTEKVSQLYNDICKENQTYLLQRYSCAFKSNVLLNDKRVCILCLNAGDGEEDICGRLVNVGCGVWVHLNCGLW
CFEVFETVSGSLMGVQLSIEKALKTQCSYCNNTGAGIPCYRSKCTSVYHVPCAYKVGCIFYTDRGMYCPPHAVLNHPMHLSSLKVD
RRVNVYRDESALVAEVILDDYQEQKCKLRIGTLILNNIGQLLPEQLLSGHYNSRSIIYPIGYFTTRLYWSSFNVNKRCKYECHIEE
ENDRPLFKVVYKDDDNNKDVVVSNQCCDTVWKSILNKIVKLRQDSGQTKLFPLYLSGEQLYGLTEPHIIRIIESLPGVEMIQSYAF
KFGKMQLITEMPLAINPTGSARTEPKLRTYIRRTQQEFSQCGILKAKCTSQQQLGLETSKISVSNTNSRTQQYKKLKSEWMNSVIL
SRSKIQGLGLFAAKDFDKSVMIIEYIGECIRNEVGNRREQLYQSRNNSIYMFRLNDDMIIDATVNGGLARYINHSCNPNCVAEFLP
LDCGGRIIISTKQRIERGEELTYDYKFDFEENKSNLIPCLCGAINCRKWMN

>TRR-1 (*D. lacteum*)
MDSNQSEQQYYLSKLTSPAQYPCTVIGSNKNTIHTIVPNISNSNSLNINVKYQNSNESKLAQSISYNCYPNNANSSYNQSDDLKFL
FSGAKPKKIPHTKRVKGVINQCIDNSNDITAAPNQLTILVQKVLPAQVPDHHPPLSAASHISAAQSSLHSVISPPRESVVSSSDII
PPVTVDTITSVINDVISKATSNDRHDDVFFSVIKSRWDTSSPDNHHVTFIPPHVSTPQFIDVNLDSNHKRPSNISLESPSLINKRS
PNKRQRLSEPIYNDNKSYQRCQLSAPNFRKLVHEPHISEKCQVLPPLGVISRRRPCISLDQLKYNCIKYRRVPFILCQPHPRTAYQ
DRYIVENLNTNIPSSRLLTSHDIIDDLDCIVAPSPNNNDLDYLSVFTSDQVSSPSVNLHGPNLEMLHPPQRTLSPQLSLFHCPNKD
YLKLDHPTSSNMTNKYVSTPTSDPLSDTSIVDSETGEMVISTIKPHPKFQLVTSLNAHHSNTDDYSIADEEAIHSKSTDFFEMPLY
SEKVYKSTSLLWKRKHRENIIGEDCTSINGDEIVSQSKDQHPDSIFIGEEPSSVVQVQGSLSSKSKDQRIFIGEDSPDSSFVIQVQ
GSISSKSKQISKRLHIILNDKLTKKKMVSVKYSPKQYKNVKYYAFHSQTLVSAHNSVISESQFIQLSRQSNTVHKLRHNIPDTRQC
SFCNQQGDVAESDAGKLLNFDIDKWLHLNCALWCNDVFETQSGALTNVPKSLSAICTACGKSGAGICCYHSRCTAKYHVPCARENN
CHFQDDRSMYCNQHQSHSNDVGLRSLSVNRRVYISRNECSQIVSALNHDRRLYIRIGSLTVISIGQLLPHQIHSGHFHNLNHVYPV
GFKSKRIHWSSRLVNMRSIYISEIVLDNDRPLFKVTATDDNLDDIVYRSPNCNTVWKEILDNIASIRTKNDLIHFFPEHVKGEYLF
GLTEPNILRLIESLPGFENVCDYAFRYGKVQLMSNMPLPLNASGSIRCEPFHRVSKHSHKLAIHQSPSFTRYSLASSQKQEKNNML
TMLNTNKLQLYRKLKQDWKSNVLLKRSKIQGLGLYAAKDLEKGTMIIEYIGELIRNEVANRREKEYESKKQGVYMFRLDDSTVIDA
TRAGDLARYINHCCDPNCFTEVVECDKESHIVVISKRKIEKNEELNYDYKFEDDGKEEKIPCLCGARSCRRWMN

>TRR-1 (*P. torva*)
MDSNQSEQPYFISKLTSNIITPAQYPCNVIGSNKNNIHTIVPNISNSSLNINVKYQNSNDSKIAQNIAYNCYSNIANSSFNIANQS
EDLKYLFSGTKPKKLHSKRVKSVNQCVNKDNDNSSDITAAPNQLTILMQEVLPAQVPDHHPLSSAHQSHISAQSLPSVISSPHNVA
VLSPVAVIPTPIVDTITNVINDVLVKAMSNDPHDDVFFSVLKSRCDTLSPENNRHIPTTFIQPHTTSPQFMAFVENNTIEENAPTN
KRRNNNLTKSVSKRPKLSEPIFNDHKPYQRCQLTAPNFRKLVHEPRISEKSAVLPPLGVTTRGLPCIDLGQLEDNCVKYKRTPFAL
GRPYPRNICQDRFIVQNLNANIPAGVLLSSRDIPDDLDCIVAPSPNNNDLDYLSVFASDQVSSPSVNLPGPNLEMLHPPLRTQSPQ
LSLFHCPNKDYLTRDSSSVVRTRKCASTPTSDTSVVDSETGEMVSKTKNRNDKLLVTFAMNQPPDDVETRYKIEDMIIDKDILQSK
SADFFISMDKGELPLYSEKVYKSTSLLWMKKCREKLSADDSGLNDDDGESVADIASDCLQNNKDCVDMIADSSQLKDRRIFIGEVP
VDTAIVVQGSVSKSKQICKRLHIVLNDKITKKKSALKIAARQFKGTKYQSYRCPPPSLAPASTIAVVSATGAFVDQSVISESQFVT
QSRVAGTVRYLTTVTDTRQCSICLRVGDHLEDGAGKLLCQGLDRWVHVNCALWSPGVHESQFGALHGVPRSAAAMAGFCAVCEQSD
ATVPCSEPHCLYHYHIDCARDHDCVLGPDHSLLCPVHSPVTNTAVTVTLRVDRRLYVARDEAAQIVAAINDEADDDDDNDDSEAGP
CLRIGALTVLSIGQLLPHQIHSGHFHTMNHVYPVGYRSQRIHWSCRVPNKRSLYLCEILASTVASGDEWPSFRVTLLDDDKNDDRY
DDCPRIFEGRDCTRTWRDILERTSELRRVNGLVQTFPDDIRGEDLFGLTEPSIVRAVESLPGLESVHDYAFKYGKMELIGAMPLPV
CVSGCARTEAVYVPRLLSRHVYRPSGWASTTTTSLAIQASTNGATTDQATTKTGLNAANSGKWQQFRRLKHEWKHNVVLKRSKIQG
LGLYAARDLEKGTMIIEYVGELIRNEVANRREKLYESRQQGVYMFRLDESTVIDATRAGDMARYINHCCEPNCVTDVVADFDRRDR
HIVIISNRRIERNEELNYDYKFDLDDPGDEKIPCLCGARNCRRWMN

>TRR-2A (*P. torva*)
MLGGDWKETSAKASSSSIHESITNMSNSHSTPEDNYIENALSSPFAKHNRLDCDRDFEVSTDQYSKNPNGILNSILVSSKNHTSID
DTINYVLSRIQEENDFGCFKSSLLLKPSSDLSVSSSSKCDKSKFSNLKTMPNCIDELDHLKRSFSSISYHNSDSQSMQIDSSVLNH
NYSTSNFNKNPQQIYNYGNYYSIPNNFNNQMNCNRQPFNKDFDKGNNVYNQYQSMTYRPYMSNSYNNQQYYRQDSQIIPMQPNYQT
LQKPIKRESYNHVMQENDLIKKAGFGINFKAPQTNAYGPFDQTVSKRRPANDQNKLSPAKKRKLINQTLLNNNYEDLVNKLKATNN
HTLLMNPSFTDGKSLLPPLEFIESKTKSIQYKKYKSSPYFIIKPKRFKDHNKCDEFIANFVSDSISGKVLLSDKCILSDGENVPDI
VCPSPDYSQSYNILCHLEIDDHLYDAQSPQLSESQTTSSTSCDHLIPKFPVKINDYPSRPVSQLSGSFLKRTKINGKSFSNVHIEE
TVSVASLAASRSQNCLLPLPCICNSCSADNNSDKVQIKTTPFKVSKNYKKVTDFEQFVQLRIPSDGDDVKKPLLGVHYRYFSDSDK
LTSIDNDESTTSDLKEQLARYKCVYRSDWLVDRRQCVLCQGYGDDDDDLCGRLMNVDCDTWVHLNCGLWCFDVFETVSGSLVGVQS
AMTKASKTRCSLCQETGAGIQCYRTKCNAVYHVPCAHRAGCIFYTDRGMYCPSHAVLNHPMQLFSLRADRRVIVGRDECAVVTEVV
LAETRSPIRLGTLTVHEVGQLRPAQVASGRYHTSSVIYPVGYHSSRIYWSTMVANERCRYDCHIDDSDDGDNDGPVFRVVFGDRTI
TDRCCDRLWHGLLTKVAQLRHEHGLTRLFPAMLSGEQLYGLTEPHIVRVIESLPAVESLVDYAFKFGRLELINGAMPAAINRSGSA
RTEPKIDTNTRHRRPSTAAAITTAGTVSGGILKTKWCSSDSPMSLATILSVKTCNSRLQQYKRLRYDWFNNVVLSRSTIQGLGLFA
AKDLEKSTMIIEYMGEIIRNEVGNRREQLYQSRNNSIYMFRLSDDLIIDATVRGGLARYINHSCEPNCQAEFLSLDSLLGGRIVIS
TKQRIESGDELTYDYKFDYEDESGDKIPCLCASENCRGWMN

>TRR-2B (*P. torva*)
MNPSLPQSKLLLPPLEFVESKTEIDIVQKQLTYEKYKSSPYFITKPKRFKDHSQSDGYLLNYVVEKMYGKVLLSDKCNADCDNLSD
IVCVTPDDDWRQSGELLSQLQCETDLPVRPDSPQLLSSHTPLTQTNVRSIGCDSQQTCDFNVPSNNYLTLVVAYSDNEARLITQQQ
TLVERDLRSYSQLSGSFLKRTKINGKSFSTVHIEEAVSVASLAATKQQNSCRQCSMYLTGLSCLRRHGGNGGTDELLFCSSSCYSQ
YLATSTGEARSLSPATGQVLIKTAPYKPPKNYKKVSDFEQYVRRRFPSEGHNELLPKMSPAVAVVLSDAQCRPLLGIKYSYWREVD
VSLSGSLYRDLDRDDQTVLLAKYSCALRTVAKDERVCVLCRGVGDGDERVCGRLMNVDGPEAWIHLNCGLWCFDVFETVSGSLMGV
QAAVERASRTRCSLCDENGAGLQCYWTRCSAVYHVPCAHRGNCVFYTDRGMYCPVHAVLNHPMLLTTLAVDRRVAVSRDECALIAE
TMILEPGENGARQMIRNGTLVLASVGQLLPEQLASGLYHTSSVIYPDGYRATRIYWSTIRPRQRCQYHCQIDSDGADGGPLFRVTY
RDEDDSDDTVLEGRSCDQIWQTILAAIARLRIDHGLTRLFPVFLTGEHMFGLTEPHVVRVIESMPGLETIANYAYKFGRMQLITEM
PLAVNPTGCARTEPKLRTVPRRSQDDSRVQCCVLNARSLSHRSSSMYDRTVSLLTSSTQQYRRLRLEWTDNVFLSRSKIQGLGLFA
AKDLEKGAMVIEYIGEIIRDHIGARRERLYQARNTSAYMFRLSDDLILDATVRGGLARYMNHSCAPNCQAEYLSLESLGAHIIIST
KQRVERGEELTYDYRFDLDDSGSDKIPCLCGSLNCRRWMN

>LPT (*P. torva*)
MFEDMNDVNDTNILINVKRIKIKMAGESIDAEIVRNIESPIDDNKNISCVKCGTDCICDNVKCILCGCGVQHSLGQGQINRWQWPE
DENINIGNLRKFGKRDEDQLCLDTARVMRENKVPGTDGKRCLTGIVNAFLDSEGRPFGLVDELNMVGISCDDDSFVKVEDGVSVVY
AHEGCVTWCKGSNLSSSLCLATITAKTRTQHCSYCLRLGASICCSAAGCLNYFHYPCASAAGCFQDYENVRLWCPDHLQQADISNE
RCICSLCERKGEISDLIYCTSCGQHYHNNCLEPPLILSPTVRAGWQCAECKTCLVCSESKDEKKMLVCDVCDKGFHTYCLQPPMAA
IPKNGWKCDSCRVCNDCGFSQKGSVTLISGAQSVTNVRWHNNYTLCDKCAQTRRRGHSICAVCDRAWRCSISLKSASGQVSCKWPG
RKCTKCRKMVHVECDNVSGIMPSPVYSLGSSSSSGGGGGPGGNGGLVGEDNVPGNNALSGNNAHYQCPQCRNRESATPSEVASTRTS
PMPGDDLADDLEEVDDEENTSMSGIDPLPPRMLAFDSQCSSVAAALSRAISTEPGGYQSNSIDDSSTISGSTTPQTYCQTTTTTTT
NGSGSTTMTVASTMAGSMQGESGGSAGGSGKFQLPQGTGRNSKDKNQYDQQGASSKKRPPVTTAAAASNNGKSKSAGGKSSSSMKK
NSTLTSKRSRRAKARKINVCESKNSEEKDEHPTTVVICRADDEFVLNQDMCVSCGSFGLDAPLLCCVQCGQCYHSYCAEVTKLTKV
IIEKGWRCLDCTVCEGCGQATDENLLLLCDDCDISYHTFCLDPPLKEVPKGSWKCQHCVLCHFCGRKDPGHNGHWQNNYTTCAPCA
SKVQCPVCQTNYSEGDLMLICDVCERSAHAACDHLNSEAELEFVFEGGGYHCLLCRERGVDVGPGHKQLLHLRATHGDSMAQTLVA
DKTSLNSLLLSSSSSSGVAMAADHGRNVLLSAVGSSFEEPSSPDASKNFFMDNVVLSEAGLNTIRGAMLKIQPRRNLNALKRAQKS
ASLEVLAEHDCKQPKIEPEIGGTTVEGGAVLGVTNHMDEDSLQSVEGDMAATDLKSPSVQSDDGSEAQGGVPASVSIASVTAASGT
GKAGASGKSTMPRNRRFSNLGIGGFRARSTRVSQVKKSAMAAANNSAGDQANAQYTIENFTTKQNRRRQQKKKSEIEENYPDYLMQ
AFLGTNLLNQSTEMLKKMQSGHKVQSGGHSQPGGHLHDNRSFGSAGSVHNLNEEVDDIDDMERGYDNDDSVNMNAAQEASEFMLL

EVPSSVSTIKCEMEVAATPKVQSVNITSNICGQQMNTSNTMLGVNATNSGVSGGGNNSGLRQMSVEVGRYTPQKTLNNHSPMIQTH
FPRKTGEIDESLRSLGNAGDGSNMFSHSSGQAQVSFHGGGGIAGGGGGGGGAQTVMSFDDDLITSDDYLMDEYWLSFNDLGIYDEG
AAGVGVSGGGVVGGNASNEATHHHQQTQQQQQQQQLLSSTSSSFTELLIDSSSDITIHRLLKTDKYIHTQESTSSSSLLSTHANFL
PVTNQNQATINEPVYDFANDL

>TRR (*S. kowalewskii*)
MRSPIPSPAPLLSTGPMAPHMQVPNQMSGQLLPGMMPVRGQAPGYSGIPGIMLGQGLPHMQGPPGQMQGPPVTTQGQLGQMQGPLG
QMQGPPGQMQGPPGQMQGPPGQMQGPPVTTQGQLGQMQGPPGQMQGPPGQMQGPPGQMQGPPGQMQVPPGQMQVPPGQMQGLPVTT
QVPPGQMQGPPGQMQGPPGQMQGPPGQMHGPPGQMQGPHGAMQMFEALPGQMLHSPRGPGEVTMDRMSMPPPPPPRPPLPSGGFAP
GMPMVRQSLTKKDQSKLIKEQKLLREQEQERQWKLLQQMRQQQQQDVPKHSSAGICTLVRQNKKKLKDIDQQNKTENLVESGDSQG
KSVETSEPAEPSGDAESQVAHTWSMANVSHVNTPSPKQGSPQSTTSPSQPVSDVNTQLDAPLSQLPHSQSQPALQQMQQMMQQAQQ
RSQIDNVSQQVFQPGDARFPVSERFPIVSLPDQNLPGEYPKAQLRELLVKQQEQRRLRRQVQEQQEQMMRDQSPVGLRHWPSDQPP
SQPPPPYPSGAHFTEQQRQQFSLSRDSMMQDQRQPPPGQRPPLQGERPSFGPGFPFGSDQRPSMPRPQWSDNFSEMHPDQGMRGRP
QWRQTGSMPPQRHPGHFIELKRHEHGGPGPNTGPFPGLGRPRNPYMALVQERLHQRSISQPSSQQMPTPRKNAPHGDLSTNPLDPY
DHLVKPRPPFSPQHQSGFMEGPGFDNTIQGKFSLKDPSQQHPPLIPTSTSAPASGYSSIPGNPMGQQTFGTPRPMFPGQPQHQTQS
VMGQDQHFRHSSQDSFPGSQDDFDTNVHGEAPLDDDLFKDGEFNILEYADPELDTNRHSEANQQLFKDLGLSEEGDVKSQSIPASG
QVDAEEKEPVCGQSETKEQIDDGLVQDEEKHITESQDGTESKNSSCSDHKSGESVNGGNKSEGCRDESCDRKDNCVQKENCSEEGC
GKKEGCGENKGCGESKGCGEKEGCGEEKGCGEEKGCGEKEGCSKKEGCGEKEGCGESTGCGENKGCGEKEGCGEKEGCGEKEGCGE
KEGCGEKEGCGEKEGCGESTGCGENKGCGEKEGCGEKEGCGKNKGCGEKEGCGKQGCGEENGCGEKEGCGEKEGCGEKEGCGEKE
GCGEKEGCGEKEGCGEKEGCGEKEGCGEKEGCTKQEGCTKQEGCTKEEYCEKKKDCDEKSDLPSTVKTEPQQETHQSTDASYTTSA
NLQHSKPVKLEPELAECDSHEDKKDKCSSGSSTPVVSTSNSTTPAVTTASIPQNLVDFIPLSQASSVKRPLLLEEQPLLLQDLLEK
EKMEQMQQQTPPGTAAPVVRHPPLNDPFLSDIDFEKLQREAKLFASKETESTPGPHAQLLKAMTKQHQENLQQQQQHQFTANPPPQ
TFSGMTSPRLPGPHGKTDPPHAFPIMRAGIQPERLPTEDPVLQGAAPRFGFPTKQSPGMGGNLRPSAITTTTSGLHSPVGINMPPA
HALTNQPFSVPSPMNVSSPGLSIPSPMGTIHNSNSTSGGPSPLGMPSPSHGVPSPRAIGSMPVMVASPLGLQSPGPNCPPAVPSPG
LQPTMSGMISPSCPPDTERQQQTQYEEWLFQHQQLLNMQQKFLEAEAAKHRKKKKTLASRQRANKKNDRPFSEHDEQELEAISLQH
STTQKQLEHIRKQQRQHGILIQEYRMKQQERMQQQMGKPPIPGQPPFMQDGSIPRFPVPSMPPGAMGVPTSLQSPPPGPPQGPPPG
PPQGLPPGPPPGLPPGPPPGPPHGPPQGPPQVSVQITQDENNPFSEVFQERERREKLERLREQQERQRIQMLQQVDRARAIRRMEE
EDHIRDHIEGDNKQPLMEMPFFDQDVLPFSQSGPQQPQQQPGMNQQPPSFQQGRPFVHSGPQGHMSMQGVMPQGSQPGAMLGMQGP
RQQFANYPPMNPDGGFPVPPNFPGAMEMCPPIEKPKKKRTRRKKKEVDPYDMPPATPMPTDALDFGSTRPLHHIDEQQHNIQDRNP
IMAMSNMNANQSRDNSLGALLEATVNKIESDMPKVDTEQSDSQSTKSDSLKDENGEQKEENGKTGGKKDSKPTQNLLLKQLLQGPA
IYSPAGGGGVPKNPQTPESVSSPVVETQNVSPIRSVKSETVDDDDVSNVVLTPQQQKQLEMLEKLPYVKPTSDSDWIAKKLQAGIS
DRDILMEIHRQEYEKQQEQQQQQQQQQQQQAEAAPKKRKKTQKQKLFDKSKRRRGSLEDEADALYDESKSLLNTIRQLPLLPLCEPE
VSVNYSLCPIYGSSTLTGENQLNGTFGHSTIDGIADYYSHFPFGPPAPLTNPPTPPASLPPSPPDKIPVPMVNGNAHPSELNQRRG
PDGIHVGPFSEPRRYQLSVIPGADITMATADDVLQEISVHQRTITKAVNPQIRNSNIDIPPSLPTPPPTSDATPKMSDDGSRPGSV
VPSSSPECVQDDEPLRLKGLRLIGHKMDLGKVQSPSPTFPIIRPTPKKALSIRGMKPVPDSPAKSTTSSTPANSSVTLEISVTAAD
DVKGVISAVSNMLKITEPSNFEVSGEMPEKKPKESTQIKKEKPSNLQSLLQAKTQFCRHCDVVVLGEGVRKRSSDVLSVNDGDSDM
IFCSTTCYMQYAITHQTQQGIREHAAATVVNHTSKDLECKQEQLQSKTQIQQKIQQNEGAIKDEQKGNKGNFRRRKNGSMEDAIIT
QPLVKKWKDIRWKKWNVNIVPMKTYKTHSDQDIDDLIRKLGTSLRPDPSLIDHRKCVFCCEMGDGCTDGPARLLNMDVDMWVHLNC
ALWSAEVYETLNGALINVEQAHKRGVLIQCCYCSKMGATVGCHKFRCTNVYHFDCAIRDKCIFFKDKTMLCPLHCPSRQSENELRT
HVVFRRVYVSRDENKQVASIMQQGEKKHTLRTGSLIFHSVGQLLPHQLQVFHSQTAIYPVGFETTRLYWSMRYTNKRCRYMCSIHD
NCGKPEFCIRVIEQGHEDVVLRDSTCKGVWSKVLESIEKLRRDAVMLKLFPSFITGEDVYGLTEPSILRVIESLPGVEGILGYMFR
YGRSPLIELPLAINPTGCARCEPRMRTYIKRPHTLTSNTSSGSYQSTVTGDLASPYTKQFAHSKSSQYRRLKQEWRNNVYLARSRI
QGLGLFATRDIEKHTMVIEYIGTIIRNEVANRREDIYEEANRGVYMFRINSDYVIDATLTGGPARYVNHSCSPNCVAEVVTFDKEQ
KIIIISNSRIAKGEELTYDYKFDFEDDKNKISCLCKAVNCKKWMN*

