MLL3/4 prevents stem cell hyperplasia and controls differentiation programs in a planarian cancer stem cell model.

Yuliana Mihaylova[1], Damian Kao[1], Samantha Hughes[2], Alvina Lai[1], Farah Jaber-Hijazi, Nobuyoshi Kosaka[1], Prasad Abnave[1], and A. Aziz Aboobaker[1*]

1. Department of Zoology, Tinbergen Building, South Parks Road, Oxford OX1 3PS, United Kingdom

2. HAN University of Applied Sciences, Institute of Applied Sciences, Laan van Scheut 2, 6525EM, Nijmegen, The Netherlands

Beatson Institute for Cancer Research, Switchback Road, Bearsden, Glasgow G61 1BD

***Author for correspondence**

Aziz.Aboobaker@zoo.ox.ac.uk

Background

The family of Mixed Lineage Leukaemia (MLL) histone methyltransferase proteins are often implicated in disease processes, particularly cancer. Here we focus on the MLL3 and MLL4 which are mutated in a high percentage of cancers implicating them as tumour suppressors, but very little is known about the underlying transcriptional and epigenetic changes that contribute to cancer.

Results

Here we make use of the highly accessible planarians model system to uncover a role for MLL3/4 in controlling stem cell differentiation and proliferation, that suggests conservation tumour suppressor over a large evolutionary timescale function for this epigenetic regulator. Knockdown of the planarian *Mll3/4* orthologs compromises stem cell differentiation and leads to hyper-proliferation and tumour-like outgrowth formation. The planarian system allows as to investigate the epigenetic and transcriptional studies changes in cells that will go on to form tumours at an early stage after loss of MLL3/4 function, identifying genome wide changes that occur early in the development of the pathology. This revealed mis-regulation of both conserved oncogenes and tumour suppressors, that together likely explain the cancer-like phenotype observed in planarians.

Conclusions

We confirm MLL3/4 tumour suppressor function and uncover a deep conservation of this role in stem cells. We find potentially conserved mis-regulated downstream targets driving the effects of MLL3/4 loss of function. Our work demonstrates the suitability of planarians for the study of epigenetic phenotypes related to cancer and stem cell function, and for capturing early causative changes in a definitive population of tumour forming stem cells *in vivo.*

Background

The pluripotent adult stem cell (pASC) population of planarian flatworms is a highly accessible study system to elucidate fundamental aspects of stem cell function[1,2]. These stem cells, collectively known as neoblasts (NBs), bestow these animals with an endless capacity to regenerate all the organs and tissues of this relatively simple organism after amputation. Comparisons of stem cell expression profiles and available functional data between planarians and other simpler animals with mammals show that key aspects of stem cell biology are deeply conserved[3-9]. Thus, studies of the NB population have the potential to inform us about the origins of fundamental stem cell properties and behaviors such as maintenance of genome stability[10], self-renewal[7,11], pluripotency[12-15], differentiation[16-18] and migration[19,20]. All of these are highly relevant to understanding human disease processes, particularly those leading to cancer. One exciting prospect is that planarian stem cells may be a suitable and simple model to study the molecular mechanisms that lead to the formation of tumor initiating cancer stem cells (CSCs).

Many conserved signaling pathways are known to be responsible for regulating growth, proliferation and other stem cell functions. In disease states changes in the activity of these pathways can be due to effects on expression levels, rather than mutations that change proteins, and this can be mediated epigenetically through chromatin modifications. It is not surprising, therefore, that mutations in chromatin modifying enzymes, like members of the Polycomb and Trithorax complexes, are implicated in cancer[21-26]. The genome-wide effects of chromatin modifying enzymes make understanding how they contribute to cancer phenotypes very challenging. Complexity in the form of tissue and cell heterogeneity, life history stage and stage of pathology make resolution of epigenetic regulatory cause and effect relationships *in vivo* very challenging. From this perspective, planarians and their accessible and relatively homogenous stem cell population may be a very useful model system, especially if the fundamental physiological effects of chromatin

84 modifying complexes are conserved. The system would be particularly
85 suitable for investigating the early transformative changes in stem cells at the
86 onset of cancer. Here, we chose to study the planarian ortholog of the human
87 tumor suppressors Mixed Lineage Leukaemia 3 (MLL3) and MLL4 to further
88 test the use of planarian NBs as a model of CSCs in the context of epigenetic
89 histone modifications.