>LPT (*S. kowalewskii*)
MQTIPKNGWKCKNCRICTDCGSRTPGSGPSSRWHLNYSVCDSCYQQRNKGLCCPICGKAYRQHTAHNAMIQCESCKKWVHVDCDES
IDISVYQQLKDDKLTTIYNCVDCRNKDPNQQHEILSDSENLLEITDSNSGYPAANVDTDSNMTETDIISDTQQIDDDPDYIPIGLH
MDMPEQSLLPRILDDGMDTESHQSSQDLCGMTNEDSMNSADVDYDPGLRSSAFMQKDLSSSSASLSSLLACNKRIQGMGKPMKRRL
GAGRPRKVFSSHGRGIALNKRRRHTAPPGKRGPKPKIRHPGLQPSIQISETKPQDSGNEKNKEEEDTGMHTTVVLFSVKDKFTLTQ
DMCVSCGSFGRDAEGRLLTCSQCGQCYHPYCVNIKITKVVLSKGWRCLDCTVCEGCGKASDEGRLLLCDDCDISYHTYCLEPPLQN
VPKGGWKCKWCVCCTKCGATSPGFNSEWQNNYTQCGPCSSLLTCPVCFKEYKEDELIIQCVQCYRWLHAECDGFHNEDDIERAADQ
GYHCLLCRPTTEPALKPPPPPPPLPIEESPSPVTQLESSKPSQQYSVEGIYLTESGFNQIKSLSLLPPHRRKQRTRNRFRPGPLSV
FKDQVKLSELDQSSEPAARDSDAEQDEEKADGSGGDDNASLTDPSFKKRKRRPYRPGIGGFIVRPRGRGGFLNKLANRKISAILDR
QLYNQNSQEGEMGGDEQGHSEQGPDGMGEKKKRKRRKKTQLEHAFPTYLQEAFFGKSLLTSSKHAKGATDLSAEFDITNKEQAALM
DTAQPKTLPVPLPVVPTSVAMTTTTTTTDVAQNENVDGNDLGNTDDPLNFLPHDDELFGILSDLNKDTSAIDIGAVTEVATPDTVA
TRDAARTTSTSDQMAELDVDDIDKIVTEGLAQMDGKEMDEILSGMLSPHEQTNTQTHLHPEGGEIPFSIGENAVVSTSTTVVTSQT
LGPQQRPTPPTTLPLHPSFPPQQPPPPPPHPHHHMGMGNPVFSPDFTPPQSPWPTPPVSEPDVETLTYNQRNVLKWEKDEPLGEMA
TISAVLYANINYPDLHKEYPDWTSRAKQIAKIWRKASTDERAPFLQRARENRAAQRLKEAQKVWKIKDLQSPSRVSVAVIVSLHPS
CFHLLTNKLLFYNLGLLSQMSIAASS*

>MLL3 (*T. castaneum*)
MEELGQALDPDDLMDLDEGSDLDEEEDELDDPLPSPFQFPSSGATSAASSPPNDDFPLPFGFSGKPLSVKRGRGRPRREGDRHDPE
TLSLLTGLLTGKPAQRRPPKIKKLKGCAPRRKFKPFEFDFNSIAPSPSDEMSLDPETGALIFEKEKPAPVFEELPYFPEQWPGKVC
AFCNLGERSQLGQGEMLRLLCPEGFTPQRVAQDQGDSGSSNLPTPERETGDKSPRGPVTCRRQKSFNKCRHPSLTSEYVDELTIIG
YTEEPHVSTLFDSSGYFYTHRSCAQWSHGVTKNENSALENVGPAVLQSSSRKCSFCNHYGASILCKIEGCTKVYHFPCATASGAFQ
VLNLLALFCSNHIGQASLECENSVCYTCKTLGDIANLMFCSSCGEHYHGICVGLAQLPGVRAGWQCRKCRICQVCRMTGDETKLMT
CEQCDKIYHSTCQRPIVTSIPKYGWKCRCCRVCGDCGSRTPGAGLSSRWHAHYTVCDSCYQQRNKGFSCPLCHRAYRAHAHREMVQ
CTLCRKFVHGTCDPEADLVTYHQRKEAHPEYEYVCLMCKNLTQPATLLAKRNSIDDPTDCVAPPPESPFDENAEFDNPFPVDVIRN
MGLGKGKPYSASKIAKKRLGLGSTRPKGVGKNIPGKVGFMKRQRLTDFGRKRGAKSKMRGVFGVPGVGLQRPVSDGKNDEEPGVEN
RLVLCSAKDKFVLTQDICVMCGALGTDHEGCLISCVQCGQCYHPYCVNVKITKVVLQKGWRCLDCTVCEGCGQRNDEARLILCDDC
DISYHIYCMDPPLDYVPHGNWKCKWCAICQTCGATDPGFNCSWMNSCSECGPCASHVNCPSCSEPYSEGDLIIQCVQCERWLHGTC
DSIKTEEDAEKCAEEGYNCLLCRPRDVPPPHLIPCGTALKPPTPTKSPEVKSNNNYYVDGVYLSEAGFSLIKSLSLEQHSTRKKRK
KIATVQDKEAGIMATIESVVAGGSSDNFPEDSAKLELVDVKDEPQEIYKEGMIWMKDDGPPPDGFTLFTMENGVSVLRRKRQRNLQ
KLGIGGFLVRMRGIRSGQDNDDIDVLPGQSNPSGSDLSMGMPVDVEKPRRKPVRRKPKSKLAETFPSYLQEAFFGKELLDSTKELY
SSSSDEEKNSSDVDKTIQLTQDEIKAVAAVTPKNDRVEAKNVLEKKVVPKEEEEENTEDLKDVLAIPGDLLDTDIVNTIMNEGDGE
FNKNGESLDALTATNLPDESDITNTLASASTSKDTKDELSDILGPNFTLESMPNINGKDVEDIFKGVLTDDSQESQESNVFPIQNN
NTFSASNSQTIAQHPVNPPPVRPTALPTVNQSNMNSPLSFPPRSPYQSEYSNSPQFSPAFSEPPSPWVAVNDSGDLDAPSAAAQST
YNQRSSEKMKADEGLGNGATISAVLYANINCPEWKTEFPNWSDRYKQILKKWRTLTPDQKAPYLQQARDNRSALRMKKQQQSQMSK

DASVPPTTPTSKAPTVPPPPAAAPTVVATTNTMISSTQVPAIVSPSVSEVCTQQQQQQSRPTPSLTTASTSSVSVSPPPPPTQPPT
TPRPPPLTPSPTTQQQFTVHGQFLGPGDHCGPAPNVTVNPLYQPNVVNEQNPTLPIIRNNVQSTPPATDQQIRVLTPSEIMRTLPS
LCQENYEPQPLDQEKVALQHKSAREAEQERQWKQLQALRQQQAQQQQNVIHEQRVQSMARVQRQISDPSPFPQDQTTPELSPVTSP
SPNPRQLLTMGIKSPTFPHPVQTPMRPPPPTPPTPLEFTPEAGDPYAKPPSTPKLPQPSFQTRLPADPYAVQPSTPRPQFQIRPNL
QGLNAATVRNESPELSRQLRDLLQRQQFKKLDDQIMAGKGQQRVWPPVEQSQEETTPVVVTSAGDTTFRQPLPPSIARPRLPLPVG
GIRQPGPHLGLRMQLDPRIQGLDPRMRLLIQQQRLIQQNTAPGQQQQPQQQQQQLFSGGTVRFAPGIVRPTSVEQYELLQRQQQAT
FQPRATEPQMAPRLPIGQMIQRPPISQTPAVPTSGTTSSENQSNDQEIPDNVTAELEKLEQETGTMPELQGVDILGGLGDDDDELL
AEMGADFNILEYADPEALPGEKTNILDLELEEEPIKNDKKVKVELPKVSTGKPTPSSQQIKKDVTPPVTSVITASPALPNTNVHPP
PLAISQAQTAHPLPAGITSHLTPQQIHQQMLHQVQQAAALGKPMPLGAKLQSPDGIIGIVMPNNLVHLQIPQGYQQRLLVSQLQNQ
KLQMRLGQNVPRMMSPATTLHTNPVALNPGPRMNVPAPPPPPYPGPPPPYPGTAPPPQQQQQQMGLRMGPVHLPLGHIHPMRPQGL
PLHPHLQRRSLLLEEQPLLLEDLLEQEKRELEKQNQNQTGGTQEVAPSSSANVEQESTLLTDHDFDRITDVLGSPTMNINPNARQN
QVVSNPAQVQQWSQNQNRPKPTPTPTPQPTTEIRVQTFNANVIPPPPLPPENIVTEQDKQKQLVYEQWLSHQSNTLNQQLKYYETE
MNKLRKLRKSLNSKQRQLRKTGGQLADADANELQKIAAEQAVLQKYLDSSRKQSRQHGLLLQEYRNKQQAKQSGPGQSPLLMQQNP
SPLGPPSSSPIHHSQSPMMSPSASPLAQHSPLHSPSPMISHSPGPGSVTNILQSPSGMSPMQPSPRIGTPHSQGDSSPGPAPSPGQ
ICLPPPAPRMTSPQHKRVVTSPVGYSGDLRPGTPQMRFVRTPMDQRRLPSPMMYQQKPGTPSPLSSPPPQQMTMNQQLIQKIEGQE
ANILTSQRAAQLIQQRNFVRQQLQQQQQLQQQQQQQQQLQQQQQQQAVAQPQPLTPQQHQMMVQQQQQLAKQQQQQIAKQQQQQIA
QIQAQLQQKFQQQQQQQQQQQQQQQQNVAPPSPMPPKSPMINQQIMSPHSMPPSPMPPRSPMMNFNSNPQTPNSPATRNNYHMNQP
PSSPMMTHNQYQPPSSPMPRSPMMGMRRPPSANSSPMQDRPRSVERTFTNQGTLDNMLDQGGGGGNPNNNPIPAPPGFGRFGYFKL
GLRGGSPMWSFGRGAKRIPTPPGTNKDDKTNVGESSMPVKPRKESHLSKVSILKRKSPAKTNLQSLAVSKVGSLVSTDYNEFDDSS
CTPPVTPPPTTSRIKTVGKKISEVTDSKDVMIVHSPDAKQMNEIMDYDDDNNTVISGEVSLSSAAQQNDADDIAVIETFSQSDLGE
AMSSPLESDQIGDEYLLFPGNMVVDLSGDGETHYSDKEEAEEEDYGGEKMHVVIRSPTTSDDEYIMQNKGKKYNIRGIETADSPEQ
EEMQADPSPEGDEEEESEIVIIDPSIKSPDEQISTKEDFEELIDEGSRKEKIMKPEMVTAKHINEIHKYTNVYSFPPKSEAVSGAN
TSRVASLITSKLGQKFSLITTPVVTVSKVDTQNKTVGTTPTSKVTNIILHSAVRSNVLNVPKLVTSGSSAITIVSANTSPIASILP
SGFTVPVISASAVRQLPNVKVIDKPSSSISLTTHDTCLPKKIFEDDSVSPDSSNCEDDKSKDSSDEKSKIDSPPIEEKAKTPPVTK
EESPKPANESEVEKVHTTSVKVQEQRTSSPVIAKTSSPVIIHNTPELKMTAQVIQESQSGIQITATETVDGFEGTSDDTKSVVISI
PSPTPSQEQMLDNIALQALENRRRDGNGMPGEFESFEDVLDMIENITGEPPTVLEDNIKDKLMSGDNKSVTIEQEQLKETPQTTQV
ETPQITKPSVVPQLSPLSQPTDLTTNMANASQQLRNLLSLHTTTATPSTNVETVVKSVVTTGPVLTASPVVTPRVIQQVTTSTVIV
PNMKPKPIVLQSPVRTQTTTVAATTSIQITTSGITFVPRSTGIQLMPEHPKLPRVTPVVTSTPVTTESETTTRKTTCSLTEMLQSH
PAAVPTTKTSADTITAASLLGTSIGLSRVSTSLVQAQPIVVSPATCNNLVVTTTSGFKPITTNTNLLHTQLTKVMRHKSLDETNEE
SKTETEPNLDDKNKQETGCKFSSIQTTIAATTRSEDSNALLKKLLQNTACASTQTPPPSSVAAASITVAAPIQFTSTLESQLKASI
LPPVVKETPKLTQKMPIMTRETSFVSSPVVPQTSTTQQLHIDIKKCMPPSRTPSRDDLLSPPTPRSSCSQDSSLQTPPLIVKKEMP
QHLQMSQDVKKEYVDETSQHSEVSDQSRSDVPMKEELDSLDPMTEKMLSEKEELKRQKRRMYQQKRRQNQILNKEAIGQPKKRPRK
SSKVDEDYDTYIDGVLAQLRTLPPMVVSEPILNRNFSICPVFGSDMVKLAANDYDSRFGELKGSYGNGCMPGYSDFYNTQPYGDRE
PLPEKPPASTQRGFYDQEFPLIKFDTEEEKRFDMFSRDDTPDSIISSSSPECPFTEPPHKFLGLKLINEDEDEEEEDMSKMRLSPV
VPIIAPIPIRLKPAGPYLKDYTEDKENVGRDVCLKPKFGSGSTTPLKDSGNVTVTLTLTSSAAEDIMGVLRDLANILHIPAPTSYQ
IVERTSTPPSQKLGLYRTKGKDGKEGAPIDIQSILNGAAKFCKHCDVVILNNMIRKKVSELPFLSKDSELLSDGDELFFCSTTCYM
QFALMHRSPSIPEDKAAAIIDHLCQKDKTEANKRSFSDHRKNFMDNVEMELQYVKREVDAMDVDLKSHSLFNKPTDIISLRRYEQN
TPKVWKGVRYKSWTPGCLQPPTKHKKPTDKETMELLYRMGITVTPPKLPEDFRKCMFCQCIGDGVADGPGRLLNFDVDKWVHLNCG
LWSDGVYETVNGALMNLENALQQSLTQICVHCNKLGATIRCFKTRCSNVYHLACAVKDGAVFYKNKTTYCTVHIPKNEKDNELTTL
SVSRRVYVNRDENRQVAAVMHHSDTNNLLRVGSLIFLNVGQLLPHQLQNFHNQNYIYPIGYKIIRFYWSMRHLNKRCKYICSIHEN
GGRPEFKVVVQETLENDVELKDSTPKAVWQRILEQIAAMRRENQCVQLFPQFVSGEDLFGLTEPAVVRVLESLPGIETLTDYKFKY
GRNPLLELPLAINPSGAARTEPRSRSQLHWKRPHTQRTGGSSVRPLFGPTPALTNNNTLGDGACPYTKQFVHSKSSQYKKMKQEWR
NNVYLARSKIQGLGLYAARDLEKHTMVIEYIGEIIRTELAETREKKYEAKNRGIYMFRLDEERVVDATLCGGLARYINHSCNPNCV
AETVEVDRDYRIIIFAKRRIQRGEELAYDYKFDIEDDQHKISCMCGAPNCRKWMN

>TRR (*B. floridae*)
MPPRVVLDDNNPFSEAFQEREKLHFPSTNNYTNIKETSTILSLNGRMFANAIQAQRMPPRVVLDDNNPFSEAFQEREKRERMERLR
EQQERQRMQLQQEIEQQRRLQQMERQDHDWHNQEVVQNNMANERRQGMMELPFFNTPEIPDIAGQLQQNPNPQQQQQQPPGQQQGP
PFMNPNPQQQQQQPGQQQGPPFMGPGQGQPPMGQGPRFPPGGPNFPQGGPGSFPRGPLMHPRMPFPGPMGPIDPHQGFPVPPNYPM
SGTPPETEKPKPKRKRNRKKKKSADSEEPPPPSMQQQQGMPPDIPQNPMNPQAPTMPHPGMVQGNPMMPPVRRDPSQPPLLDQSFT
VPLRHIADGSQGQLDMNKHSLLLQHLTTEQKDALRMSATTGDNAPRIDSIDPDKLKAETLKLDPQKLETGNGEGEKGKEEDMEEGK
GDGNGDGKKDSPEKPAANQLLKALLQGTPTQNLLAKAGLPPCKALQEAQDVTVEASTPEKVKKEEQEDSDDEEVINKLKLTPEQQK
QLEMLEQMPETVPKGQNKWPTLNKKKEWDPKSLSSLPPAHSAPDVIVREQEEFERRRQEYQQQLKKRKETQTKRASKKRKKDEEDE
ITKMKQNRFPSTEMIMSTVKQLSLCEPEISVNFALFPPYGSGPLNGDNQLKGTFGHAFLDGVQDYYSNFLKSETLSNPPTPPASLP
PTPPPPGMARPPMVNGHLHPDEILGTRRAKLIGDQRSHDQDELMKEMVSHQRTITRAENSTVLGLDGGKTMQDVNVPPSLPTPPPS
NLAPGEAPRFSRDATTPDSIVPSSSPESVIDDEPPPQPPKLISLNRGPNRGPSPTIPLIAPTPRRDYSDDISKKPGLPSSCQLTFP
KTLSKPKLLSTHPPSGGLPPGPTNPTLVIQSLHHHRAPKLTSQTTQATALPQCGMPRHKSADSAEPQRSKVQGIHPPSAIQTAQLA
FQLSEPGKEAASASTSSTSFTVVVVNKTFLLQPNNAKLRGLPLVDSLPLPFAAKPPTEFGPPRDHPIPPTSGQSSFPSSQAGVFVR
PSFSVPQSVIPPNTSTVDIRTSVSMPRATDAVQVTRTSESAFSVSNLERMRTLVGQAGGALPPQRPDGNMSVSLTLSATAAEDING
IVAAVADLVRVPVPTSYEISEGVDYPSSMAEHFRMQQQQRKGDGGVLPPQGVSLETLLQQGKPKFCRHCDVVVQGDGIRKASSDFP
FLRDQEGRSDQVRTRFVAQDSSDSDSELTFCSSTCLMQFAISLQSRGRRETKEKAGSIVDHRSRDIVRPQHSEIPIHLSPTYVNNT
SKPFSEGQGSEGKSERPRLKRRSDSNTSQTSQDLPPEPPKVLIKKWKGVRWRRWEVSILVPKSTYRPPSEKEIDELMSKLGTCLKP
DDLPVDSRCCVLCGRAGDGDTEAAARLLNMDLDMWVHLNCALWSTEVYETLNGALINVEMAYKRGQTLACTACNKFGATITCHRYN
CKRIYHLICAIKENCMFFKDKTVMCPVHAPNKHENELVSLSVFRRVYINRDDSKQIAKIMRNYGEKKYTLRVGSLIFHNVGQLLPH
QLQAFHTRTAIYPIGFEVTRLYWSMRYANKRCRYVCRVEENQGRPQLVIRVIEQGHEDVVFKGSTPKLVWLNILEPIEKMRRGSDV
LKLFPNFITGEDLFGLTEPAVLRIVESLPGTEMLQDYFFRYGRHPLIELPLAINPTGCARSEPKMRTYIRRPHTLTSSNTSRSSQT
TLTGELLSPYHKQFAQSKSAQYRKLKQEWRNNVVLGRSRIQGLGLFAAKDIDKHVMVIEYIGVIIRNEVCNKREHIYEEQNRGVYM
FRIDSDLVIDATLAGGPARYINHSCNPNCVAEVVNFEKEQKIIIISSRRLSKGEELTYDYKFDIEDDEQKIPCCCGAPNCRKWMN*

>LPT (*B. floridae*)
DAGRMLYAGQDDWVHANCALWSAEVYEEADGSLQNVHTAISRGRQMRCERCNKLGATVGCCTRGCPANFHFMCARLEDCLFQEDKK
VFCKEHKDKVDGELVKEDNFEVHRRVCVNMDNIKLNRKWMRGLESTIIQVLAGKVFMVFSLDIMQKER*

>LPT (*P. hawaiiensis*)
MRGGTAAPGKGGKGGSSSLSRKPLGSGSTIRGNKRGKMVGSGVVNSRANRGRGGSGSVPVRGGAGGGYRSYFNYQQQAVMAMERET
SGGKEEGEENKMIVLDGSDTFSIWQDVCAMCGAFGQDQEGRLIACAQCGQCYHPFCANVKVNRVVLEKGWRCLDCTVCEGCGERND
EARLVLCEDCDISCHIYCMTPPLLQVPQGIWKCKWCAYCHYCGSKEPGLNSSWKQNYCMCGPCHSLTTCPVCLGSYKEGELIVQCE
ECNRWMHGVHDYIHTEDDAERCAEKGYLCQECRPPDTQPAHLLPASPTPSLPSSTATTTSPAHSSNSTATVVAGAGGHGASAAPGA
STIGSFMVGSGNSVDSPARSDSPSDLLPRQTNSKAEHWVDGICLSDSGMLQIKNMTLEPLKKIRNPPKPLRPPFRGPSVDSTSSRD

MDDDEDGDDEDRASNKEGEEPRHWVYGLKEGTVLQAQEDGSPPVLPEGFYTYPGQEPGSFILRKRRYRNIKTLGIGGFQVRSRIKR
DEERAKEEAQLDPKELEARRKRTHWRTKKKIKLLETFPTYIQEAFFGRNLMDTTTMEDTARVLEDMDVPPPQTEDSPPGDCPSPLP
GVLSSHQHIKLNKDELDLVAAAREKQQQLLQQKQQQLLRKASTSGAPGQSAAGAEAAANTSGSSNKSSSSKAPVKDEAKEEDLGDA
DEAIEALFGGGRNLLTDDIMRDIMDAEISHEEEGEESKEESSTPRGAAPKLTDILGPHFDLDDVADIMKNLPDEALEEVNNAGLAT
STAASSESVTSTAAASAYTGTVEVKTEDTGSAIVTAPSQTADQMCQLSSASGTAAAITASVAPAVAGMSPSASTFPQPLQDIKPTL
VDQPLQPPAQQLMASSASGVLPREAQAGYHVARSPLVHPQRVAPQQQQPVTNVILQQLQQPTQPHPTLIRMPSMGSEVNASSHGAA
AGSLVPHSPVHQQPQRHLHMQPVGEGSGTVAGSPVRTPTSGHFHHGAANSTSSIAADPIHSIISMPATSTIPPGAINAPKGQLIPL
ARALPSVPTPLMLPTSSSPVSTPSTSQTAAFPSPTAPPTTPVSAGSPHNTLVSAMQQQMLQQQQQQQVGDTMPVAGQPSGGVTTAS
SSDPSSLSGFQSSTEVPQRTNSTDAPLSGSLPATGEVQNRMLLQLQQQPTTVQQQHQGQPLQQQNPGLQLHQQNTALQPVQQQGMG
LQPLQQSGAGVQSIQQQQNVGIPAIQQQNSGIPSQPSVVQQSNVLPVQQQQRSL

>TRR (*P. hawaiiensis*)
MAGMMHLQQQQRSPQSMGMPHPNASPLGHAATSPMHSTPQSPLMSPSPSPAVGGIMSPMVPSPHTPVASPNVVAAPQHSPHPNVPL
HSPSNNAASADGSYSPHAMPSPSYSATSPYSQVPLYPPRPPFPTTGAPPGLLRQQQQQQLVRRASMPSPSASPSSIPPHMMNPAPN
MPIHPSKQSPAHEATGHSPLPNPSPRQNPSPLQNPSPRQNPSPLPNPSPRQNPSPLQNPSPRQIPSPLQNPSPRQNPSPLQNPSPL
HNPSSVPPPSNPGSNPNSNPATPRSTPTPDNSSMYQPQHPNSVPSSSPSGGGGGGGPIPYVSGPFSGGVKLGGLKGGAFFAKSRWG
YYKFGLKGGQETTSSTMNSTATITASTAVTTTTSNTITTSSALASSASVTNNNTTPSCATVTVKSEHEATSTTESRIVSGMGPASD
SDSKNNNDINNGSSSTEGMNTTNCGLLGNPISSHVSSQSAAPQVTLISSYSLPATYQLASHTAGSALTASVAPHVGRSLSNVGYST
GGGVMPALQAQPIRGGAMSSVGVISTGIGTCTGVLSVAGTTVGVMATGIRTTSSVIGTSIGLGPSLGVTGTSSEGVAVTGIAGIVN
ATSAVSGILGTSAASAHHGVVDSVVRSTHPALPAATAVLGAAAALHTDVSSTAAGFPGAPLNIAPSRLGAVPPLSGVHSIVSSVGSL
GTTPSSVTVAEGAAVLPQPPKSSIPPSSRHQAVIGIPGGGAIPSHVPTYASHLIPHTVDAHNPSATATAASTLTTPALASRPLPSL
PHLPQQVPRSHVVVGGNQPPVRTMLVKGGPVRMMMIRQQLPQTMSPMPPPPSRSPQGASPMPPPPSPSPHLPSSPNGGGVPQTTTN
RVSPAQSPSGNSPLMVSPAPSPNPVATREEGTQNALLKQLLQNTSSNATLPAGLQPPQPQQFVVVRSPVNGSQVPIQSSQVLVSQR
PPQQQQLYSTAPQQAFISQQQQVRIAGAGGIRFASAAPPTSQVVIATNVPQQQAVGAGANSPQHQPQSQQQQQPQQQHAQQHSPIH
TVHPALRQLPPSPTPHLDELVEEDRNSNKKRLNTSGGSASGASGSGAGTSQQPRRRSQSKEGANKSSARRPRPEEDYDSFLESTMT
KLCAMPALTVMDPIVHRNFNVCPIYGSDELVKLGRPDHDHRSGKLEGTAGKLKLKGWFDYYDTKPLGKSLPVLPPLKNPLTSKGFY
SMEFPPPKIGFPMDETCEEVSSLTNSAAVTATSGVGGASTPTRPSSATPQSMADSPDTVLSSSSPECVMPDSPPRFRGLRLIDMEE
EVPQDRGPSPSIPLLAPIPMKQGELPRMSSGQFFQLSFKGNKPPSSPKCDVKKEVSDENDPKANVPNGIKEESKAADDPKDGSTAV
STENFASIILKNRVGENPAMPLKHEGNVEVELTLNSNTSEEIAESLRALATLLNIPPPQVYDIKLQDDLPKMPKSWKVSCKNGQSV
EVNVERILNGQTRFCKQCSDIADKTPVIRRSCEMPYVSREELIGVDELTFCSEQCCLQYSMQHRIDMPHLSSLLPPGMEPDPQEAD
DSDSDSDASESSRGRKRKLEDCDDDDKTDEEEYPMQADGKRYRGHKYRIVNEDSMLPEKKEKPTDREMTELLFRMQITYRKQGLPQ
DSRRCTLCQLYGDCVADGPSRLLNHDLDKWVHLNCALWGEDVYETMSGSLVNVDTAVKKCLALVCEHCEMPGASVKCFKVRCSKVF
HLNCAVKEGCTFYKNKTVYCHEHSTKGEKESELSTLAAWRRVYVERDESRQVASVMHHTELNHLLRIGSLVLLNIGQLLPTQLQAF
HTPYCIYPVGFRVIRFYWSMRRLHKRCAYQCSILEVEGQPEFEIVVLEDGYEDLTLVSRSPKGVWNKVLTPITEMRQAADLVKLFP
QYISGEDLFGLTEPAIVRILESLPGIDTLADYNFKFGRNPLFELPLAVNPTGCARSEPFNRTQLKRYHGLRTSAGSIRSGGDIRTA
ATIANITMTMDYLKPFTHSRSFHYKKMKTEWRSVVYLARSKIQGLGLYAARDIEKNTMIIEYIGEIIRAELAEVREKRYEAMNRGV
YMFSLGDNHVVDATQTGGLARYINHCCDPNCYTEIIQVEKDRKIVIISNRKIMRGEELAYDYKFDEEEDCKIPCMCGADNCRKWMN

>SET-16 (*B. malayi*)
MNSSSLTCALCTRPIISASNVRSCSSCRKFTHSDCDSDSLNANTIETEYVCPTCRNSPLVADAMMISAGNSSVAASPTSISTHDEA
SRSIPLDNPYCMDSPNRSSPLMSLSQLDTFGDFFNIHQSTGQVPSMSPSGSTALSGPVSSCRASPGSFPEFNNTGDTADELRFSSK
RERSDNSSTVSRTSRGGGVKKPCGRGVTPRTRSLAGVNNDLTEVTHSHRGKRGGRGSNKNGRGRRPRGGKTSAMLAMVNAAAANRD
ADQDSSEMDDTRHRMEENDYIRTVVVTCAENSYFLQMPLCLICGSIGKGAEGTMVTCVTCSQSYHTYCVGLHDKLNSTIVRRGWRC
LDCTVCEGCGDGHDESNLILCDECDISYHIYCLEPPLERIPHGPWRCKWCSACRRCGNQISNITDNQNFCETCFTLRKCPKCLRLY
EIGDNIIKCQHCARWLHGKCEELYGEEMFETASENGFRCSLCRPQGNVVGDAFNILVCDNVTMNKCALEILHLRCGGALFRYPSSA
DSVYDFPSYFFAPRSHSFEEPIEEDDLESSGFGMSVGVSGRGRGMRSGGPGRRVMKLGVGGFFVKQVPRHRLSNAEDENGVEEDSA
KPKIKRFRKPRRSQLEDNYPPVIQESFFGMRPVEGRCLLDVVVDEPNLVEHQKPTLVIDRQKEACELNEMDSEALHESESLLGNIT
ENDILGNMEEIDFDNLDLNELLLDGDEDDDDDEALENALSDKMGPPSEDGQITAHGGPMRSQTFCSGQQTRRFTIASTSNLTQGSI
RSNVHMPPGYNQTGIYTQSRSSSQLGDTCTERVNQATEKWEEDEPLGERATKAAVLYANINYPQWKEQYPEWPERVKHIHRMWRSL
DASSRQQYVNKARENRANRVKQPRAKRITTSATTPTGVVRAIVGPNRYGIIPGQDNMAGYCRSEVEYNESFKVPTMPYGAGECMMN
QSASVGSVDVSIVQPYGAAQQHYYHSEQEGGSFASASQVRSAPVRNPQSLIAHLSKEVLDNYSLLQKRSNELAKYQLAIENDLARM
RKQKKNLTAKRRQLNKNNQQYDAHGQRIEIELSETDRHSLTTLANAIPIRQKDLENCKRDIKLHSSNVRDFEVKNNIPPELIPQMT
PVVQPVGLHVNAARAKQLQSGTPPTPNDMMSSNNGIRPPPSYPAFGMVGNYRPPYQSGAQMSEYNHSGYGTTPKSETETAQICSDS
PTPPSGRPGSSSSTEWQCRGGRGKRKRRDDDRNDECKRPAMLGDVAAETLTTQFERDIYDVVDRLITQAVIMVDGAEAARKSGMLK
RLLESQMKKEYTSQRNSFSGLEFNIPPKPKKKRTQQKKLGVGNSNEYELLLERISTQLRLCPQLPRSAIEPISRVGRSSYFMFGIT
DLPDRIDKPSIKGSEIGDVRLTFMGDYYTRLFQCPFDETSLYQTMSDFAADVKLPVLYRTRYIAPRGIQEREKPVSKMRALSVVRR
SGSPPLRQISRIPESVVPEAPPRHAFFKKRESPKKVGVSIVFEGPEGSLDAEAGLRQLMVLLGMEQLPDYEFDTPPQTPEPSKRDL
GEIQKTESPVADDFTCRHCGALTDQPAVRQTARQLGIIDHEPKDETLAFCSMNCYYNYMANAKAALSPEDLSRAERFIGAETLAKL
RQASADNFAKYLHLGKFKGQLESPGNSNMCSPLSDGRVESDARNNIQIIQARELSLLGDQNIQKALEKKWKGQLWMQYDATILESF
RELSIQTRQKQLIADLQSVLHSHCIDPTEDKRQCAFCGQHGDGALDDCGRLLNADANTWVHVNCALWSEEVYETSAGALINVEDAL
RRATNVICSICQLMGASLRCYKLNCEKNFHIFCAKSTRGKFMKDKTFICEEHDDYRNDLVLDHLKALRRIYIDREENQLIAKLFHS
EADRLVLRVGSLIFSHIGQLLPEQLKAFHNSDHIFPIGYNVKRIFWSVTNANNRIQYQCTICDKSSAPEFVISYEDGQEIRENSAS
AAWNHILVEVERTRDKDDGLLRLFPANLTGDSLFGLTEAAVSKMTESLPGVDQLITYTFKHGGVPLMDLPLAVNPSGCARCEPRFR
TLIKHRRRNQPPAATARSSVSTDVLSRDTRTRGSSSSLSSVDVSAGRGSCSTTLDCYGGRWEPGQLSTAYYSQYQKMKKEWKNTVY
LARSRIQGLGLYASRDIEMNSMIIEYKGEVIRSEVGEMREKKYEAQNRGVYMFRIDDEKLIDATMAGGPARYINHSCDPNCSTRLV
DSGPCGDDKKIIIIANRPISAGEELTYDYQFDIEDVADKIPCLCGAPNCQKWMN

>LPT (*H. robusta*)
MCHVHAHCQASKDEQSSKDEEHASTLVIVSFKDQFTLRQDMCLSCGSFGLGDEGRLIACVQCGQCYHPYCVGIKVSPSMLEKGWRC
LDCTVCEGCGKPHDESRLILCDECDISYHIYCLDPPLQEVPQGTWKCKWCIMCVVCGSNNPGTNCTWLENFSKCGPCSSLEICPVC
CMKYADTDLVLQCAHCDR

>TRR (*H. robusta*)
MLCGSSGDAESNGPGRLLDMHAAVWVHCNCSLWSHDVYETLDGTLVGSKRHINSMKHTGFCAVCARPGATVACFLPSCGQSFHLIC
ARKQGMCFLENKTTYCKHHCPRHLLGTMIHDLSIKRRICCERNEIKQVARMLKEDEPENYFIRIGSLLVNNIGHLLPQHVQTFKFN
TDVNIYPVGFKSTRFYWSYRNLHQRCQYHCSIEECSGEPLFKVKVVEHGEENVTFESNTCTGAWKDILDHLVELRRRHGLAHLFRQ
YITGEDLFGLNEPAIRKVIESLPGVELMTNVQMRYGKSSIYVLPLAVNPSGCVRTEPKSRTFFKRFPSASHNFNAGGSSEKSTWMG
SGEFSITQYMKQFTATKSQRYRSLRTEWRNNVYLRRSNIQGLGLFAARELEQFTMVIEYIGEIIRNEIAEIREKVYEQQNRGVYMF
RLDENKVVDATMAGGPARYINHSCDPNCFTEVVATEKGEKIIIIAKRKIRSGEELSYDYKFNFEDEKSKIPCLCGAAKCMKWMN

>MLL3 (*N. vectensis*)
MPHSCLMSTNLSRKLKTSAAPSPTRVWLAAGTGESIIHTTHQKKDVYKESTESSGSTHLNDFKRERLLLIGRRNVNALPPLSESFP
SSVALLTSVRSSCNIWVLTVGRSKRGPRRPRNADTPFTPSRCIPPASPLTPDSEEGFVITEPCTTKNRGRSKPSPPVRSSHPVKKT
RVNAQSKLKSPIHAAGAMGIDMPRREGECMTDHRKEPCSLMVDAMGQPVPLDASSAVQPSDLEEELLLKDKLCAFCNCGPERRMFQ
GKLQRCAPTPGFNPFKKTAQNTTQQQPNFEMVPSEPADMGKEAMDDVPLPQPMVHRRGPGRPPGRGRKKGAVVIGLPLRNSSQNMT
PDKYTFGTVKEHEVHEAFESDGSVWAHHCCASWSEGVCQTDSYDLVNVDKAVYQAMTEEIHSMTMLCPDHLEDAGRLVTVDGSPVK
QVETTKSLGVTIDDKLNWKYQIENITKKIASGIGHQAYEARAYRQISDVAMVKCEICEKWLHAECDNVDPDTYRCIDSRKIPYHCP
SCRVKKHEIVTDNGSFYDNNQMSDDPVMPQSLSDEDKRTLTSAAIGRITPSPMDDELPLLHSSLDNNLDDMNPLRTFSDSLMDTEM
SMHSLDKMVSDTKLPMEDEMEVFGSSSGLPDSPPHSDNMQEEKLFVSASYCAPTRYTTTPPKNSGKRKPGPKFKGKTGTPVRKKPK
HEKRGRPPTKLKKQNSVQPSPGKENVSGVANSAEPPTKEEESEGLHSTLVLFSADDEFTTKQDMCLCCGSFGKGPEGQLIVCSQCG
QCFHPYCVGVKVNKMILSKGWRCLDCTLCEGCGKGSDEARLLLCDSCDISYHTYCLDPPLEKVPPGGWKCKWCVSCDDCGATSAGT
QCEWQSNYTQCGPCASKTSCPVCNIKYNLNDLMIQCLHCDRWLHGSCDGLMTEEEVDRAADYGYQCLYCRPKTKCSLGALSGPPSP
RATALHHYRPQFLSELHPPQPQAKPIHPCQRLTNSGLVEMKLPIQSTSPNLGMTPQPVADNILLTESGIKQMSRLKISPASRRRAK
VKNKIKSKMETPLSPLKKGVLSPLNDEQMNIDYYNDNAVDYTRQLTATTLKKNAHRPGVGGFLVRGRGAAARKFGGRGSRGPRRKG
LDKMLGRGRHRESQEQLHEAMMDTTSNAAWADEADKTRETSLQEAYPSWMQDAFFGKELLQPHSKKKKAKHEPVTPSDHQTNKQPS
QGLHQTPMADLSAMNQDQTGRDQQTSWQSAGAEAALPTEGIPIPDDLPNPDALYRSLDMLTAEVMPIGTQDGPTVEQNMHQEHSLV
NSTHTRTNHGQPQTSDPLIQQPHPAQANDLDFLPSDLEILPHVDGQEVEDIFAGIEMEMKCRETGFIGIENNSQGCLPNNSNDQIS
RQPTDMQRPMSYHGNHPEIGPYTPGPHGPGFERSFPTQYPQVSQRFSQFAAERDSFQVQPDGRVSHAMSNIGQYPFENDPPGFPKK
QEFDLASDLPRERRRPPSPRADPMKQQQKWQEDEKLGATATISPVLYANLEHPNLSAEYPEWTNRYRAIARLWRKLTVEQREPYRL
KARDNRVRLKDQEKRLLRQSESEKKHAALEQPPTPAPQVETQPETPIHQAASPFVFRDNQQPKPAPPPLTLTEQHQMAVGDLPSTT
HQEMQPLSVPPRSTTPTTPLSSSGPSTPISTVAPMTPLTPTSPRGLDMLAMQPGYDASPASKNAASARSASMKAKRSKEEKEREKE
RRKQKLKIQQEQERQWKLLQQRRQLEHQRQQLQQQQETHMMRQELKRTQQAKGLRMGEEKKKKSPQKAGKKFSSGPSFGDKEEELS
GSSMDHHSEGAMQPSPAVSTSSSLDSHLSASLQDSFGGSMPSDGLMDTTSSTGISHGLEDMGPLSPATSVDNVVAMVNQSPMQQVA
GGNWNEWNTNSKTSNREGPQQTHTTPSRMGQPPQAVDQSMVIAPRQQAPSVSQEDTGVMQSWQGRPIPQPSDCRVSHEQALPHHSV
GFTYHNQEQMWKGSALSQPVQLKGMQVSSEDIGMSLTPSVSGGHMTAETTPSSHQLPKSQGQWHSSNIPSSMAQQVAPPAMENVDM
HQEWQSQTQVSSGVPPLKHQQQSGPSLPMSAQQPVMQQPINQQVQIPQQGMPVNIQPGPPPAPPPPVQPVGAQQQPDDQQQLELL
QKQQREQLMRAQQQQFLQMQMQQQQALQQQQQQQQQQQQQQAIAVQPSFVRGRMNPVVNTKDMPEIDSEEEHQERLRFLYRQRQLQK
QQMQLQAQVAHQTMFQQQWNTVQGFVPGALPADPNNPQMLQPQIAQMRPQAPDMMQQQLMMEYQRRMQQRHDLGQQTLQYDKVVQE
LHQQQGVQMQVNQPQGVMEVSHQRFALRHPAPGLPPPPQQVPGILPGQVPLPGVDGQQLYQMQLQQQQILMQQQQHPKSTIHPEAE
ASQQTIKDQNLAAKETLSNSTVCGADKEKLVPVKCDAEEEAKEGAVDEKSDESKTSTEAKESKTDDGNQVLAGKENTEEGSKSLPG
SADKPLAKEELGEISDKCDNQTENSCGRKADDSRKDHGVDSKDTSITTALNDSPTKEGTIESVSNQSQANPNTEKKSQETGPDVNK
VENNIKTESCTANEQLNKLSNQQTQQQQVMQTQAHAMQPSQPNPQQLQQQQIHPLSQPNPFPTPTSGPPQSVANNQQQVMQQQLVV
LQQKFFQMQQQRQLMMQQYQHLQQQLQQNQDPMIGAQLIALQQQGQFLQQSLFQVWQHMQMVQQQLQVPGKPGIMVPGQQPLFPQP
VSSPGPNNQALPWSQDQQGMVRSGMPGAPLMPGMMPIQPLQRPEGVAPSPKPEKPQRKPRARAKSKDSLTPQASPLPGQGPPADWM
NSPPAASKPGPKPRGKKSKKQSDLDDPEIQLQQQRLQQQILQQIEQRHRLIQQQQEQQQQGPVGPALGMNQQMMGPLHMANPRPQM
QQPTAQPMGPAAPRHMGPITSSMATSSPNQISIVSQNMGSVSSAPQGTHLGPPMSGTVMSQPMVSHASQNFGPQSQPSGFSMPQQI
VTQGHMAASSSPAVLSSASPSQFMSMSHHQQPGTPAPHMSQAPTPQIGVSQGDTVPTQMGPPPSPHLPVMMNQQQGRPSLDEQRRS
EPGSPFVEPQSPRTPSSTDPFGQRFPTPDQYLQFTQQQARQRQPIVEYVADANNPFSDDFRVMPVKAKPARAASKRKRNTKEKNSK
LPLPSVPEGKILTDIPPSLPSTASQIGEQVVNTRTSSNEPSLPEAPAKDKADSVIEDGGVQEKKPECDLQVPDSSIDSTAATQASK
DPTSCEETTQAKTKKVACSGALQKLESMVADIATEEEACKDQETRDGDEFDFDLYNEGTDQEVANLCRLAELRSPLQQMERDHNMQ
PSGSQANSVISLECHEEVPSFVSPVQDDQQTGKKSAVNSREGSPLPTNLEIGNASKSVSTVKSEEQEQSEASCVQSSSSICEENTS
KNNDDCAKQGTLSDRVLSDDQAAKAVSKTNQDSQQIEQDKEIVSSQEPNDTATMIQQDPTSITRESDSIPGTHEAKNESTVKPNSG
TLNHSGNISSSLDLEQEAKAIISTASSTVPSIQTESVISVSHGTLEEHLAPTGPSVLSFNQINPGEPSYNFAQREPNISAPVPDHD
AVMVSGNDPMLGGQIIQAEGQGALMTSRLAPKAVAPKPRTPSKSRSRKTSKTENIKPVVASVMDFRKAEYERKKREFEEQQSKKRM
QQKQLREQKIRE REERKRQRMFGNSPKTKTRENQLQKSRKLENIEAESVGDSVSLPLCEPKLLLTHALMHPYGSAPFNGQCTLNGT
FGNASIENTADYYSRFPSPDLDVVLSKYAQRKPLVNGDVTVEGQVEVMGLESKVPLKLRAMNMVPLKSVRELKTERFHSAENGFST
EHSSAASSPDTQYIMSSSPEIDASSKPKNLSFPALSNREISESPVFSPLAPVNIKREQISAFGTNSCSGSSRADNPNDSRTCQPSR
AGPSLRLDQDIDDLDSINVTLTLSPSSEMRVSETVHSVAELIGCSPPRPSDIVIEPSWKTSVTSPTSNSMPSRSYSFPVTSPTVSK
TPYPYSQFDVTLDKKERKKPDGPYCRHCDVLIIGIGVTRDPIDQADIEQNEQSAQKDFRVQSVNVEAGDSRDAFCSEACLKQYYSH
LEATSSGMKTDDVKSLSAASLAVSIGQSTNSTSMEAVMAVGQENRGQHAAEGLPASAMLKLRRPSWKEEDSLGEKEGKVRRDTRWK
RWQKSDMKDSKPKQELNAMEVAELMEKYGAVFRPMDLPEDTRKCTLCGASGDGESEGTSRLLNLEVDVWVHLNCALWSLEVYETLN
GALMNVMLAVKRGINTTMLCHEHRPDGDREDRLPSYAVFRKVFIHRDEIEQIARMLRHNQESGGDKLQTLRIGSLVVKSLGQLLLH
QLQTFHTRNAVYPVGFNSIRFYWSMREINKRCQYHCKVEECEGMPLFTLRVNEQGHEDTVFKGKSPRDVWKQVLEPIHKLRKEAVL
VNLWPAYITGEDLFGLTEQFVLRIIESIPGVDYMQGYNMRYGPSSIMEMPLAINPPX
XMLRHNQESGGDKLQTLRIGSLVVKSLGQLLLHQLQTFHTRNAVYPDLFGLTEQFVLRIIESIPGVDYMQGYNMRYGPSS
IMEMPLAINPPGNNRFSCYLPRARALRTTTGGSSSSSSSRATAGSNAAGNSSREAEESTSLYMKQFVYSKASQYRKLKTEW
RQNVFLGRSNIQGLGLFANRDMEPGCMVIEYIGSIIRNEVANKKESIYESQNRGIYMFRIDSDSVIDATIAGGPARYINH
SCMLTYDYKFDFEDDEHKISCLCGAPNCRKWMN