90 The human MLL proteins are the core members of the highly conserved
91 COMPASS-like (complex of proteins associated with Set1) H3K4 methylase
92 complexes, and they all contain the 130-140 amino acid SET domain
93 (Su(var)3-9, Enhancer of zeste and Trithorax). An extensive research effort
94 has now established the evolutionary history and histone modifying activities
95 of this extended protein family (**Additional File 1**[27-42]). Perturbation of MLL-
96 mediated H3K4 methylase activity is characteristic of numerous cancer types.
97 While the most prominent examples are the translocations widely reported in
98 leukaemias involving the *Mll1* gene[43-46], the mutation rate of *Mll3* across
99 malignancies of different origin approaches 7%, making *Mll3* one of the most
100 commonly mutated genes in cancer[24]. In attempts to model the role of *Mll3* in
101 cancer, mice homozygous for a targeted deletion of the *Mll3* SET domain
102 were found to succumb to urether epithelial tumors at high frequency[32], an
103 effect enhanced in a *p53+/-* mutational background. Heterozygous deletions
104 of *Mll3* in mice also lead to acute myeloid leukaemia, implicating *Mll3* in dose-
105 dependent tumor suppression[26].

106 Recent studies have revealed an increasingly complicated molecular function
107 of MLL3, its closely related paralog MLL4, and their partial *Drosophila*
108 orthologs – LPT (Lost PHD-fingers of trithorax-related; corresponding to the
109 N-terminus of MLL3/4) and Trr (trithorax-related; corresponding to the C-
110 terminus of MLL3/4). LPT binds chromatin via its PHD (Plant Homeodomain)
111 finger domains and targets the H3K4 methylating function of Trr and,
112 potentially, the H3K27 demethylating action of UTX (Ubiquitously transcribed
113 tetratricopeptide repeat X), to specific places on the genome[35,39]. LPT-
114 Trr/MLL3/4 proteins have a role in transcriptional control via monomethylating
115 and/or trimethylating H3K4 in promoter and enhancer contexts[29,31,33-35,40,47,48]
116 (**Additional File 1**).

Links between cellular hyperplasia, molecular function, and potential downstream targets of LPT-Trr/MLL3/4 remain to be elucidated. Given the accessibility of NBs, planarians could provide an informative *in vivo* system for identifying conserved aspects of MLL3 and MLL4 function relevant to cancer. We find that the planarian *Mll3/4* homolog, like its *Drosophila* counterpart, has undergone gene fission leading to split orthologs that are all expressed in stem cells. Loss of function experiments result in failures in stem cell differentiation and formation of tumor-like tissue outgrowths caused by stem cell hyperplasia. These data suggest that fundamental roles in controlling stem cell behavior might be conserved between planarian and human *Mll3/4* genes. We performed both RNA-seq and ChIP-seq for key histone modifications in pre-outgrowth forming NBs to identify downstream effects. We find that early regulatory changes driving the *Mll3* loss of function cancer-like phenotype are rooted in mis-regulation of pathways that drive proliferation and differentiation. Mis-regulated genes include well-established oncogenes and tumor suppressors and suggest a potentially deep conservation of MLL3/4-mediated epigenetic regulation in stem cells. We find that concomitant knockdown of planarian *Mll3/4* and either a *pim-1-like* oncogene ortholog or a *utx* ortholog, both overexpressed in the *Mll3/4* loss of function over-proliferation phenotype, can rescue tumor outgrowths. This implicates these genes as key early regulatory downstream targets important for controlling stem cell proliferative activity. Our data demonstrate the power of the planarian model system for actively informing studies in mammalian systems.

Results

The planarian orthologs of *Mll3/4* are expressed in stem cells

We found 3 partial orthologs of mammalian *Mll3* and *Mll4* genes. The planarian gene homologous to *Drosophila* LPT and the N-terminus of mammalian *Mll3/4* was named *Smed-LPT* (KX681482) (**Additional File 2a**). SMED-LPT (LPT) protein contains two PHD-fingers and a PHD-like zinc-

binding domain, suggesting that it has chromatin-binding properties[49] (**Figure 1a**). There are two planarian genes homologous to *Drosophila* Trr and the C-terminus of mammalian *Mll3/4* – *Smed-trr-1* (KC262345) and *Smed-trr-2* (DN309269, HO004937), both previously described[36]. Both SMED-TRR-1 and SMED-TRR-2 contain a PHD-like zinc-binding domain, a FYRN (FY-rich N-terminal domain), FYRC (FY-rich C-terminal domain) and a catalytic SET domain. SMED-TRR-1 (TRR-1) contains only a single NR (Nuclear Receptor) box at a non-conserved position and SMED-TRR-2 (TRR-2) has no NR boxes (**Figure 1a**). This could indicate that the planarian members of the trithorax-related family are not broadly involved in recognition of nuclear receptors like their arthropod and mammalian counterparts [33,35,50,51]. It is also possible that some functional divergence exists between TRR-1 and TRR-2, where only TRR-1 is capable of interacting with nuclear receptors.