>MLL3 (*H. vulgaris*)
MNTVNGVSNDGLGLVNNSPESFSGHTSSQEFEPSSPSKDDDISNEVCALCNCGYEMTLILGQLTRYRVSQSTLKKSDKNKYDTPKE
ETEIYSSDSLNNRKDTKEKVFHKRRRRKQSIDADLNLPPDMTAFGTATMQEFADVLDNDGFVWAHHCCAAWSAGVCQTDSYDLENV
DKAIAKALLKKCSHCLRYGASVECQVQRCSKLFHYPCCASAGAFQNITSMILLCPEHLDRSEKLVDAEAFCDTCQKSDNMQSQLFC
TSCGRHFHSYCVDMNIPITPVVRMGWQCSFCKVCQGCKQPGDEEKMLCCDQCDKGYHIYCLNPPISVVPKSVWKCVSCRKCSDCGS
SKPGSGPSCRWHNNFSLCDRCYQQRKKGQSCPICKRAVRLFNNGDAIQCKKCFKCVHGECHSPLDDGAEYICPDCIEEDKDINADS
YENESNSWYDQEFLSSGQKVVSIEEAELFSLAASMNFIDSCLQLDNDNKVYSGDEQKFDGKESESTLDSPYSSSLISDSYQSAQSF
RGRGRGHEQKMRTFKKSFRGRGHFSRNPNYTKDSQITSQDTEKEEKDLSNIVKYSDKAENGTRTEEDEENEVVGLHSTVVIVSTRD
SFTLKQDICMCCGSLGKGSEGYLIGCVQCGQCFHPFCVNVMINEVILTRGWRCLDCSVCEGCGRNSDEGHLLLCDICDISFHTYCL
DPPLEEVPIGGWKCQWCVVCEDCGSKDAGFDNLWMANYTQCGPCSSEQYCSICSYAYEANNLVIQCEKCNRWLHGECDGLYTEEDV
ELACNLIYHCIRCRPKTKHLGIYTNLKASTISKVSIKKAISCQNQIQRKTKPLVSLKAKKLSRSELKEQKRIELEAMPLIPQHAEI
NECGTFAIDLSNLKARLRMDDIEKPNFHLTETGLKEIKTQVIRAPIHRSNLRKERNISRKREETKNNSIMNEEVSCSIDPEDPSAS
VLGASFDLEGLGVENDLLSPGMLSVKQEDLDMDLMQPDTTQELDDFLNDEENEENVFAPEPEDVNDPFEFFGVNLLEKHLNMELES
LETSTNSKQLKKEKKKERRMKRAAKLAEAARIAFDTMRGPSVTDPNAMTIPLQKKPELVETLDLITPEMNPQLMITELEQPASDVK
VEEELENKTPSSPVDIAAVPSSNLSPNHPISLDQNEDSSHLELMSVLDLPHVAGHEVEQMFAGLEQEMNSNKFITNKMSKSDIPSR
MILSSLSSSGMPQPGLENQIKYSPSLLMSNISPHSSNLIQNIYPMDNDFEGKKSDGNGLNPKIDTQKQLRKWEIDEQLGENATISP
VLYTNLMHKNLKNDFIDWPNRARAIARLWRKLPNEQREPFRIRARDNRVKLKERKGIRRHASRESPMTQNPQLQNSSLHDSQHNSQ
MKNSSMITSSFQNLQQNKNLHVSMPTMQMCNRFDDGEIIGSPKYSVSNSLSPTGLTTSLQANIRSGFTESPVASNISTPPSIFTVP
EPQKLSPLLQRLSPHCAISSNNDLSPQLANTPTHANISSATEPTKIIIKSDSKIQLPSHKQEHYQQPIENALQEPDKKFQPLQQKR

VFEQQKQTLQKDEIISKAEQKETLCLSQRDDTQTISTQSNISSRNRALNDNRNKSPLSNVNITSRIMKSPAIVSPPDSSVSNHDSD
RNADIERYFNSRKNLDYSGSQMHNVDQQQAYMLQQQYMVAQQQYIQLQLAQQSTAPDGMYGIQKQPDKDLPDVDNEHEHQERLRFL
FHQRQKEPSFGPQTMTPIKNTLANQDSLTHHYDPMHGQGSSIMGGNYMPRLPHFPVPNPADSPAYHQMVVEYMERYRKSLEMDTAG
PPLSQVNSTMMAEFTKQYQQHHYNKYYHSRPHQSYFNFSPRPDYPNYDERMYTGYAEKGPHDYQQAFPRSNILPNKSQSVIIDEKN
RGYGTATTLQRSSSFQISRIAADSTSNDEFLPGKEKDSEQKNCNISHKNDSNESEISSSKATDGLVDNEGSITSNREHYGSDDEDD
IFNRNQNKPFMKRSNRAENELKKINERLLKKVSVMLSDDEEAFLGRNQDEYISDCSSHVCSDSEEEILRVSNKKLCSDHKKDENEF
GNNKKTDIVHQNEQEILGRGTSLRVNETYIRDQQIAENHNSEEFKHHLYHNEKGTVQSSSFPSTDKYTIEQPGLTNGGSLSRLQNN
HIPLFTNNQPTSYLSNNLYSSQNSYPHPSVYQLGVMGYPNEGNDKHFESRQHSALKNESQPDWGSFSGSNKNNYQKAGVLHQSQTK
YSGFSSPSQNYNAESVSTSETLKTSFSHFHSKSSLPQSVSSVSSASNPSFSNSPDSSYFSPTITSSANNSINSDKATVSFSTTTSS
NIYSTCHPNSMEISGNSMFQQGSISGSMQFPSSGIVNPHYSNPNGPYIPQSTNPLFPESGMAPFTQSSSLPPLSSSITSNPSLIQS
SPIPGSVPQSSSPFKINSSPQKSNIVSLHPNSQRYHQIFFYQHQLLIMQFQQYQFQLQLQFQQLNQQQIGQQQQFLLQQQFHQQLM
LLQQQLVQQQNQLYRMAANGIQFTEHQLNAMYQQQMLTQQSYSQWIVALSPEQQVLFQQRMQACMQQVQHQMAQILQQSSRKEEEQ
KKKPKPRRSLNKKKSAALLAADSVQTSQPSTKFPVPELSEEMKIRTGQNLVVSKEESSHSEPNQITSKPDSILPTNSVSSDSKPNQ
QNTCKPKTEEKTINPNIISPQYEKIHESFIRHQKQFRGEIPPGPSPLSGYSRCDTQPGQDFQRTQQNDISKPVTTVLSGEISPGSM
HQKAEQLLNPRNTAIPPKPSFFPTHPSKLFKQATEKCEKSPSQESLVSNAILPMGETPSFGSRPGINQLSSEPSRIRNPFLNTKTG
ESSPSNNSAFSVSNLLGTPKKQSLVSTNSQPLFEPAKKAIITSEHLSVSQVIKQNDSDPGTNSPNDLACDNMYHMTDGETEKYEGM
NTQLKNLIKLHNKKKDKILEEKKRLRELERIKKKEEKEKIKVIREQRKAEKYQQRLLKQKALQEAKAQQLVIKKHNKKIIVPLEIP
DMPSGSKTPTIVSPLSSFTPKTPPMVVTPKIPLCEPDVHLVCPLVQPYGFNSSNTFKITGKFGYALIENVIDVYNEKPYTTSPIAT
SSFLKPISQCSVVVNDILNDVSLFPSPEPNNESSLMHSDIQKDKLLGSIVKEESESVRTISSTQNIMNFKDEDVDSEIEDTISRCN
LCYGLLHNVVFYVDKNMENVTECMYDKFELMSSEQLAFCSNECVEAYQRKVDYTFENDDQELIDAVKVSSMVTTDELHVFLPLPSD
EIRGDVCASGMSNKLVSKSNCRWKRWRVSSDNEKKTKIKSVKEDINQLMKKYSVALKPSESIRDRRQCCFCLVVGDGEANQSSRLL
NCNVDSWVHLNCALWSSEVYESLNGGLHNVDKALARSKQMRCSYCSKIGATLNCSYTFGAQQRLHCEKSFHFHCALQSHAALYKDK
TILCSEHKSLGIRKEELDSYAVQRKVYIERNITEQIARMLSLNNSFERKEYTFRVGNLIFHELGHVPMQLLDNFSSKDFVYPVGFE
TTSIFWDNHDIQKRRYYTCSISECMKSPRFKVKAHSNGSEGSLYTSSTAKGAINEIIANIQKLREKENCLKVFPEFICGEDFFGLS
EPTIQRMIESLPGIDTISNYNFRFGRTPILIAELPLALNPTGCIRTEKCRKHHTKPLAMRTAAEHTHTSDAIITDDSWLVYSQRNV
HIMNTKTGQYRRMRQEWRNSVVLGRSRIQGLGLFTKKDIEENTFIIEYIGSLIRNEVANRLEKHYERQNRGIYMFRIDNDTVVDAT
MAGGPARYINHSCDPNCVAEIVSIEKEKRIIIISSRKIDKGEELTYDYKFELEDDSEKIACLCGSFNCRKWMN

>TRR (*L. anatina*)
MQDDMDMVNLTPEQRRQLEEIDNMPLMLEREISQEEWDCLPEEEKFRIKELQRQMQKSKELLSPASKLKKMKGKKSPKEGSSSSGK
DGPPKKRSRKPKPTEDYEQYIQGIIQHLRTMPAVHLQEPEVAVNHSLCPAYGSGEVMTGENQLRGTFGSSQAGDFPDYYGHYPFTN
VHPLPLAPSGGHISRFMTEIAQRRQMMGEIAKQNQRAGDQQQQQRYRLLPGTSGAAGEGGAAGAEEGATKEPEAIARLISRAGSIT
TFAPSPVPSPPIIGPRLDIPSRLSNTPSSIVSSSSPENLEEDVKYPALRVIDPVDNTQEDNRTSPAIPLIKPTPVRATQMKVEKEE
MEAEVKMEKMESGPASQALAQIIKDKENMELMADKQSMANEKSESLSSVLGKSFADSVEKQVSVTLTLSAAAAEDMSGVIARLANL
LKIAVPPSYEISRSPSPDLFRTGRKHREEAINIQTLIKAKQKYCRHCDVLVPRSGIQKKVSEMPFLSKEEQESLDEDDVMFCSMNC
YMQFALSHRTSAGTAESKEASDVVEYKSSSAVQDDAKLPPPSIPASLSVSLTESGTGVVIKKDVKALKHRRSSSSLDGVPATPKPL
EKKWRGKCYKHWEPKMVPKAQAYTPPSDEELTKLIRKIALALVPNPVPEDTRKCAFCHEKGEGETDGPGRLLNLDTDKWSHLNCAL
WSHEVYETCAGALINVDASLKRGMTLECVLCKRSGATLGCFKVRCTNIYHVGCAQKDGCVFFQDKTILCPVHKPKGVIEGELKDLT
VFRRVYIQREESKQVASFSRIVHRQDTQHAVRIGSLILHSVGQLLPHQIASGKFHTRDTIYPVGFRITRFYWSMHHLKKRCKYMCS
ISDKDGHPEFEVKVMEEGMPEVAFKEATPKEAWMHLLTPMEKMRRDADLVKMFPAYMDGEDLFGLHEPIIVQLIESLPGTDLLHNY
SFKFGRSPLIEMPLAINPTGCARSEPKLRTHFRRPHTLHSSKTHSLPSTVTGVTSEVDSPYFKQFVHSKSQQYRRMKTEWRSLVYL
ARSRIQGLGLYAARDTEKHTMVIEYIGELIRNETAEMREKIYEASNRGIYMFRINDDTVVDATMSGGLARYVNHSCNPNCVAEVVP
FEKESKIIIIAKRRIATGEELTYDYKFDFEDEQHKIPCCCGAPNCRKWMN

>LPT (*L. anatina*)
MSSRNRQRLRKQDPERSTPTPEPPSGQAVGMEGDDIAKKVSGSRSQKDASSPTPGKKKRIRPPKIKVKIPEPPPPSAQPSTTEVGA
ELEISTAESELLESSNPASLDSSTVSQFDSSKEEVVELGEDEGSKGLSGSTGCLSGLGEEGGSEEGDQPICSLCACGKKSLLGQGD
LTRYEPTPNFNPFKKQLAKANRRSSSELDDRSGEKSPKHLTWRRARGPLKPGRERSKSPRRPNPHSITDADGRPIATVDELNTVGY
PEDTDIHRLFEPSGHTSAHHCCAAWSDGVIQAEDYTLQYVDKAVFAGTSQKCAYCRKFGATINCRYQKCNKKYHYPCASGGGCFQD
IKTLSLLCPDHTDQAVSIAGQEALCVVCDQAGDVLEQLFCTSCGQHYHGTCLDPPVQVNPIVRAGWQCPECKICQTCRQPGDDNKM
LVCDTCDKGYHTFCLRPVMTTIPKNGWKCKNCRVCGDCGSRTPGSGPSSRWHLNYSVCDSCYQQRNKGLSCPVCGKAYRQFTQKAM
AQCMTCKKYVHNDCDSAVDFGAFTPQQQAEYVCSVCKDRPPDDVPMETNLTESLNSMIEETDLVDMQDFSIFQQTEGSVNEESLQE
ADLQSSQESLLLDDNSMSSMDIDLNVLEKGAAQLAQLTNQQSPHLSAMLKSGSGAVAAASKIAGPGKGKPSSVGKKKVPSTTTGRA
RSTGKKAQKRPTGTGEKKRPAVKLKNKGGGGGGASSGASGGQAAASTTQVSASSNTDSGKEDDHHSTVVVCSKDDDYVLKQDMCYS
CGSYGKGDEGKLIVCTQCGQCYHSYCVQIRPTQVVLKKGWRCLDCTVCEGCGKPHDEGRLLLCDDCDISYHIYCLDPPLDHVPKGN
WKCKWCVICYTCGSTTPGYGGQWQSNYTQCGPCHSLTACPVCKKHYKEEELIIQCCHCDRWLHSICDGMKNEDDTERAADAGYHCL
MCRPKTGAPPPPPPPPPPKVESLPASPIHKEMDTPKSPSVQFVVDGICLSESGMEQIKALTVPPPPAKKRQRRPSKLPGDSFKHPH
MAGLLSPGEKSEQGDFIEGKEYDHIMDEETKSIAEETAEIAAVKSPEVPGIPEKKKRQRKPFKPGVGGFIAKPRARSHMIKKTLLA
KAAGIDLPQQRDTSPTKMEGGDDLKDILGGVDLSHDVHDDELMDILDSEDISKDEELMGGQHGEQDMDKMDVDKVLSSLAQIDSKT
VEDLFKGVLEHNPLTSPGAEGDGQSMMPMPPQSGHDQQLRLPGHGPMHMGQGEGHSPGGMSPRSGPPPFGMPPHSPYGAEMGSGQF
SPGFAGGPATPQWPHLPGDDFEPDSASQNKRNLAKWEHDEPMGDKATITPVLYCNIMHPELKTQYPEWSERIKQITKLWRKVPTEE
KQPYLQKARENRASSRLSKAQKIQDIRKKQQDPSAAGGMPPPGMPPPPPGMPPTGMLPPPPPPHPQAHVPGPKSQEQLEREQKFQE
QERQWKLLQQIRAQERKMELEALNSDSPRGSMTPGSTQGSPHPSTPESQGVPPFGPQQQQQQPPPQQQQWDPYAMPPGTPRPGSQD
GGNTNKPDNLNNQPGASPQYSQSQMGMPPGRPGEQRMPPDPYSQPPGTPQPGGPHPGQGRPQEGYPGPPPQQALKGMPDPYAQPPH
TPRPPGPEDPYARQPMTPRPQQQQQQHPMAPHMQGPPHPGHGLPPGHQQGPRPGVPYSQAQGMPPHPGDPFNRPPGGPRPPHEVWH
GPHPGMPHPYSQAGAVPHPTSPREFRLPGPGMIRPKEGGYGQGGMTPTPRPPTSGQNSQLQQMLSTAPRFADPNLQGLDPSTSTAV
LHQMFSSQRLYYLSQRHQQPQGRNATDSTVYKTGAGPGDGDPKSDAHRQHLRELLATKAKESQKQKQMEQQFRDGGGANVPLHHWP
SEEELGPQQPQGSQGQMSVSQGQMPVSQGQMPGPQGQMPGPQGQISGPQGQMPGPQGQMPGPQGQMPGPQGQIPGPQGQ
MPGPQGQMQRPQGQMPIPQGQMPGQISGMPRPPPPGYPGAQFGPRGPQGEGGWQEQMMRPPHPGQIRPDMEGFGPRGPQGEVMRGP
GPEMMRPGYPGQWRMTPPGMQGPMPPHGGPYPQGQYPPGFNPRHPGQPPWEFHPQHMTPAEQHRTPYLKMFQERLRQTQLQAAQQQ
QQLQQQXXXXXXXXXXXXXXXXXGTPGTPPHTSGNPLDPYDHLVRQKMQGTPPAVGEARVPPHPPSRPSSQQASPRGTPYSTPPTT
PHPGQSTPSANHGQQASDNLNLSDYSDPEIQRALSEAQDKTIFDEHLGLREDAEKSAAEETSGRVGAFQAQFLEFSKKNSNKESEG
NVEDSVGQEMPGDKASPVADVGSQEQIAGGEPSSADADKPPETEPSEASESTQCPDGHNCPGDENVPATNSGAGQCPDGHNCPGDE
NDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPG
DENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNC
PGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGH
NCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPD
GHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTKNVGAGQCPDGHNCPGDENDPTTNSGAGQC
PDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAG

QCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSG
AGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTN
SGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPT
TNSGSGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTNSGAGQCPDGHNCPGDENDPTTSSGAGTCPDGHNCPGDENE
PTTNSGSGQCPDGHNCPGDENDPTTNVGAGQCPDGHNCPGDENDPTTNSGSGQCPDGHNCPGDENDPTTNAGSGQCPDGHNCPGDE
NEPTTNSEAGQCPDGHNCPGDENDPVTSASSQSNTTLTSTVPTSDQHVSPGPHEAAPSSTTTSATAVDAQSATGVSDTAIPVAGGA
EPMPGSNVPIPAVPVSMPQATVSMPCVPEVSQIGHSPMSASSQGAQLTTPSGAPVNTEAQHAIQATIASTVSAPASMETTSSPAVQ
TTASTTEHVAQPVVTNMPPSIQPVVPSMGQVEHPGGSEVKVEPQGPTAVGEQVVPGQASEEPGQIEQKVASGTEIQKYIKTEADLT
VPQTCQEETKPPPALAVATTPHSAVPNTTHPPMPSAVAISGSAGTKPTTNDMMSSLAGMMPTTPGMMPASNTSMSMATGMMPSSGA
GPTSVGMPTSAAQGMPPLSMMQGMPRPGMPVATSGMGATPHPGMGPGMHFPPGMRPRMMPPGMHHQPGPPFPPHPGQSRMPGSGMV
PPTSGMVPPTSGMVPPGQQPQFSPQMPGGMGAPRKDKPVLLEEQPLLLQELLEQERRDQQKQAEQQVLIQKRDSGLMSDAEFEKRR
AELAITESRRPPMPGPRMPGMGMEGMPRPGDQWQGPMGGPREGMFPPPYPGRSQGPHMMPRTPSPFGSSPAEGGMGPRPPPPQGTD
PLSEAERREQYKYEEWLMQHAHLLSMQVKYLEQQIAKQRKTKKSIQNKQKQARKAGAELKPEEQQDLDKANQEITLVQKQLDQVRK
QQRQHQTLIQEYRQKQQERYGVNWPAPTPPTKPGQNPMGAQQPPMMQQPNPMMQLGPGHMMQGQGHPRPMVPGSPQMQQPPTTMMP
QPPSSMMPQSTTGMMPQSGAGM

>TRR (*D. pulex*)
DTPDTVVSSSSPECVMPEPVVRFPGLRFIDLAMDSAGRSSSCNSPDIPIIICAPVPIRPTPPTQPGKSSSFAPSAFGLMTGGSGQL
PLRDSTNVTVTLTLNASAAEDVNQVLCRLAALLRVPPPTDYRIIERPGAGTPGGGGGPPGPSQRLGLYRFKGKDGKEGAPVDIQSI
LNGTTKFCRHCDVVILGNNKITKKAQLFSAAGLDSSSAEPKKTTTATTIADPDDDPDGEDEDYHFCSSVCFVQFAVAHNTRPSAIEA
KEANAVVAHISSSSGGGGGGGQGQQVCVQDSGVSSHSHTAASAAAASVVAAAAAAAAVAVPAASQAQTPVPVTPPAPKWKGLRYRN
WSPNFQQQVARKQQQQQQLPGNHNKKMTDNELTEMLYRAAICIRPTGEKAIDKRQCLFCSGQGDGVSDGPARLINYDVDRWVHLNC
ALWHEEVFEMVNGALMNVDTALKQSLTQTCLHCGRNGASVKCFKLRCSSVFHLGCAVKEGCVFLKNKSVYCSQHLPKCSSTSSTSV
VDIKEEQLTTLSVFRRVYINRDENRQVATVMHSGGVALDAGEQSYLLRVGSLTFLSVGQLLPHQLAAFHSPNCIYPIGYQVVRFYW
SPRQVHKRCRYVCSIEECEGRPQFNVVVQEPDRPNLPDLSFKDWTCQGVWQQILGAIEEMRRQEGVVRLFPQFLNGEDLFGLNEPA
IVRVLESLPGIESLADYNFKYGRNPLLELPLAVNPTGSARSEPKLRTHVKRFHAHRTVRFSHSSSLASSLVMVDLPGPYSKQFVHS
KSSQYKKMKQEWRHNIYLARSKIQGLGLYAARDIEKHTMVIEYIGEMIRAELAECREKRYEAANRGIYMFRLDEQRVIDATLCGGL
ARYINHSCGPNCVAEAVEVERDLRIIIFASRRIARGEELSYDYKFDIEDDHKIPCLCGAASCRKWMN*

>LPT (*D. pulex*)
MCGSFGLDQEGRLISCAQCGQCYHPFCANVKVTKVILQKGWRCLDCTVCEGCGERHDEARLLLCDECDISYHIYCMEPPLDYVPQG
NWKCKWC

>MLL3 (*D. plexippus*)
MDGQDEGVLDVDLLADDGSDEECEAQPSSDFYAGPASTPTSCASSPRGEEPASPLAVVSQPTFFHHQTYTRPFFTSGRRGPGRPRK
EGAKLAREGKIVRRNRGSAGSVRGAKRHRTSRDEALDDMMDEDDFTMPAPEEPPYAPEKWPGKLCALCNLSERSQLGQGEMRQIVC
NTESEGGTTPGVSNSGGATPTSITTPSTPTFPLPPGLTSPSPEMLDSNQPQQSLPLSRRQKAFMKCKTPLYNMEHTDELSIIGWTE
SLELPAVVSSGMFYVHRCCLEFSPPYQDQVAIACNNDDDKEQLEEEARIRGIVIAALGRKCAFCQRHGASIPCKMSCSKYFHLPCVL
ASGGFMDFQTKSSFCKDHLHQAPLICTADIDCRTCRTIGDMANLMTCVTCGGHYHGTCVGLAQLPGVRAGWSCRSCRVCQVCRGEA
GGGAGGEARAVACEHCDKLYHAACLRPVMATVPKYGWKCKCCRVCSDCGARSPGAGPSSRWHAHYTVCDSCYQQRNKGSCCPLCRR
AYRAAAYRDMIRCSACRRYVHGMCDPEAEPQNYKQKKGENSSYEYTCPICKTNQQMAGSKSGSFDEDTMASTSQDSSFGDENTQEQ
DPLAIETKPDVGLGKGKPYTVSSKVAKKKIGGYKLKGGFPPAGKIGFQKRQRSVLDFGRKRGSKPKMRGVFGVPGLGLQRPQAPDT
KTSEDDPGVENKLVLCSSKDKFVLTQDLCVMCGAVGTDSEGCLIACSQCGQTYHPYCVNIKVSQVIVSLGWRCLDCTVCEGCGSRG
DEPLLVLCDDCDTAWHTYCARPALAEVPRGAWRCGRCRRCLVCGTRDTALWCDNYTECAPCASLVMCCVCSEPYSDGELIIQCTAC
SRWLHAACDSIRSEADAETCCRAGYKCTWCRGREVPPPHALPGSRAPSPRPRLTPHALGLAGEYYVDDVCLSQRGAHHMKQLEAEL
GITHTRRKRRFKNDNTEKDAEIMASIETVVQNVEGEMGDDTKPDSTAEVKEEPGLSSTNFKEGILWNVANDGPVPEGFTIYTTDLG
LTVLRRKRQRNLNKLGIGGFVVRQRQTTKAQADEEKDGDGTQGSGESPSNKRKPRRKPRSKLMEQFPSYMQEAFFGKDLLEPAKPA
MGSTGSADGEGLLGDRDNSDSESDDVLSALHTFNEHDTTVFITLNTEETELLQSLKPKDEKDDPSTLNLEHVKIKTECEEGQLKHT
EDSTALKNAILGPQPSPDHADHAPHPPANEQNTESAPQSAVMSAKEELSLLGVSLDTLPDMDSNDVDEIFKGVLTDDSQESQESSV
SYVNSMAGTPYSQQRQQLQSPMDYASPYHTEFGSSVDLSLMTRNVSYSSSLSPLYECGSGAGSWPDAPAPPPSYNQRSADKMRADE
SLGSAATISAVLYANTNHPEWKTEFPNWVDRCKQILKKWRALPSEQKAPYLQKARDNRSAIRMKKAQQSSGGADDGGRSPAGPPPS
ASPAPPPPPPPQPAPPQSPAVPPPGPAGAPVRAPNVTVTAGGLLEADAHIRVLTPSEIMRTLPRLAPPPAPPAHPAHSATPTPHSA
DTEQERACGQQRSAREAEQERQWKQLQQMRQQQTQQQQQIIHDQRIQASPHGGQPAGRDPAPTPTALPGPAPHPPAHPPPHATPHR
FPLYSTNEDINRQLRDLLQRHPEKMWPGPSGEESPAGCVEGGGPSFRPPLPVPLRPRAPLHHAHRPEHMIIQHRLVFSDGGQMQQN
VQNSSLQQIMEQKQVTQGNMSSNDKQDSGPEQGTDEMDIGDMDKLEQDAGNIGEVDILTGLGAEDDEELLESLTAEIGEQFNILEY
ADPELAALNDETLLDGLDLPDDVQHHKRGNTDDADQKGDNKMDSDLQTAVKNVEPKTEMKSEPEVKTELQDVKPEIPVSSSMGQMF
PDGVPRVITQNQQIALHSAMQAMQAMQAVKAEVKAEGDDKQDGDWPRPVFSHMYGLQLPAQTQVPNVRAGPPATTPQRPQGQFVNQ
HVQFSQGGVSASTQSIVNAMTAQVQAALAAGRAIAPGTRLLGADGAVGVVRHDRSVALAVLPHSAARGMVNAARPHAPPPLRQDTP
VAVIITEYCRPYKGIHSTVLPAGAARSAPPPPYPGLRAPPPPPYPQLQQEQPLLLEELLEQEKREQEREGGDWGAAGPGPAPPAPP
AAAAAPTVPLFAGSAPPAPPAPPDRALSDADHHARLVYEQWLKQYNTFAADQLRYYEQEVQKLRKIRKSLNSKQRQLRKSGNELMP
NDAAELQRVSTEQQALQKHLEAARKQARQHSMLIQEYETKQRQTNPQLAQQTIINQSQIGSQTVFAQNQNITQNQQMQQQTINRTM
QIGLQGSQIQQQTLIRTQQTVLGSDGQPMSQQRIGLVTSPMNTQGRILQGGRTLVIEQAGAPQQQLVRQLSGVQDRPQSVGGIVQF
SQQQLGGVRGPMPAGGPAPRTPARPAMSPLHVQSPHSQSPHHSLAPQSPLHPMQSPMQSPLHSQQSPLHHTSQSPLHPSTQSPLHS
QQSPMHVQQNSNMTQQSPMHPQQSPMHPQQSPMLPQQSPMHPQQSPMHPQQSPIHQQSPMHSQNQMTTQNSPMHPQQSPIHPQQSP
MHPQQSPMHPQQSPMHSQQSPMHPQQSSMHPQQSPMHSQQSPMHMQQSPMHTQQSPMQSSMSPQHFLQQLQAQRTSPQLPTPSRSP
QIYQQSQIQSPVASHSPQMQNADYTSGRFPRPTSGCSPLPTRFARPPHHQQGMRVGVPFVNRQQTSPLGSPQPLPSPGNNTNEMTR
QQLLQRQQYSPMGGAPSSPSLSRSPHVARTPTPSASPAASPAPHHDHGGGHHHHVAPIPPGYRYFKPGLYGGAPVWPNKDDDRRRA
THINKVSILKRRTPPRPRFTGKTAPCTDRSEDTGDSTDTRKYVLYSSDSVEYSTTSDDSARMSFSKDDDDDDIQEFVEHLDDDDIV
VVEPGAISPEERIATEEDFEELIDSGKPEDEVEEEVSQEKPITRTETTTKTVAVISLAPAKQPSSNMQQYKLLNPSQSSALTRLPV
LSATVISPHSNAKILNEVSQATVASVSIANHTISVPVLKSLSVPIANVTGHASGMAKSQIKKVPLNISNSPMTVNMSSSLSKPISS
VISVISSNVKNATVNPVITLATSNTVAGATRVPVPLLTQHQVKQKIVKTIEKPMALSLSNTKLSSLSVTHDATLPAKIFQDDVASP
DSTVATESNLEMQNPTSSRHIMSISPQDHSQKEQKSSILKESTKDIDIESEISQQTELPHTICLSDRTPLLLGKADIKSPDPIPEK
IPDNIMEGDDEDKPEDDLQSHLLLSYNKTIQDQSILNPPSTDKKPEENVVKTIKAIANQTKEMVIDSNMQITSDLTQDSVQISIPS
PTPSQERYLNDITMREHSETVEGNQKHVESFEEMLCLIENITDEAKKFAMKQSNPKTSSAPTQSSVSQAKSRNLAPKQEVKEQEQS
AFPPREVERLNLQSATVPQLSPLSQPAELTSNMANVSQQLRTIMSSLNTNTSKVETQTVMRKSSEPTTPTQVNFENLLPSSKVEVT
TSRPSPIQRMEKPQTSTSCSDPVPMSAAQVIGSRVNALSSVGQVRKSPTVSPISSPVGIGNQNLMKSPAQSPLINNQSYNVNDEPQ
TTSMSTQILQMPALSKIHSSSSVPTSLMHGNNIITTSVSNFQKSQTLPSSILGHTLLQPTRQINTSNLQFNPQSISSSQPPALVMT
SRTIIGNKEPAPNVTVRTHSIVSPGMNQMQNKVSQGSINFITSSKLLHTQLTSPLKRSKSTDEPKSEVIVGHIQPTKRHSVESVTV
KSEPMETEESNNVPTSTPESSVNKIIQQNVVNQRNDESQNVLLKQLLQTTTTSSSVVPQRTMTIQRTAPALGTIPSLEAQLARPSI

PPPTLSLSQEVDIPKNSPRHITSVSSHFATRSMQSTMSTSTLSPLIHTTTQSLMDVRKQPMKIMNKEETTPLPECSPTMKTLPMYP
MSIDNQHQPIKKEVAPPQQSPVHRPFTPLDVKKELLDESSQQSATSGVSTASDQGKLEQPMKEEYPESISMEHSSDINPSETPSEA
KKRKRREYQQKKRKQMQLNMKAAAENNMSSINAKKRPRKGSRYEEDYDTFIDNLMAQLRLLPPMQIQEPSLTTNFAVCPLFGSGDL
TKLKNQDYDILKGDLLGEFGNAKIPNVADYYNTKPFGDEEPLPEKSTASTQRGFYDQEFQPIIFDEDPEDKKLDFICKERDTETPD
TIVSCSSPEGYEIELDDSFPFLKPIDDDDEDEDEEASPSGKVSPIIPLIAPIPIRVKPCSLYQLKDEEDNQKSLKGLDSDAPLKTK
GDDSPSSTESNENVTVTLTLTSGAAEDILGVLKELAGILHIPPPTSYQIIERTATPPSHKLGLYRSKGKDGKEGTPIDIQSILNGA
AKFCRHCDVVILDSVVRAKASEFPLLAANKGNAEEILCDSDSELYFCSTQCYERFAWRPTNIILDGKTKIANVKEDNKSDVDTNLS
KDRDDFDTASTESMETDDLDIKPDIKDEKMDLSFIDSLDNDELMKEVGDDVSALNEDLKSMEQDEKSNQSVETEKFKGIRYKTWSP
GCIGPPIKYKRPTDRELTELVFRTGVAIMPVTNEDTRKCELCGIQGDGVADGVSRLLNCDVDRWVHLNCALWSEGVYETVSGALMN
VDTALVAGSNATCAVCRRLGATVRCFKVRCGSVYHLGCAVKDNCVFYKNKTAFCASHAPKNEKDNELTTLSVGRRVYVCRDEQRQV
ASVMLHSDTNHLIRVGGLIFLSPGHLLSHQLAAFHTPNYIYPIGYKIVRFYWSMRRANSRCRYVCWISEDDGRPRFHVRVQDESRH
EMSAPTPRAAWAAVLDAVAALRDGRKEAGEEGVLKLWPNYVTGEDLFGLTEPAVVRVLESLPGIETLTDYRFKFGRSALLEGGLAV
NPSGCARGEPRCARSWRRATRSLLPAPPVPTPQPTALTPHDPTHQTHPACPYSKQFVHTKSSQYKKMKLEWRNNVYLARSKIQGLG
LYAARDLEKHTMVIEYIGEIIRSELSEIREKKYEAKNRGIYMFRLDERRVVDATLCGGLARYINHSCQPNCVAETVEVDRHLRIII
FAKRRIARGEELAYDYKFDIEDDAHKIMCMCGAPNCRKWMN

8.7. APPENDIX G

Gene clones used in Chapters 3 and 4.

Clone	Vector	Origin of clone	Use
Chapter 3			
GFP	pGEM-T	Dr. Daniel Felix	dsRNA
Smedwi-2	pGEM-T	Dr. Mette Handberg-Thorsager	RNA probe
Smed-H2B	pCR II	Dr. Robert Blassberg	dsRNA
Smed-LPT	pCR II	Dr. F. Jaber-Hijazi	dsRNA, RNA probe
Smed-LPT(2)	pGEM®-T Easy	Dr. Samantha Hughes	RNA probe (pooled detection)
Smed-trr-1	pCR II	Author	dsRNA, RNA probe
Smed-trr-1(2)	pGEM®-T Easy	Dr. Samantha Hughes	RNA probe (pooled detection)
Smed-trr-2	pCR II	Author	dsRNA, RNA probe
Smed-trr-2(2)	pGEM®-T Easy	Dr. Samantha Hughes	RNA probe (pooled detection)
Smed-trr-2(3)	pGEM®-T Easy	Dr. Samantha Hughes	RNA probe (pooled detection)
Chapter 4			
Smed-NB.21.11E	pCR II	Dr. R. Blassberg	RNA probe
Smed-AGAT-1	pCR II	Dr. R. Blassberg	RNA probe

Smed-SP6-9	pCR II	Author	RNA probe
Smed-chat	pCR II	Author	RNA probe
Smed-TPH	pGEM®-T Easy	Author	RNA probe
Smed-TH	pGEM®-T Easy	Author	RNA probe
Smed-GAD	pGEM®-T Easy	Author	RNA probe

8.8. APPENDIX H

Phenotype scoring during regeneration and homeostasis following knockdown of *Smed-LPT*, *Smed-trr-1* and *Smed-trr2*.

Regeneration

	Smaller blastema	**Delayed eye formation**	**Absent eyes**	**Dented posterior blastema**	**Regenerat-ed pharynx**	**Out-growths**	**Bloated posterior**	**Inching/ Immobility**
***Smed-LPT* (RNAi)**								
Head	16/20	-	-	9/20	5/20	15/20	0/20	20/20
Middle	17/20	16/20	5/20	1/20	-	7/20	0/20	16/20
Tail	15/20	20/20	12/20	-	4/20	8/20	0/20	20/20
***Smed-trr-1* (RNAi)**								
Head	4/20	-	-	1/20	16/20	0/20	0/20	12/20
Middle	10/20	13/20	1/20	1/20	-	3/20	0/20	15/20
Tail	10/20	17/20	7/20	-	12/20	1/20	0/20	11/20
***Smed-trr-2* (RNAi)**								
Head	0/20	-	-	0/20	20/20	0/20	1/20	0/20
Middle	0/20	6/20	0/20	0/20	-	0/20	3/20	5/20

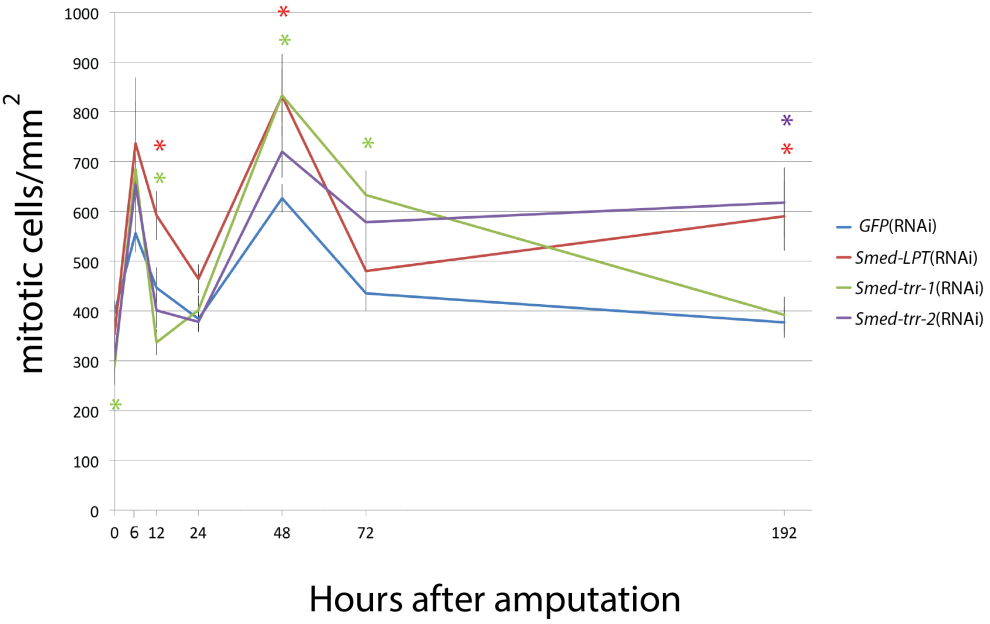
Tail	0/20	5/20	3/20	-	20/20	0/20	0/20	8/20

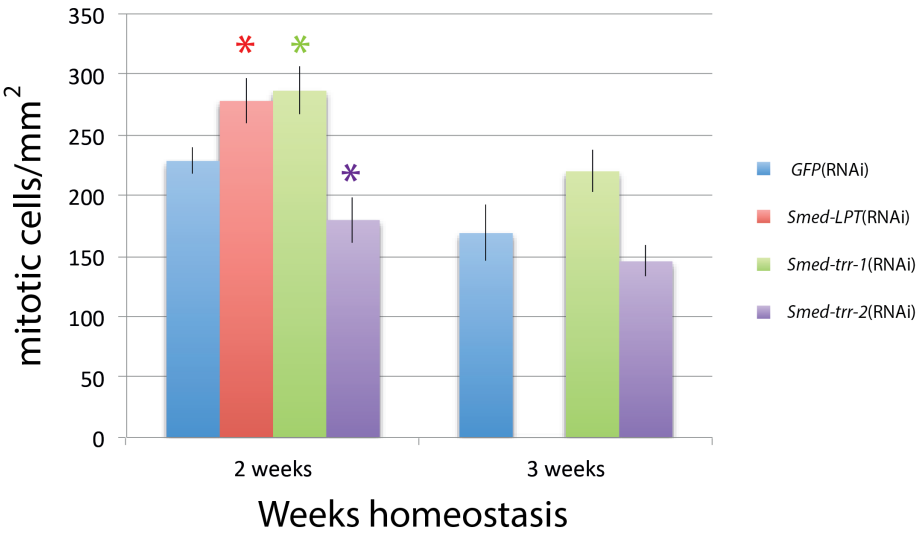
Homeostasis

	Head regression	**Pharynx regression**	**Movement defects**	**Outgrowths**	**Bloated posterior**
***Smed-LPT* (RNAi)**	4/20	13/20	20/20	2/20	4/20
***Smed-trr-1* (RNAi)**	0/20	0/20	8/20	0/20	3/20
***Smed-trr-2* (RNAi)**	0/20	0/20	3/20	0/20	3/20

8.9. APPENDIX I

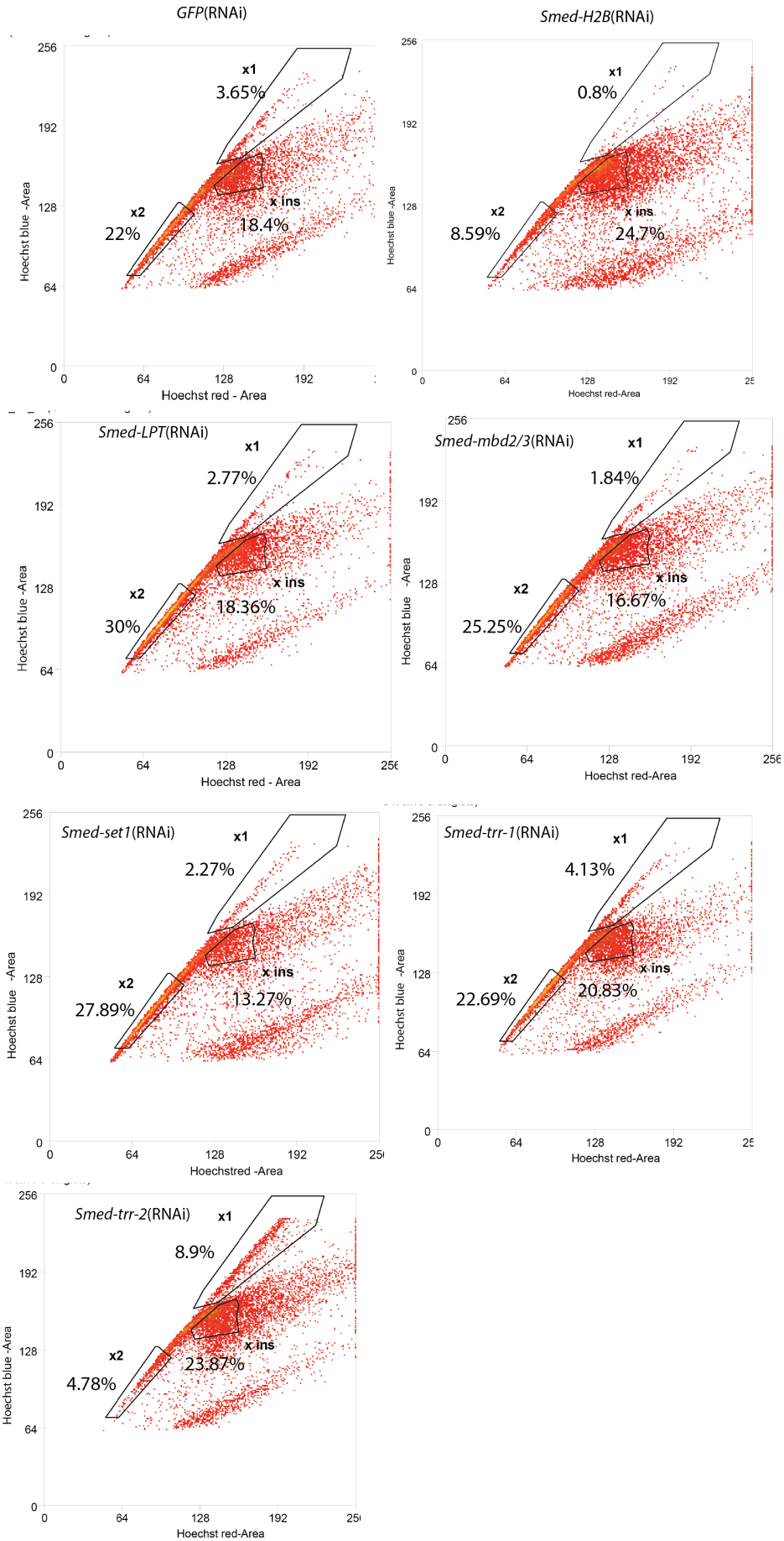
Number of mitotic cells following knockdown of trithorax-related genes. Mitotic cells are labeled with anti-H3P antibody. Regeneration and homeostatis phenotypes are shown.


mitotic cells/mm2
1000
900
800
700
600
500
400
300
200
100
0
0 6 12 24 48 72 192
Hours after amputation
GFP(RNAi)
Smed-LPT(RNAi)
Smed-trr-1(RNAi)
Smed-trr-2(RNAi)


mitotic cells/mm2
350
300
250
200
150
100
50
0
2 weeks
3 weeks
Weeks homeostasis
GFP(RNAi)
Smed-LPT(RNAi)
Smed-trr-1(RNAi)
Smed-trr-2(RNAi)

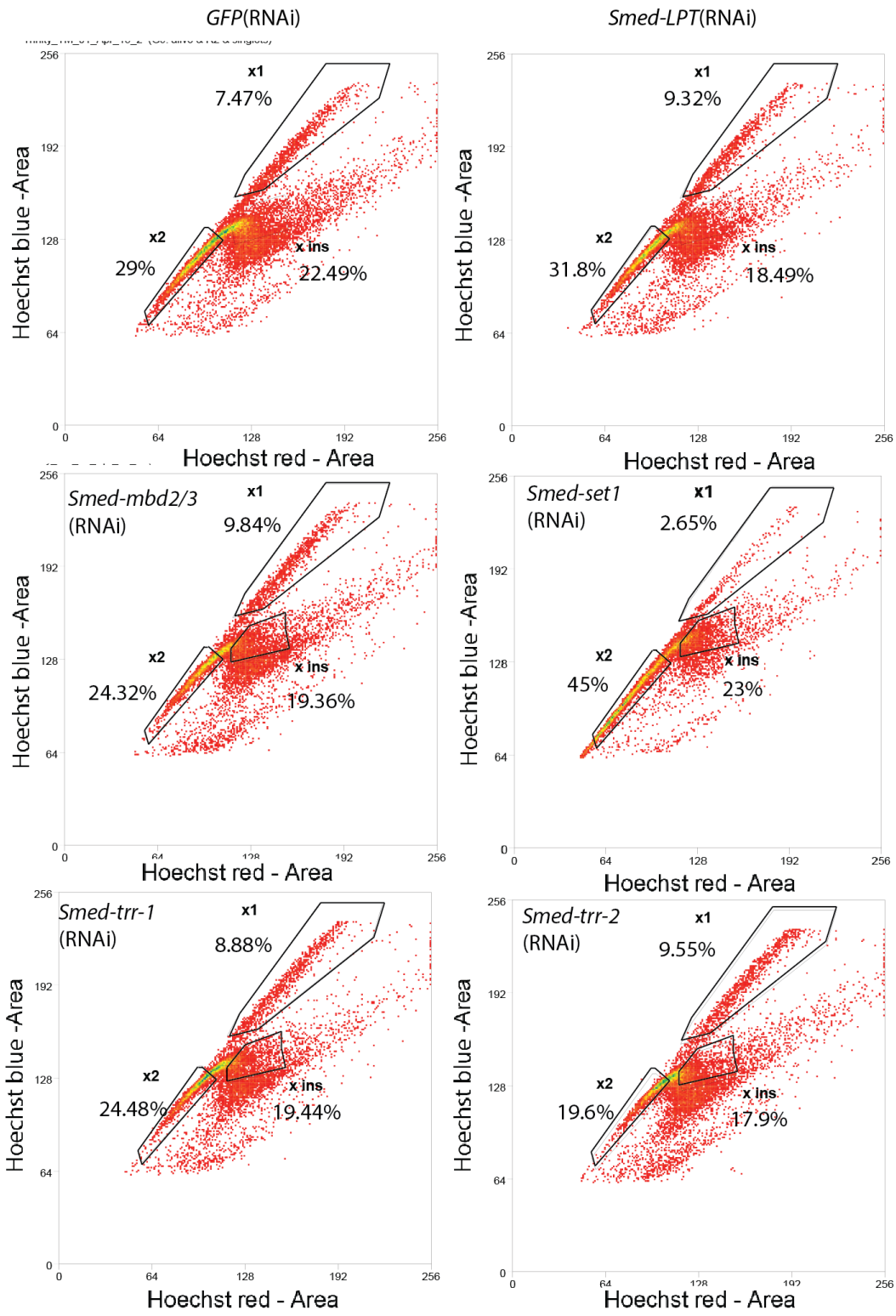
8.10. APPENDIX J

FACS graphs corresponding to each knockdown under the 5 days homeostasis condition. *Smed-H2B(*RNAi) was used as a positive control in the assay. The x axis shows Hoechst 33342 detection in the red channel (cytoplasmic complexity), while the y axis shows Hoechst 33342 detection in the blue channel (nuclear size).


GFP(RNAi)
Smed-H2B(RNAi)
Smed-LPT(RNAi)
Smed-mbd2/3(RNAi)
Smed-set1(RNAi)
Smed-trr-1(RNAi)
Smed-trr-2(RNAi)
x1
x2
x ins
3.65%
22%
18.4%
0.8%
8.59%
24.7%
2.77%
30%
18.36%
1.84%
25.25%
16.67%
2.27%
27.89%
13.27%
4.13%
22.69%
20.83%
8.9%
4.78%
23.87%
Hoechst blue -Area
Hoechst red - Area
Hoechst red-Area
Hoechstred -Area
0
64
128
192
256

8.11. APPENDIX K

This appendix shows the FACS graphs corresponding to each condition under the 3 days regeneration condition.



9. BIBLIOGRAPHY

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. *18*, 1923–1938.

Aasland, R., Gibson, T.J., and Stewart, A.F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. *20*, 56–59.

Abnave, P., Mottola, G., Gimenez, G., Boucherit, N., Trouplin, V., Torre, C., Conti, F., Ben Amara, A., Lepolard, C., Djian, B., et al. (2014). Screening in Planarians Identifies MORN2 as a Key Component in LC3-Associated Phagocytosis and Resistance to Bacterial Infection. Cell Host & Microbe *16*, 338–350.

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

Adler, C.E., Seidel, C.W., McKinney, S.A., and Sánchez Alvarado, A. (2014). Selective amputation of the pharynx identifies a FoxA-dependent regeneration program in planaria. eLife *3*.

Agata, K., Soejima, Y., Kato, K., Kobayashi, C., Umesono, Y., and Watanabe, K. (1998). Structure of the Planarian Central Nervous System (CNS) Revealed by Neuronal Cell Markers. Zoological Science *15*, 433–440.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell *9*, 1191–1200.

Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B., and Issa, J.P. (1998). Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res. *58*, 5489–5494.

Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., et al. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. Science *336*, 736–739.

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Françoijs, K.-J., Stunnenberg, H.G., and Veenstra, G.J.C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev. Cell *17*, 425–434.

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. *29*, 585–590.

Allis, C.D. (2007). Epigenetics (CSHL Press).

Andersen, E.C., and Horvitz, H.R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval. Development *134*, 2991–2999.

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183–197.

Annunziato, A.T., and Seale, R.L. (1983). Histone deacetylation is required for the maturation of newly replicated chromatin. J. Biol. Chem. *258*, 12675–12684.

Ansari, K.I., Shrestha, B., Hussain, I., Kasiri, S., and Mandal, S.S. (2011). Histone methylase MLL1 and MLL3 coordinate with estrogen receptors in estrogen-mediated HOXB9 expression. Biochemistry *50*, 3517–3527.

Ansari, K.I., Hussain, I., Kasiri, S., and Mandal, S.S. (2012). HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4. J Mol Endocrinol *48*, 61–75.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. *30*, 2817–2828.

Aryee, M.J., Liu, W., Engelmann, J.C., Nuhn, P., Gurel, M., Haffner, M.C., Esopi, D., Irizarry, R.A., Getzenberg, R.H., Nelson, W.G., et al. (2013). DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. Sci Transl Med *5*, 169ra10.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532–538.

Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., and Mills, A.A. (2007). CHD5 is a tumor suppressor at human 1p36. Cell *128*, 459–475.

Baguna, J., Saló, E., and Auladell, C. (1989). Regeneration and pattern formation in planarians. III. that neoblasts are totipotent stem cells and the cells. Development *107*, 77–86.

Baguñà, J., Romero, R., Saló, E., Collet, J., Auladell, C., Ribas, M., Riutort, M., García-Fernàndez, J., Burgaya, F., and Bueno, D. (1990). Growth, Degrowth and Regeneration as Developmental Phenomena in Adult Freshwater Planarians. In Experimental Embryology in Aquatic Plants and Animals, H.-J. Marthy, ed. (Springer US), pp. 129–162.

Balakrishnan, A., Bleeker, F.E., Lamba, S., Rodolfo, M., Daniotti, M., Scarpa, A., van Tilborg, A.A., Leenstra, S., Zanon, C., and Bardelli, A. (2007). Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. Cancer Res. *67*, 3545–3550.

Ballabeni, A., Park, I.-H., Zhao, R., Wang, W., Lerou, P.H., Daley, G.Q., and Kirschner, M.W. (2011). Cell cycle adaptations of embryonic stem cells. PNAS *108*, 19252–19257.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bapat, S.A., Jin, V., Berry, N., Balch, C., Sharma, N., Kurrey, N., Zhang, S., Fang, F., Lan, X., Li, M., et al. (2010). Multivalent epigenetic marks confer microenvironment-responsive epigenetic plasticity to ovarian cancer cells. Epigenetics *5*, 716–729.

Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R., Dynlacht, B.D., and Simon, I. (2008). Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. PNAS *105*, 955–960.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Barth, T.K., and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. Trends Biochem. Sci. *35*, 618–626.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome — biological and translational implications. Nat Rev Cancer *11*, 726–734.

Belinsky, S.A., Nikula, K.J., Palmisano, W.A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S.B., and Herman, J.G. (1998). Aberrant methylation of p16INK4a is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A *95*, 11891–11896.

Berdasco, M., and Esteller, M. (2010). Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. Developmental Cell *19*, 698–711.

Berdasco, M., Ropero, S., Setien, F., Fraga, M.F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N.-K., Rahman, N., and Esteller, M. (2009). Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proc. Natl. Acad. Sci. U.S.A. *106*, 21830–21835.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., et al. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat Genet *44*, 40–46.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Best, J.B., and Morita, M. (1982). Planarians as a model system for in vitro teratogenesis studies. Teratog. Carcinog. Mutagen. *2*, 277–291.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. Trends in Biochemical Sciences *31*, 35–40.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bird, A.P. (1984). DNA methylation versus gene expression. J Embryol Exp Morphol *83 Suppl*, 31–40.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Blanpain, C., Horsley, V., and Fuchs, E. (2007). Epithelial stem cells: turning over new leaves. Cell *128*, 445–458.

Bode, H., Dunne, J., Heimfeld, S., Huang, L., Javois, L., Koizumi, O., Westerfield, J., and Yaross, M. (1986). Transdifferentiation occurs continuously in adult hydra. Curr. Top. Dev. Biol. *20*, 257–280.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med *3*, 730–737.

Bonuccelli, L., Rossi, L., Lena, A., Scarcelli, V., Rainaldi, G., Evangelista, M., Iacopetti, P., Gremigni, V., and Salvetti, A. (2010). An RbAp48-like gene regulates adult stem cells in planarians. J Cell Sci *123*, 690–698.

Borodovsky, A., Salmasi, V., Turcan, S., Fabius, A.W.M., Baia, G.S., Eberhart, C.G., Weingart, J.D., Gallia, G.L., Baylin, S.B., Chan, T.A., et al. (2013). 5-azacytidine reduces methylation, promotes differentiation and induces tumor regression in a patient-derived IDH1 mutant glioma xenograft. Oncotarget *4*, 1737–1747.

Brockes, J.P., and Kumar, A. (2008). Comparative Aspects of Animal Regeneration. Annual Review of Cell and Developmental Biology *24*, 525–549.

Brøndsted, H.V. (1969). Planarian regeneration, (Oxford; New York: Pergamon Press).

Brun, M.-E., Gasca, S., Girard, C., Bouton, K., De Massy, B., and De Sario, A. (2006). Characterization and expression analysis during embryo development of the mouse ortholog of MLL3. Gene *371*, 25–33.

Capo-chichi, C.D., Cai, K.Q., Simpkins, F., Ganjei-Azar, P., Godwin, A.K., and Xu, X.-X. (2011). Nuclear envelope structural defects cause chromosomal numerical instability and aneuploidy in ovarian cancer. BMC Med *9*, 28.

Cebrià, F., and Newmark, P.A. (2007). Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. Development *134*, 833–837.

Cebrià, F., and Newmark, P.A. (2005). Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. Development *132*, 3691–3703.

Chaffer, C.L., and Weinberg, R.A. (2015). How does multistep tumorigenesis really proceed? Cancer Discov *5*, 22–24.

Chandebois, R. (1976). Histogenesis and morphogenesis in planarian regeneration. Monogr Dev Biol *11*, 1–182.

Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., et al. (2010). The dynamic genome of Hydra. Nature *464*, 592–596.

Chauhan, C., Zraly, C.B., Parilla, M., Diaz, M.O., and Dingwall, A.K. (2012). Histone recognition and nuclear receptor co-activator functions of Drosophila Cara Mitad, a homolog of the N-terminal portion of mammalian MLL2 and MLL3. Development *139*, 1997–2008.

Chauhan, C., Zraly, C.B., and Dingwall, A.K. (2013). The Drosophila COMPASS-like Cmi-Trr coactivator complex regulates dpp/BMP signaling in pattern formation. Dev. Biol. *380*, 185–198.

Chen, C., Liu, Y., Rappaport, A.R., Kitzing, T., Schultz, N., Zhao, Z., Shroff, A.S., Dickins, R.A., Vakoc, C.R., Bradner, J.E., et al. (2014). MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. Cancer Cell *25*, 652–665.

Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S.C., Falck, J.R., Peng, J., Gu, W., and Zhao, Y. (2007). Lysine propionylation and butyrylation are novel post-translational modifications in histones. Mol. Cell Proteomics *6*, 812–819.

Cheng, J., Blum, R., Bowman, C., Hu, D., Shilatifard, A., Shen, S., and Dynlacht, B.D. (2014). A role for H3K4 mono-methylation in gene repression and partitioning of chromatin readers. Mol Cell *53*, 979–992.

Chestier, A., and Yaniv, M. (1979). Rapid turnover of acetyl groups in the four core histones of simian virus 40 minichromosomes. Proc Natl Acad Sci U S A *76*, 46–50.

Chinwalla, V., Jane, E.P., and Harte, P.J. (1995). The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. EMBO J *14*, 2056–2065.

Cho, R.W., Wang, X., Diehn, M., Shedden, K., Chen, G.Y., Sherlock, G., Gurney, A., Lewicki, J., and Clarke, M.F. (2008). Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. Stem Cells *26*, 364–371.

Chu, P., Clanton, D.J., Snipas, T.S., Lee, J., Mitchell, E., Nguyen, M.-L., Hare, E., and

Peach, R.J. (2009). Characterization of a subpopulation of colon cancer cells with stem cell-like properties. Int. J. Cancer *124*, 1312–1321.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., et al. (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353–360.

Clarke, M.F. (2005). A self-renewal assay for cancer stem cells. Cancer Chemother. Pharmacol. *56 Suppl 1*, 64–68.

Clevers, H. (2011). The cancer stem cell: premises, promises and challenges. Nat Med 313–319.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., et al. (2011). The ecoresponsive genome of Daphnia pulex. Science *331*, 555–561.

Collins, A.T., Berry, P.A., Hyde, C., Stower, M.J., and Maitland, N.J. (2005a). Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res. *65*, 10946–10951.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005b). In Vitro and in Vivo Analyses of a Phe/Tyr Switch Controlling Product Specificity of Histone Lysine Methyltransferases. J. Biol. Chem. *280*, 5563–5570.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R., and Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat. Struct. Mol. Biol. *15*, 245–250.

Consortium, T.U. (2015). UniProt: a hub for protein information. Nucl. Acids Res. *43*, D204–D212.

Corpet, A., and Almouzni, G. (2009). Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. Trends Cell Biol. *19*, 29–41.

Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat. Genet. *24*, 132–138.

Cougot, N., Babajko, S., and Séraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. J. Cell Biol. *165*, 31–40.

Coward, S.J. (1974). Chromatoid bodies in somatic cells of the planarian: observations on their behavior during mitosis. Anat. Rec. *180*, 533–545.

Cowles, M.W., Brown, D.D.R., Nisperos, S.V., Stanley, B.N., Pearson, B.J., and Zayas, R.M. (2013). Genome-wide analysis of the bHLH gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. Development

140, 4691–4702.

Craig, J.M. (2005). Heterochromatin--many flavours, common themes. Bioessays *27*, 17–28.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U.S.A. *107*, 21931–21936.

Crook, J.M., Dunn, N.R., and Colman, A. (2006). Repressed by a NuRD. Nat Cell Biol *8*, 212–214.

Currie, K.W., and Pearson, B.J. (2013). Transcription factors lhx1/5-1 and pitx are required for the maintenance and regeneration of serotonergic neurons in planarians. Development *140*, 3577–3588.

Dearolf, C.R. (1998). Fruit fly "leukemia." Biochim. Biophys. Acta *1377*, M13–M23.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biology *6*, 227.

Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc. Natl. Acad. Sci. U.S.A. *102*, 5501–5506.

Dolle, P., Ruberte, E., Leroy, P., Morriss-Kay, G., and Chambon, P. (1990). Retinoic acid receptors and cellular retinoid binding proteins. I. A systematic study of their differential pattern of transcription during mouse organogenesis. Development *110*, 1133–1151.

Du, Y., Lu, F., Li, P., Ye, J., Ji, M., Ma, D., and Ji, C. (2014). SMG1 Acts as a Novel Potential Tumor Suppressor with Epigenetic Inactivation in Acute Myeloid Leukemia. International Journal of Molecular Sciences *15*, 17065–17076.

Dubois, F. (1949). Contribution a l'étude de la migration des cellules de régénération chez les Planaires Dulcicoles.

Dutnall, R.N., and Ramakrishnan, V. (1997). Twists and turns of the nucleosome: tails without ends. Structure *5*, 1255–1259.

Easwaran, H., Johnstone, S.E., Van Neste, L., Ohm, J., Mosbruger, T., Wang, Q., Aryee, M.J., Joyce, P., Ahuja, N., Weisenberger, D., et al. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. Genome Res. *22*, 837–849.

Easwaran, H., Tsai, H.-C., and Baylin, S.B. (2014). Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. Molecular Cell *54*,

716–727.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet *38*, 1378–1385.

Eisenhoffer, G.T., Kang, H., and Alvarado, A.S. (2008). Molecular Analysis of Stem Cells and Their Descendants during Cell Turnover and Regeneration in the Planarian Schmidtea mediterranea. Cell Stem Cell *3*, 327–339.

Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev. Biol. *339*, 240–249.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A *87*, 9923–9927.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C., et al. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature *486*, 353–360.

English, C.M., Maluf, N.K., Tripet, B., Churchill, M.E.A., and Tyler, J.K. (2005). ASF1 binds to a heterodimer of histones H3 and H4: a two-step mechanism for the assembly of the H3-H4 heterotetramer on DNA. Biochemistry *44*, 13673–13682.

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. *17*, 1086–1093.

Ermakova, O.N., Ermakov, A.M., Tiras, K.P., and Lednev, V.V. (2009). Retinoic acid as a regulator of planarian morphogenesis. Russ J Dev Biol *40*, 367–372.

Escriva, H., Bertrand, S., and Laudet, V. (2004). The evolution of the nuclear receptor superfamily. Essays Biochem. *40*, 11–26.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Molecular Cell *38*, 452–464.

Esteller, M., Sparks, A., Toyota, M., Sanchez-Cespedes, M., Capella, G., Peinado, M.A., Gonzalez, S., Tarafa, G., Sidransky, D., Meltzer, S.J., et al. (2000). Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. Cancer Res. *60*, 4366–4371.

Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G. (2001). A gene hypermethylation profile of human cancer. Cancer Res. *61*, 3225–3229.

Estève, P.-O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev. *20*, 3089–3103.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution *39*, 783–791.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. Current Biology *12*, 1052–1058.

Fernandéz-Taboada, E., Moritz, S., Zeuschner, D., Stehling, M., Schöler, H.R., Saló, E., and Gentile, L. (2010). Smed-SmB, a member of the LSm protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation. Development *137*, 1055–1065.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucl. Acids Res. *42*, D222–D230.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fisher, K., Southall, S.M., Wilson, J.R., and Poulin, G.B. (2010). Methylation and demethylation activities of a C. elegans MLL-like complex attenuate RAS signalling. Developmental Biology *341*, 142–153.

Forbes, A., and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. Development *125*, 679–690.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., et al. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucl. Acids Res. *43*, D805–D811.

Foster, J.A. (1963). Induction of Neoplasms in Planarians with Carcinogens. Cancer Res *23*, 300–303.

Foster, J.A. (1969). Malformations and lethal growths in planaria treated with carcinogens. Natl Cancer Inst Monogr *31*, 683–691.

Fraguas, S., Barberán, S., and Cebrià, F. (2011). EGFR signaling regulates cell proliferation, differentiation and morphogenesis during planarian regeneration and homeostasis. Developmental Biology *354*, 87–101.

Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., Dimeco, F., and Vescovi, A. (2004). Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. Cancer Res. *64*, 7011–7021.

Garcia-Corrales, P., and Gamo, J. (1986). The Ultrastructure of the Gastrodermal Gland Cells in the Freshwater Planarian Dugesia gonocephala s.l. Acta Zoologica *67*, 43–51.

Gardner, K.E., Allis, C.D., and Strahl, B.D. (2011). OPERating ON chromatin, a colorful language where context matters. J Mol Biol *409*, 36–46.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. New England Journal of Medicine *366*, 883–892.

Geyer, K.K., Rodríguez López, C.M., Chalmers, I.W., Munshi, S.E., Truscott, M., Heald, J., Wilkinson, M.J., and Hoffmann, K.F. (2011). Cytosine methylation regulates oviposition in the pathogenic blood fluke Schistosoma mansoni. Nat Commun *2*, 424.

Geyer, K.K., Chalmers, I.W., MacKintosh, N., Hirst, J.E., Geoghegan, R., Badets, M., Brophy, P.M., Brehm, K., and Hoffmann, K.F. (2013). Cytosine methylation is a conserved epigenetic feature found throughout the phylum Platyhelminthes. BMC Genomics *14*, 462.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395–398.

González-Estévez, C., Felix, D.A., Smith, M.D., Paps, J., Morley, S.J., James, V., Sharp, T.V., and Aboobaker, A.A. (2012). SMG-1 and mTORC1 Act Antagonistically to Regulate Response to Injury and Growth in Planarians. PLoS Genet *8*, e1002619.

Goo, Y.-H., Sohn, Y.C., Kim, D.-H., Kim, S.-W., Kang, M.-J., Jung, D.-J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T., et al. (2003). Activating Signal Cointegrator 2 Belongs to a Novel Steady-State Complex That Contains a Subset of Trithorax Group Proteins. Mol. Cell. Biol. *23*, 140–149.

Goss, H.M. (2014). Histone modifications across the cell cycle in undifferentiated and differentiating mouse embryonic stem cells. d_ph. University of Birmingham.

Greer, E.L., Maures, T.J., Hauswirth, A.G., Green, E.M., Leeman, D.S., Maro, G.S., Han, S., Banko, M.R., Gozani, O., and Brunet, A. (2010). Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. Nature *466*, 383–387.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in C. elegans. Cell *161*, 868–878.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35–46.

Grewal, S.I.S., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798–802.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., et al. (2011). The Genome Portal of the Department of Energy Joint Genome Institute. Nucl. Acids Res. gkr947.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol. Cell. Biol. *24*, 7483–7490.

Grimwade, D., Hills, R.K., Moorman, A.V., Walker, H., Chatters, S., Goldstone, A.H., Wheatley, K., Harrison, C.J., Burnett, A.K., and National Cancer Research Institute Adult Leukaemia Working Group (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood *116*, 354–365.

Groth, A., Corpet, A., Cook, A.J.L., Roche, D., Bartek, J., Lukas, J., and Almouzni, G. (2007). Regulation of replication fork progression through histone supply and demand. Science *318*, 1928–1931.

Guedelhoefer, O.C., and Alvarado, A.S. (2012). Amputation induces stem cell mobilization to sites of injury during planarian regeneration. Development *139*, 3510–3520.

Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li, X., Zhou, L., et al. (2011). Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet *43*, 875–878.

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

Gurley, K.A., Rink, J.C., and Alvarado, A.S. (2008). β-Catenin Defines Head Versus Tail Identity During Planarian Regeneration and Homeostasis. Science *319*, 323–327.

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

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381–385.

Han, L.-L., Nan, H.-C., Tian, T., Guo, H., Hu, T.-H., Wang, W.-J., Ma, J.-Q., Jiang, L.-L., Guo, Q.-Q., Yang, C.-C., et al. (2014). Expression and significance of the novel tumor-suppressor gene SMG-1 in hepatocellular carcinoma. Oncol. Rep. *31*, 2569–2578.

Hanahan, D., and Coussens, L.M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell *21*, 309–322.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. *43*, 768–775.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291–1300.

Harris, T.W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., Done, J., Grove, C., Howe, K., et al. (2014). WormBase 2014: new views of curated biology. Nucleic Acids Res. *42*, D789–D793.

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

Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: A public knowledge-base of divergence times among organisms. Bioinformatics *22*, 2971-2972.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. (2015). Tree of Life Reveals Clock-Like Speciation and Diversification. Mol Biol Evol. *32*, 835-845.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heitz, E. (1928). Das heterochromatin der moose. I Jahrb wiss Bot 69: 762–818. Find This Article Online.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet. *27*, 389–396.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends in Genetics *20*, 320–326.

Herz, H.-M., Madden, L.D., Chen, Z., Bolduc, C., Buff, E., Gupta, R., Davuluri, R.,

Shilatifard, A., Hariharan, I.K., and Bergmann, A. (2010). The H3K27me3 demethylase dUTX is a suppressor of Notch- and Rb-dependent tumors in Drosophila. Mol. Cell. Biol. *30*, 2485–2497.

Herz, H.-M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.-H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. *26*, 2604–2620.

Herz, H.-M., Hu, D., and Shilatifard, A. (2014). Enhancer malfunction in cancer. Mol. Cell *53*, 859–866.

Hewitson, T.D., Kelynack, K.J., and Darby, I.A. (2006). Histochemical localization of cell proliferation using in situ hybridization for histone mRNA. Methods Mol. Biol. *326*, 219–226.

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

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.

Hori, I. (1997). Cytological approach to morphogenesis in the planarian blastema. II. The effect of neuropeptides. J. Submicrosc. Cytol. Pathol. *29*, 91–97.

Hsieh, J.J.-D., Cheng, E.H.-Y., and Korsmeyer, S.J. (2003). Taspase1. Cell *115*, 293–303.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013a). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat Struct Mol Biol *20*, 1093–1097.

Hu, D., Gao, X., Morgan, M.A., Herz, H.-M., Smith, E.R., and Shilatifard, A. (2013b). The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. Mol Cell Biol *33*, 4745–4754.

Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M.T., Jenuwein, T., et al. (2007). p53 is regulated by the lysine demethylase LSD1. Nature *449*, 105–108.

Hubert, A., Henderson, J.M., Ross, K.G., Cowles, M.W., Torres, J., and Zayas, R.M. (2013). Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone methyltransferases. Epigenetics *8*, 79–91.

Huen, M.S.Y., Sy, S.M.-H., van Deursen, J.M., and Chen, J. (2008). Direct Interaction between SET8 and Proliferating Cell Nuclear Antigen Couples H4-K20 Methylation with DNA Replication. J Biol Chem *283*, 11073–11077.

Ishizu, H., Siomi, H., and Siomi, M.C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. *26*, 2361–2373.

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

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and Function of a Human TAFII250 Double Bromodomain Module. Science *288*, 1422–1425.

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

James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol. Cell. Biol. *6*, 3862–3872.

Jamieson, C.H.M., Ailles, L.E., Dylla, S.J., Muijtjens, M., Jones, C., Zehnder, J.L., Gotlib, J., Li, K., Manz, M.G., Keating, A., et al. (2004). Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. N. Engl. J. Med. *351*, 657–667.

Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem *3*, 274–293.

Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63–66.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jerez, A., Sugimoto, Y., Makishima, H., Verma, A., Jankowska, A.M., Przychodzen, B., Visconte, V., Tiu, R.V., O'Keefe, C.L., Mohamedali, A.M., et al. (2012). Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. Blood *119*, 6109–6117.

Jeschke, J., Collignon, E., and Fuks, F. (2016). Portraits of TET-mediated DNA hydroxymethylation in cancer. Curr. Opin. Genet. Dev. *36*, 16–26.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519–1529.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683–692.

Jones D.T., Taylor W.R., and Thornton J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences *8*,

275-282.

Jones, R.S., and Gelbart, W.M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol. Cell. Biol. *13*, 6357–6366.

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., and Izpisúa Belmonte, J.C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. Nature *464*, 606–609.

Juliano, C.E., Swartz, S.Z., and Wessel, G.M. (2010). A conserved germline multipotency program. Development *137*, 4113–4126.

Kanda, H., Nguyen, A., Chen, L., Okano, H., and Hariharan, I.K. (2013). The Drosophila Ortholog of MLL3 and MLL4, trithorax related, Functions as a Negative Regulator of Tissue Growth. Mol. Cell. Biol. *33*, 1702–1710.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333–339.

Kao, D., Felix, D., and Aboobaker, A. (2013). The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. BMC Genomics *14*, 797.

Kastan, M.B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316–323.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol *30*, 772–780.

Kelly, P.N., Dakic, A., Adams, J.M., Nutt, S.L., and Strasser, A. (2007). Tumor Growth Need Not Be Driven by Rare Cancer Stem Cells. Science *317*, 337–337.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Kido, T., Kishida, Y., and Asakura, H. (1964). Observation on the histological configuration of the pharynx in the planarian, Dugesia gonocephala. Ann Rep Noto Mar Lab *4*, 121–127.

Kikuchi, K., Holdway, J.E., Werdich, A.A., Anderson, R.M., Fang, Y., Egnaczyk, G.F., Evans, T., MacRae, C.A., Stainier, D.Y.R., and Poss, K.D. (2010). Primary contribution to zebrafish heart regeneration by gata4+ cardiomyocytes. Nature *464*, 601–605.

Killmann, S.A., Cronkite, E.P., Robertson, J.S., Fliedner, T.M., and Bond, V.P. (1963). Estimation of phases of the life cycle of leukemic cells from labeling in human beings in vivo with tritiated thymidine. Lab. Invest. *12*, 671–684.

Kim, D.-H., Lee, J., Lee, B., and Lee, J.W. (2009). ASCOM controls farnesoid X receptor transactivation through its associated histone H3 lysine 4 methyltransferase activity. Mol. Endocrinol. *23*, 1556–1562.

Kim, D.-H., Kim, J., and Lee, J.W. (2011). Requirement for MLL3 in p53 regulation of hepatic expression of small heterodimer partner and bile acid homeostasis. Mol. Endocrinol. *25*, 2076–2083.

Kim, D.-H., Rhee, J.C., Yeo, S., Shen, R., Lee, S.-K., Lee, J.W., and Lee, S. (2015). Crucial Roles of Mixed-Lineage Leukemia 3 and 4 as Epigenetic Switches of the Hepatic Circadian Clock Controlling Bile Acid Homeostasis in Mice. Hepatology *61*, 1012–1023.

Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J. Cell Biol. *153*, 1341–1353.

King, R.S., and Newmark, P.A. (2013). In situ hybridization protocol for enhanced detection of gene expression in the planarian Schmidtea mediterranea. BMC Developmental Biology *13*, 8.

Knopf, F., Hammond, C., Chekuru, A., Kurth, T., Hans, S., Weber, C.W., Mahatma, G., Fisher, S., Brand, M., Schulte-Merker, S., et al. (2011). Bone regenerates via dedifferentiation of osteoblasts in the zebrafish fin. Dev. Cell *20*, 713–724.

Kobayashi, C., Watanabe, K., and Agata, K. (1999). The Process of Pharynx Regeneration in Planarians. Developmental Biology *211*, 27–38.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaöz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res *17*, 691–707.

Koelle, M.R., Talbot, W.S., Segraves, W.A., Bender, M.T., Cherbas, P., and Hogness, D.S. (1991). The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell *67*, 59–77.

Kornberg, R.D., and Thonmas, J.O. (1974). Chromatin Structure: Oligomers of the Histones. Science *184*, 865–868.

Kotaja, N., and Sassone-Corsi, P. (2007). The chromatoid body: a germ-cell-specific RNA-processing centre. Nat Rev Mol Cell Biol *8*, 85–90.

Kotaja, N., Bhattacharyya, S.N., Jaskiewicz, L., Kimmins, S., Parvinen, M., Filipowicz, W., and Sassone-Corsi, P. (2006). The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. Proc. Natl. Acad. Sci. U.S.A. *103*, 2647–2652.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Kragl, M., Knapp, D., Nacu, E., Khattak, S., Maden, M., Epperlein, H.H., and Tanaka, E.M. (2009). Cells keep a memory of their tissue origin during axolotl limb regeneration. Nature *460*, 60–65.

Kreso, A., and Dick, J.E. (2014). Evolution of the Cancer Stem Cell Model. Cell Stem Cell *14*, 275–291.

Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M., and Shilatifard, A. (2002). COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J. Biol. Chem. *277*, 10753–10755.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet *4*, e1000242.

Kumar S., Stecher G., and Tamura K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution (submitted).

Kummerfeld, S.K., and Teichmann, S.A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. Trends in Genetics *21*, 25–30.

Kunert, N., and Brehm, A. (2009). Novel Mi-2 related ATP-dependent chromatin remodelers. Epigenetics *4*, 209–211.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083–5090.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different Ezh2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3. Molecular Cell *14*, 183–193.

Labbé, R.M., Irimia, M., Currie, K.W., Lin, A., Zhu, S.J., Brown, D.D.R., Ross, E.J., Voisin, V., Bader, G.D., Blencowe, B.J., et al. (2012). A Comparative Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians and Mammals. STEM CELLS *30*, 1734–1745.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lagasse, E. (2008). Cancer stem cells with genetic instability: the best vehicle with the best engine for cancer. Gene Ther. *15*, 136–142.

Lambrus, B.G., Cochet-Escartin, O., Gao, J., Newmark, P.A., Collins, E.-M.S., and Collins, J.J. (2015). Tryptophan hydroxylase Is Required for Eye Melanogenesis in the Planarian Schmidtea mediterranea. PLOS ONE *10*, e0127074.

Lanzuolo, C., and Orlando, V. (2012). Memories from the Polycomb Group Proteins. Annual Review of Genetics *46*, 561–589.

Lapan, S.W., and Reddien, P.W. (2011). dlx and sp6-9 Control Optic Cup Regeneration in a Prototypic Eye. PLoS Genet *7*, e1002226.

Lapan, S.W., and Reddien, P.W. (2012). Transcriptome Analysis of the Planarian Eye Identifies ovo as a Specific Regulator of Eye Regeneration. Cell Reports *2*, 294–307.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M.A., and Dick, J.E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature *367*, 645–648.

Laube, F., Heister, M., Scholz, C., Borchardt, T., and Braun, T. (2006). Re-programming of newt cardiomyocytes is induced by tissue regeneration. J. Cell. Sci. *119*, 4719–4729.

Lee, H., Kim, J.H., Chae, Y.J., Ogawa, H., Lee, M.H., and Gerton, G.L. (1998). Creatine synthesis and transport systems in the male rat reproductive tract. Biol. Reprod. *58*, 1437–1444.

Lee, J., Saha, P.K., Yang, Q.-H., Lee, S., Park, J.Y., Suh, Y., Lee, S.-K., Chan, L., Roeder, R.G., and Lee, J.W. (2008a). Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. Proc. Natl. Acad. Sci. U.S.A. *105*, 19229–19234.

Lee, J., Kim, D.-H., Lee, S., Yang, Q.-H., Lee, D.K., Lee, S.-K., Roeder, R.G., and Lee, J.W. (2009). A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proc. Natl. Acad. Sci. U.S.A. *106*, 8513–8518.

Lee, J.-S., Smith, E., and Shilatifard, A. (2010). The Language of Histone Crosstalk. Cell *142*, 682–685.

Lee, S., Lee, D.-K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.-Y., Lee, S.-K., Roeder, R.G., and Lee, J.W. (2006a). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proc Natl Acad Sci U S A *103*, 15392–15397.

Lee, S., Lee, J., Lee, S.-K., and Lee, J.W. (2008b). Activating Signal Cointegrator-2 Is an Essential Adaptor to Recruit Histone H3 Lysine 4 Methyltransferases MLL3 and MLL4 to the Liver X Receptors. Molecular Endocrinology *22*, 1312–1319.

Lee, S.-K., Anzick, S.L., Choi, J.-E., Bubendorf, L., Guan, X.-Y., Jung, Y.-K., Kallioniemi, O.P., Kononen, J., Trent, J.M., Azorsa, D., et al. (1999). A Nuclear Factor, ASC-2, as a Cancer-amplified Transcriptional Coactivator Essential for Ligand-dependent Transactivation by Nuclear Receptors in Vivo. J. Biol. Chem. *274*, 34283–34293.

Lee, S.K., Jung, S.Y., Kim, Y.S., Na, S.Y., Lee, Y.C., and Lee, J.W. (2001). Two distinct nuclear receptor-interaction domains and CREB-binding protein-dependent transactivation function of activating signal cointegrator-2. Mol. Endocrinol. *15*, 241–254.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006b). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell *125*, 301–313.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p Modulates Silencing in Yeast by Methylation of the Nucleosome Core. Cell *109*, 745–756.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. Nucl. Acids Res. *43*, D257–D260.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91–95.

Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K., and Clarke, M.F. (2007). The Prognostic Role of a Gene Signature from Tumorigenic Breast-Cancer Cells. New England Journal of Medicine *356*, 217–226.

Liu, W., Tanasa, B., Tyurina, O.V., Zhou, T.Y., Gassmann, R., Liu, W.T., Ohgi, K.A., Benner, C., Garcia-Bassets, I., Aggarwal, A.K., et al. (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature *466*, 508–512.

Liu, Z., Li, F., Ruan, K., Zhang, J., Mei, Y., Wu, J., and Shi, Y. (2014). Structural and Functional Insights into the Human Börjeson-Forssman-Lehmann Syndrome-associated Protein PHF6. J Biol Chem *289*, 10069–10083.

Logan-Klumpler, F.J., Silva, N.D., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., et al. (2012). GeneDB—an annotation database for pathogens. Nucl. Acids Res. *40*, D98–D108.

Loyola, A., and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? Trends Biochem. Sci. *32*, 425–433.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol. Cell *24*, 309–316.

Luo, R.X., and Dean, D.C. (1999). Chromatin Remodeling and Transcriptional Regulation. JNCI J Natl Cancer Inst *91*, 1288–1294.

Luo, Y.-J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M.,

Yamasaki, S., Endo, K., and Satoh, N. (2015). The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nat Commun *6*, 8301.

Lynch, M.D., Smith, A.J.H., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. *31*, 317–329.

Mahgoub, N., Parker, R.I., Hosler, M.R., Close, P., Winick, N.J., Masterson, M., Shannon, K.M., and Felix, C.A. (1998). RAS mutations in pediatric leukemias with MLL gene rearrangements. Genes Chromosomes Cancer *21*, 270–275.

Malanchi, I., Peinado, H., Kassen, D., Hussenet, T., Metzger, D., Chambon, P., Huber, M., Hohl, D., Cano, A., Birchmeier, W., et al. (2008). Cutaneous cancer stem cell maintenance is dependent on beta-catenin signalling. Nature *452*, 650–653.

Mann, K.M., Ward, J.M., Yew, C.C.K., Kovochich, A., Dawson, D.W., Black, M.A., Brett, B.T., Sheetz, T.E., Dupuy, A.J., Initiative, A.P.C.G., et al. (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. PNAS *109*, 5934–5941.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.

Marcotte, E.M., Pellegrini, M., Ng, H.-L., Rice, D.W., Yeates, T.O., and Eisenberg, D. (1999). Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. Science *285*, 751–753.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol. Cell *32*, 503–518.

März, M., Seebeck, F., and Bartscherer, K. (2013). A Pitx transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. Development *140*, 4499–4509.

Mason, P.B., and Struhl, K. (2003). The FACT Complex Travels with Elongating RNA Polymerase II and Is Important for the Fidelity of Transcriptional Initiation In Vivo. Mol. Cell. Biol. *23*, 8323–8333.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. Int. J. Biochem. Cell Biol. *41*, 108–116.

Medyouf, H., Mossner, M., Jann, J.-C., Nolte, F., Raffel, S., Herrmann, C., Lier, A., Eisen, C., Nowak, V., Zens, B., et al. (2014). Myelodysplastic cells in patients reprogram mesenchymal stromal cells to establish a transplantable stem cell niche disease unit. Cell Stem Cell *14*, 824–837.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Merlos-Suárez, A., Barriga, F.M., Jung, P., Iglesias, M., Céspedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Muñoz, P., et al. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell *8*, 511–524.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*, 105–116.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., et al. (2009). New insights to the MLL recombinome of acute leukemias. Leukemia *23*, 1490–1499.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U.S.A. *98*, 12902–12907.

Milutinovic, S., Zhuang, Q., and Szyf, M. (2002). Proliferating cell nuclear antigen associates with histone deacetylase activity, integrating DNA replication and chromatin modification. J. Biol. Chem. *277*, 20974–20978.

Minc, E., Courvalin, J.C., and Buendia, B. (2000). HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. Cytogenet. Cell Genet. *90*, 279–284.

Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010a). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. *24*, 574–589.

Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010b). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat Rev Cancer *10*, 721–728.

Mohan, M., Herz, H.-M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., and Shilatifard, A. (2011). The COMPASS Family of H3K4 Methylases in Drosophila. Mol. Cell. Biol. *31*, 4310–4318.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393–396.

Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the Maestro of the Replication Fork. Cell *129*, 665–679.

Molina, M.D., Saló, E., and Cebrià, F. (2007). The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. Dev. Biol. *311*, 79–94.

Morgan, T.H. (1901). Regeneration. Columbia University Biological Series, vol. 3 (New York: The Macmillan Company).

Morita, M., Best, J.B., and Noel, J. (1969). Electron microscopic studies of planarian regeneration. I. Fine structure of neoblasts in Dugesia dorotocephala. J. Ultrastruct. Res. *27*, 7–23.

Muller, H.J., and Altenburg, E. (1930). The Frequency of Translocations Produced by X-Rays in Drosophila. Genetics *15*, 283–311.

Murakami, Y. (2013). Heterochromatin and Euchromatin. In Encyclopedia of Systems Biology, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, eds. (Springer New York), pp. 881–884.

Nelson, C.M., and Bissell, M.J. (2006). Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. Annu. Rev. Cell Dev. Biol. *22*, 287–309.

Newmark, P.A., and Sánchez Alvarado, A. (2000). Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. Dev. Biol. *220*, 142–153.

Newmark, P.A., Wang, Y., and Chong, T. (2008). Germ cell specification and regeneration in planarians. Cold Spring Harb. Symp. Quant. Biol. *73*, 573–581.

Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P., and Losson, R. (2001). Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. Mol. Cell *7*, 729–739.

Nishimura, K., Kitamura, Y., Inoue, T., Umesono, Y., Sano, S., Yoshimoto, K., Inden, M., Takata, K., Taniguchi, T., Shimohama, S., et al. (2007). Reconstruction of dopaminergic neural network and locomotion function in planarian regenerates. Devel Neurobio *67*, 1059–1078.

Nislow, C., Ray, E., and Pillus, L. (1997). SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. Mol. Biol. Cell *8*, 2421–2436.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., and Dubchak, I. (2014). The genome portal of the Department of

Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. *42*, D26–D31.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23–28.

Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. Cell *105*, 487–497.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Önal, P., Grün, D., Adamidi, C., Rybak, A., Solana, J., Mastrobuoni, G., Wang, Y., Rahn, H.-P., Chen, W., Kempa, S., et al. (2012). Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. The EMBO Journal *31*, 2755–2769.

Org, T., Chignola, F., Hetényi, C., Gaetani, M., Rebane, A., Liiv, I., Maran, U., Mollica, L., Bottomley, M.J., Musco, G., et al. (2008). The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep *9*, 370–376.

Oro, A.E., McKeown, M., and Evans, R.M. (1990). Relationship between the product of the Drosophila ultraspiracle locus and the vertebrate retinoid X receptor. Nature *347*, 298–301.

Oviedo, N.J., Pearson, B.J., Levin, M., and Sánchez Alvarado, A. (2008). Planarian PTEN homologs regulate stem cells and regeneration through TOR signaling. Dis Model Mech *1*, 131–143; discussion 141.

Özdağ, H., Teschendorff, A.E., Ahmed, A.A., Hyland, S.J., Blenkiron, C., Bobrow, L., Veerakumarasivam, A., Burtt, G., Subkhankulova, T., Arends, M.J., et al. (2006). Differential expression of selected histone modifier genes in human solid cancers. BMC Genomics *7*, 90.

Palakodeti, D., Smielewska, M., Lu, Y.-C., Yeo, G.W., and Graveley, B.R. (2008). The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA *14*, 1174–1186.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Parsons, D.W., Li, M., Zhang, X., Jones, S., Leary, R.J., Lin, J.C.-H., Boca, S.M., Carter, H., Samayoa, J., Bettegowda, C., et al. (2011). The genetic landscape of the childhood cancer medulloblastoma. Science *331*, 435–439.

Partanen, J.I., Nieminen, A.I., and Klefstrom, J. (2009). 3D view to tumor suppression:

lkb1, polarity and the arrest of oncogenic c-myc. Cell Cycle *8*, 716–724.

Parvinen, M. (2005). The chromatoid body in spermatogenesis. Int. J. Androl. *28*, 189–201.

Pasek, S., Risler, J.-L., and Brézellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics *22*, 1418–1423.

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O., et al. (2011). Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet *43*, 830–837.

Pawlik, T.M., and Keyomarsi, K. (2004). Role of cell cycle in mediating sensitivity to radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. *59*, 928–942.

Pearson, B.J., and Alvarado, A.S. (2010). A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. Development *137*, 213–221.

Peiris, T.H., Hoyer, K.K., and Oviedo, N.J. (2014). Innate immune system and tissue regeneration in Planarians: An area ripe for exploration. Semin Immunol *26*, 295–302.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J *30*, 4198–4210.

Pelegri, F. (2003). Maternal factors in zebrafish development. Dev. Dyn. *228*, 535–554.

Pellettieri, J., Fitzgerald, P., Watanabe, S., Mancuso, J., Green, D.R., and Sánchez Alvarado, A. (2010). Cell death and tissue remodeling in planarian regeneration. Developmental Biology *338*, 76–85.

Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100–103.

Peña, P.V., Hom, R.A., Hung, T., Lin, H., Kuo, A.J., Wong, R.P.C., Subach, O.M., Champagne, K.S., Zhao, R., Verkhusha, V.V., et al. (2008). Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J. Mol. Biol. *380*, 303–312.

Pesavento, J.J., Yang, H., Kelleher, N.L., and Mizzen, C.A. (2008). Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. Mol Cell Biol *28*, 468–486.

Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A. (2012). TrxG and PcG Proteins but

Not Methylated Histones Remain Associated with DNA through Replication. Cell *150*, 922–933.

Petruk, S., Black, K.L., Kovermann, S.K., Brock, H.W., and Mazo, A. (2013). Stepwise histone modifications are mediated by multiple enzymes that rapidly associate with nascent DNA during replication. Nat Commun *4*, 2841.

Plaks, V., Kong, N., and Werb, Z. (2015). The Cancer Stem Cell Niche: How Essential Is the Niche in Regulating Stemness of Tumor Cells? Cell Stem Cell *16*, 225–238.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol *10*, 192–206.

Pullirsch, D., Härtel, R., Kishimoto, H., Leeb, M., Steiner, G., and Wutz, A. (2010). The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development *137*, 935–943.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064–1071.

Quail, D.F., Taylor, M.J., and Postovit, L.-M. (2012). Microenvironmental regulation of cancer stem cell phenotypes. Curr Stem Cell Res Ther *7*, 197–216.

Quintana, E., Shackleton, M., Sabel, M.S., Fullen, D.R., Johnson, T.M., and Morrison, S.J. (2008). Efficient tumour formation by single human melanoma cells. Nature *456*, 593–598.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rajendran, P.S., and Dalerba, P. (2014). Theoretical and Experimental Foundations of the “Cancer Stem Cell” Model. In Cancer Stem Cells, V.K. Rajasekhar, ed. (John Wiley & Sons), pp. 1–16.

Raška, O., Kostrouchová, V., Behenský, F., Yilma, P., Saudek, V., Kostrouch, Z., and Kostrouchová, M. (2011). SMED-TLX-1 (NR2E1) is critical for tissue and body plan maintenance in Schmidtea mediterranea in fasting/feeding cycles. Folia Biol. (Praha) *57*, 223–231.

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

Reddien, P.W., Bermange, A.L., Murfitt, K.J., Jennings, J.R., and Sánchez Alvarado, A. (2005a). Identification of Genes Needed for Regeneration, Stem Cell Function, and Tissue Homeostasis by Systematic Gene Perturbation in Planaria. Developmental Cell *8*, 635–649.

Reddien, P.W., Oviedo, N.J., Jennings, J.R., Jenkin, J.C., and Sánchez Alvarado, A.

(2005b). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science *310*, 1327–1330.

Reddien, P.W., Bermange, A.L., Kicza, A.M., and Sánchez Alvarado, A. (2007). BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. Development *134*, 4043–4051.

Reese, B.E., Bachman, K.E., Baylin, S.B., and Rountree, M.R. (2003). The Methyl-CpG Binding Protein MBD1 Interacts with the p150 Subunit of Chromatin Assembly Factor 1. Mol Cell Biol *23*, 3226–3236.

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

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Resch, A.M., and Palakodeti, D. (2012). Small RNA pathways in Schmidtea mediterranea. International Journal of Developmental Biology *56*, 67–74.

Resch, A.M., Palakodeti, D., Lu, Y.-C., Horowitz, M., and Graveley, B.R. (2012). Transcriptome Analysis Reveals Strain-Specific and Conserved Stemness Genes in Schmidtea mediterranea. PLoS ONE *7*, e34447.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Rheinbay, E., Suvà, M.L., Gillespie, S.M., Wakimoto, H., Patel, A.P., Shahid, M., Oksuz, O., Rabkin, S.D., Martuza, R.L., Rivera, M.N., et al. (2013). An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. Cell Rep *3*, 1567–1579.

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

Robb, S.M.C., Gotting, K., Ross, E., and Sánchez Alvarado, A. (2015). SmedGD 2.0: The Schmidtea mediterranea genome database. Genesis *53*, 535–546.

Roberts, T.L., Ho, U., Luff, J., Lee, C.S., Apte, S.H., MacDonald, K.P.A., Raggat, L.J., Pettit, A.R., Morrow, C.A., Waters, M.J., et al. (2013). Smg1 haploinsufficiency predisposes to tumor formation and inflammation. PNAS *110*, E285–E294.

Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. *6*, 597–610.

Rodriguez, J., Muñoz, M., Vives, L., Frangou, C.G., Groudine, M., and Peinado, M.A. (2008). Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. Proc. Natl. Acad. Sci. U.S.A. *105*, 19809–19814.

Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.A., Vultur, A., Basu, D., Gimotty, P., Vogt, T., and Herlyn, M. (2010). A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell *141*, 583–594.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. *20*, 7137–7148.

Romero, B.T., Evans, D.J., and Aboobaker, A.A. (2012). FACS analysis of the planarian stem cell compartment as a tool to understand regenerative mechanisms. Methods Mol. Biol. *916*, 167–179.

Rouhana, L., Weiss, J.A., King, R.S., and Newmark, P.A. (2014). PIWI homologs mediate Histone H4 mRNA localization to planarian chromatoid bodies. Development *141*, 2592–2601.

Ruault, M., Brun, M.E., Ventura, M., Roizès, G., and De Sario, A. (2002). MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. Gene *284*, 73–81.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol. Cell *25*, 15–30.

Sakabe, K., Wang, Z., and Hart, G.W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc. Natl. Acad. Sci. U.S.A. *107*, 19915–19920.

Saló, E. (2006). The power of regeneration and the stem-cell kingdom: freshwater planarians (Platyhelminthes). Bioessays *28*, 546–559.

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

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

Salvetti, A., Rossi, L., Bonuccelli, L., Lena, A., Pugliesi, C., Rainaldi, G., Evangelista, M., and Gremigni, V. (2009). Adult stem cell plasticity: neoblast repopulation in non-lethally irradiated planarians. Dev. Biol. *328*, 305–314.

Sanchez, R., and Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader. Trends in Biochemical Sciences *36*, 364–372.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell *15*, 595–605.

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

Sato, Y., Kobayashi, K., Matsumoto, M., Hoshi, M., and Negishi, S. (2005). Comparative study of eye defective worm “menashi” and regenerating wild-type in planarian, Dugesia ryukyuensis. Pigment Cell Research *18*, 86–91.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809–818.

Schaefer, M., and Lyko, F. (2010). Solving the Dnmt2 enigma. Chromosoma *119*, 35–40.

Schaeffer, D.J. (1993). Planarians as a model system for in vivo tumorigenesis studies. Ecotoxicol. Environ. Saf. *25*, 1–18.

Scharf, A.N.D., Barth, T.K., and Imhof, A. (2009). Establishment of Histone Modifications after Chromatin Assembly. Nucl. Acids Res. gkp518.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in Drosophila Embryos. PLoS Biol *7*, e1000013.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nature Reviews Molecular Cell Biology *12*, 799–814.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. PNAS *95*, 5857–5864.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Scimone, M.L., Meisel, J., and Reddien, P.W. (2010). The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian Schmidtea mediterranea. Development *137*, 1231–1241.

Scimone, M.L., Srivastava, M., Bell, G.W., and Reddien, P.W. (2011). A regulatory program for excretory system regeneration in planarians. Development *138*, 4387–4398.

Scimone, M.L., Kravarik, K.M., Lapan, S.W., and Reddien, P.W. (2014). Neoblast

Specialization in Regeneration of the Planarian Schmidtea mediterranea. Stem Cell Reports *3*, 339–352.

Sedkov, Y., Benes, J.J., Berger, J.R., Riker, K.M., Tillib, S., Jones, R.S., and Mazo, A. (1999). Molecular genetic analysis of the Drosophila trithorax-related gene which encodes a novel SET domain protein. Mechanisms of Development *82*, 171–179.

Sedkov, Y., Cho, E., Petruk, S., Cherbas, L., Smith, S.T., Jones, R.S., Cherbas, P., Canaani, E., Jaynes, J.B., and Mazo, A. (2003). Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. Nature *426*, 78–83.

Seenundun, S., Rampalli, S., Liu, Q.-C., Aziz, A., Palii, C., Hong, S., Blais, A., Brand, M., Ge, K., and Dilworth, F.J. (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. EMBO J. *29*, 1401–1411.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893–897.

Selman, K., and Kafatos, F.C. (1974). Transdifferentiation in the labial gland of silk moths: is DNA required for cellular metamorphosis? Cell Differentiation *3*, 81–94.

Senoo, M., Pinto, F., Crum, C.P., and McKeon, F. (2007). p63 Is essential for the proliferative potential of stem cells in stratified epithelia. Cell *129*, 523–536.

Shackleton, M., Quintana, E., Fearon, E.R., and Morrison, S.J. (2009). Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. Cell *138*, 822–829.

Shah, N., and Sukumar, S. (2010). The Hox genes and their roles in oncogenesis. Nat Rev Cancer *10*, 361–371.

Shen, H., and Laird, P.W. (2013). Interplay between the Cancer Genome and Epigenome. Cell *153*, 38–55.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Shibahara, K., and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell *96*, 575–585.

Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., and Agata, K. (1999). Expression ofvasa(vas)-Related Genes in Germline Cells and Totipotent Somatic Stem Cells of Planarians. Developmental Biology *206*, 73–87.

Shibata, N., Rouhana, L., and Agata, K. (2010). Cellular and molecular dissection of pluripotent adult somatic stem cells in planarians. Dev. Growth Differ. *52*, 27–41.

Shilatifard, A. (2012). The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annu Rev

Biochem *81*, 65–95.

Shilatifard, A., Lane, W.S., Jackson, K.W., Conaway, R.C., and Conaway, J.W. (1996). An RNA polymerase II elongation factor encoded by the human ELL gene. Science *271*, 1873–1876.

Siebold, A.P., Banerjee, R., Tie, F., Kiss, D.L., Moskowitz, J., and Harte, P.J. (2010). Polycomb Repressive Complex 2 and Trithorax modulate Drosophila longevity and stress resistance. Proc. Natl. Acad. Sci. U.S.A. *107*, 169–174.

Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science *283*, 1164–1167.

Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res. *14*, 6711–6719.

Sims, R.J., and Reinberg, D. (2006). Histone H3 Lys 4 methylation: caught in a bind? Genes Dev. *20*, 2779–2786.

Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., Paro, R., James, T.C., and Gaunt, S.J. (1991). A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. Nucleic Acids Res. *19*, 789–794.

Siomi, M.C., Mannen, T., and Siomi, H. (2010). How does the royal family of Tudor rule the PIWI-interacting RNA pathway? Genes Dev. *24*, 636–646.

Smith, C., Matheson, T., Trombly, D., Sun, X., Campeau, E., Han, X., Yates, J., and Kaufman, P. (2014). A separable domain of the p150 subunit of human Chromatin Assembly Factor-1 promotes protein and chromosome associations with nucleoli. Program in Gene Function and Expression Publications and Presentations.

Sobel, R.E., Cook, R.G., Perry, C.A., Annunziato, A.T., and Allis, C.D. (1995). Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. Proc. Natl. Acad. Sci. U.S.A. *92*, 1237–1241.

Solana, J., Lasko, P., and Romero, R. (2009). Spoltud-1 is a chromatoid body component required for planarian long-term stem cell self-renewal. Developmental Biology *328*, 410–421.

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

Stassen, M.J., Bailey, D., Nelson, S., Chinwalla, V., and Harte, P.J. (1995). The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal

proteins. Mech. Dev. *52*, 209–223.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Suh, E.-K., Yang, A., Kettenbach, A., Bamberger, C., Michaelis, A.H., Zhu, Z., Elvin, J.A., Bronson, R.T., Crum, C.P., and McKeon, F. (2006). p63 protects the female germ line during meiotic arrest. Nature *444*, 624–628.

Sun, X.-J., Xu, P.-F., Zhou, T., Hu, M., Fu, C.-T., Zhang, Y., Jin, Y., Chen, Y., Chen, S.-J., Huang, Q.-H., et al. (2008). Genome-Wide Survey and Developmental Expression Mapping of Zebrafish SET Domain-Containing Genes. PLoS ONE *3*, e1499.

Surapureddi, S., Rana, R., Reddy, J.K., and Goldstein, J.A. (2008). Nuclear receptor coactivator 6 mediates the synergistic activation of human cytochrome P-450 2C9 by the constitutive androstane receptor and hepatic nuclear factor-4alpha. Mol. Pharmacol. *74*, 913–923.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

Suzuki, H., Watkins, D.N., Jair, K.-W., Schuebel, K.E., Markowitz, S.D., Chen, W.D., Pretlow, T.P., Yang, B., Akiyama, Y., Van Engeland, M., et al. (2004). Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nat. Genet. *36*, 417–422.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev. *16*, 1779–1791.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51–61.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011a. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

Tan, T.C.J., Rahman, R., Jaber-Hijazi, F., Felix, D.A., Chen, C., Louis, E.J., and Aboobaker, A. (2012). Telomere maintenance and telomerase activity are differentially regulated in asexual and sexual worms. PNAS *109*, 4209–4214.

Tanaka, E.M., and Reddien, P.W. (2011). The Cellular Basis for Animal Regeneration. Developmental Cell *21*, 172–185.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from

professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Teshirogi, W., and Ishida, S. (1987). Biology of planarians—foundation, application and experiments. Kyoritsu Shuppan Inc., Tokyo 65–83.

Tie, F., Banerjee, R., Conrad, P.A., Scacheri, P.C., and Harte, P.J. (2012). Histone demethylase UTX and chromatin remodeler BRM bind directly to CBP and modulate acetylation of histone H3 lysine 27. Mol. Cell. Biol. *32*, 2323–2334.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. U.S.A. *96*, 8681–8686.

Trojer, P., and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Molecular Cell *28*, 1–13.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. *13*, 3822–3831.

Tweedie, S., Ng, H.-H., Barlow, A.L., Turner, B.M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in Drosophila melanogaster? Nat Genet *23*, 389–390.

Tyagi, S., and Herr, W. (2009). E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases. EMBO J. *28*, 3185–3195.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. *35*, W71–W74.

Vaissière, T., Sawan, C., and Herceg, Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research/Reviews in Mutation Research *659*, 40–48.

Valekunja, U.K., Edgar, R.S., Oklejewicz, M., van der Horst, G.T.J., O'Neill, J.S., Tamanini, F., Turner, D.J., and Reddy, A.B. (2013). Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. Proc Natl Acad Sci U S A *110*, 1554–1559.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922–926.

Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T., and Shapiro, L.J. (1982). Transformation with DNA from 5-azacytidine-reactivated X chromosomes. Proc. Natl. Acad. Sci. U.S.A. *79*, 2352–2354.

Vermeulen, L., De Sousa E Melo, F., van der Heijden, M., Cameron, K., de Jong, J.H.,

Borovski, T., Tuynman, J.B., Todaro, M., Merz, C., Rodermond, H., et al. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat. Cell Biol. *12*, 468–476.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M.V., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J.P., Göttgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. Nat. Struct. Mol. Biol. *17*, 617–619.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. N. Engl. J. Med. *319*, 525–532.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science *339*, 1546–1558.

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

WADDINGTON, C.H. (1956). Principles of embryology. x + 510 pp.

Wagner, D.E., Wang, I.E., and Reddien, P.W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. Science *332*, 811–816.

Wagner, D.E., Ho, J.J., and Reddien, P.W. (2012). Genetic Regulators of a Pluripotent Adult Stem Cell System in Planarians Identified by RNAi and Clonal Analysis. Cell Stem Cell *10*, 299–311.

Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M., et al. (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol. Cell. Biol. *29*, 6074–6085.

Wang, Y., Zayas, R.M., Guo, T., and Newmark, P.A. (2007). nanos function is essential for development and regeneration of planarian germ cells. Proc Natl Acad Sci U S A *104*, 5901–5906.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., and Jia, S. (2009b). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Mol. Cell *33*, 428–437.

Wang, Y., Krivtsov, A.V., Sinha, A.U., North, T.E., Goessling, W., Feng, Z., Zon, L.I., and

Armstrong, S.A. (2010). The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. Science *327*, 1650–1653.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

Weirich, S., Kudithipudi, S., Kycia, I., and Jeltsch, A. (2015). Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. Clinical Epigenetics *7*, 36.

Weissman, I. (2005). Stem cell research: paths to cancer therapies and regenerative medicine. JAMA *294*, 1359–1366.

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

Wenemoser, D., Lapan, S.W., Wilkinson, A.W., Bell, G.W., and Reddien, P.W. (2012). A molecular wound response program associated with regeneration initiation in planarians. Genes Dev *26*, 988–1002.

Wenger, Y., Buzgariu, W., and Galliot, B. (2016). Loss of neurogenesis in Hydra leads to compensatory regulation of neurogenic and neurotransmission genes in epithelial cells. Phil. Trans. R. Soc. B *371*, 20150040.

White, J., and Dalton, S. (2005). Cell cycle control of embryonic stem cells. Stem Cell Rev *1*, 131–138.

Williams, R.T., den Besten, W., and Sherr, C.J. (2007). Cytokine-dependent imatinib resistance in mouse BCR-ABL+, Arf-null lymphoblastic leukemia. Genes Dev. *21*, 2283–2287.

van Wolfswinkel, J.C., Wagner, D.E., and Reddien, P.W. (2014). Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. Cell Stem Cell *15*, 326–339.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol. Cell. Biol. *28*, 7337–7344.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90.

Xiao, B., Wilson, J.R., and Gamblin, S.J. (2003). SET domains and histone methylation. Current Opinion in Structural Biology *13*, 699–705.

Xu, M., Wang, W., Chen, S., and Zhu, B. (2012). A model for mitotic inheritance of histone lysine methylation. EMBO Rep *13*, 60–67.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a

novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851–861.

Yamamoto, H., and Agata, K. (2011). Optic chiasm formation in planarian I: Cooperative netrin- and robo-mediated signals are required for the early stage of optic chiasm formation. Development, Growth & Differentiation *53*, 300–311.

Yamauchi, K., Yang, M., Jiang, P., Yamamoto, N., Xu, M., Amoh, Y., Tsuji, K., Bouvet, M., Tsuchiya, H., Tomita, K., et al. (2005). Real-time in vivo dual-color imaging of intracapillary cancer cell and nucleus deformation and migration. Cancer Res. *65*, 4246–4252.

Yamazaki, J., Estecio, M.R., Lu, Y., Long, H., Malouf, G.G., Graber, D., Huo, Y., Ramagli, L., Liang, S., Kornblau, S.M., et al. (2013). The epigenome of AML stem and progenitor cells. Epigenetics *8*, 92–104.

Yilmaz, M., and Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. Cancer Metastasis Rev. *28*, 15–33.

Yilmaz, O.H., Valdez, R., Theisen, B.K., Guo, W., Ferguson, D.O., Wu, H., and Morrison, S.J. (2006). Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. Nature *441*, 475–482.

Yoshida-Kashikawa, M., Shibata, N., Takechi, K., and Agata, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. Dev. Dyn. *236*, 3436–3450.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res *21*, 564–578.

Zang, Z.J., Cutcutache, I., Poon, S.L., Zhang, S.L., McPherson, J.R., Tao, J., Rajasegaran, V., Heng, H.L., Deng, N., Gan, A., et al. (2012). Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet *44*, 570–574.

Zeng, A., Li, Y.-Q., Wang, C., Han, X.-S., Li, G., Wang, J.-Y., Li, D.-S., Qin, Y.-W., Shi, Y., Brewer, G., et al. (2013). Heterochromatin protein 1 promotes self-renewal and triggers regenerative proliferation in adult stem cells. J Cell Biol *201*, 409–425.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283.

Zhang, Y. (2006). It takes a PHD to interpret histone methylation. Nat Struct Mol Biol *13*, 572–574.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-Methyladenine DNA Modification in Drosophila. Cell *161*,

893–906.

Zhang, L., Eugeni, E.E., Parthun, M.R., and Freitas, M.A. (2003). Identification of novel histone post-translational modifications by peptide mass fingerprinting. Chromosoma *112*, 77–86.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189–1201.

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

Zhang, Z., Shibahara, K., and Stillman, B. (2000). PCNA connects DNA replication to epigenetic inheritance in yeast. Nature *408*, 221–225.

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

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

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. *39*, 61–69.

Zink, D., Fischer, A.H., and Nickerson, J.A. (2004). Nuclear structure in cancer cells. Nat Rev Cancer *4*, 677–687.

Zuckerkandl, E., and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. In Evolving Genes and Proteins, V. Bryson, and H. Vogel, eds. (New York: Academic Press), pp. 97–166.