We performed wholemount *in situ* hybridization (WISH) and found that *LPT*, *trr-1* and *trr-2* are broadly expressed across many tissues and organs. Gamma irradiation, used to remove all cycling cells in *S. mediterranea* within 24 hours, revealed that the three transcripts are also likely to be expressed in stem cells (**Figure 1b**). This was supported by using an alternative method for stem cell depletion – *H2B*(RNAi)[7] (**Additional File 2c**). The genes also showed clear expression in the brain, pharynx and other post-mitotic differentiated tissues (**Figure 1b**). During regeneration *Mll3/4* gene orthologs are expressed in structures like the brain and pharynx as those are being reformed (**Additional File 2b**).

In order to confirm expression of all three transcripts in planarian stem cells we performed double fluorescent *in situ* hybridization (FISH) with the pan-stem cell marker *Histone2B* (*H2B*). We found conspicuous co-expression between *LPT*, *trr-1*, *trr-2* and *H2B*, with over 90% of all *H2B*-positive cells co-expressing the three transcripts (**Figure 1c-d**). These results confirmed the expression of all three transcripts in cycling cells. Analyses of RNA-seq experiments consolidated across multiple published Fluorescence Activated Cell sorting (FACS) datasets[9] revealed that 28% of *LPT*'s total expression is in the X1 FACS fraction (S/G2/M stem cells), 44% in the X2 fraction (G1 stem cells and stem cell progeny) and 28% in the X ins fraction (irradiation-

insensitive; differentiated cells) of FACS-sorted planarian cell populations (**Figure 1e**). Both *trr-1* and *trr-2* showed similar distribution in expression through FACS-sorted cell populations. This is in agreement with the observed ISH patterns, suggesting all 3 transcripts are widely expressed and co-expressed in cycling stem cells, stem cell progeny and in neuronal cells (**Figure 1b-e**, **Additional File 2d**). These data support the hypothesis that these proteins act together, with LPT binding chromatin to serve as a scaffold for TRR methyltransferase activity [35,39].

Loss of *Mll3/4* function leads to regeneration defects and tumor-like outgrowths

In order to study the function of planarian *Mll3/4,* we investigated phenotypes after RNAi-mediated knockdown. Following *LPT*(RNAi), there was a clear failure to regenerate missing structures, including the eyes and pharynx, with regenerative blastemas smaller than controls (**Figure 2a-b**). After 8 days of regeneration we observed that, as well as failure to regenerate missing structures, animals began to form tissue outgrowths (**Figure 2c-d**), with this phenotype being most pronounced in head pieces (75% of head pieces, 35% of tail pieces 40% of middle pieces) (**Additional File 3a**). Intact (homeostatic) *LPT*(RNAi) animals also developed outgrowths, but with decreased frequency compared to regenerates (**Additional File 3b**).

Following individual knockdown of *trr-1* and *trr-2*, milder differentiation defects were observed compared to *LPT*(RNAi), with no obvious outgrowths (**Additional File 3**, **Additional File 4a-f**), confirming results from an earlier study[36]. However, *trr-1/trr-2* double knockdown recapitulated the phenotype of *LPT*(RNAi), but with higher penetrance and increased severity (**Additional File 5**). Thus, functional redundancy between the two *trr* paralogs likely accounts for the reduced severity after individual knockdown. All double knockdown animals developed outgrowths and started dying as early as day 5 post-amputation. Based on these observations, we decided to focus our attention on the *LPT*(RNAi) phenotype as regeneration defects and the

formation of tissue outgrowths were temporally distinct and could be studied consecutively.

A more thorough study of the differentiation properties of *LPT*(RNAi) animals following amputation showed that the triclad gut structure failed to regenerate secondary and tertiary branches and to extend major anterior and posterior branches (**Figure 2e**). Cephalic ganglia (CG) regenerated as smaller structures, the two CG lobes did not join in their anterior ends in *LPT*(RNAi) animals (**Figure 2f**) and the optic chiasma and optic cups were mis-patterned and markedly reduced (**Figure 2g-h**). We found that 80% of *LPT*(RNAi) animals did not regenerate any new pharyngeal tissue (**Figure 2i**). We interpreted these regenerative defects as being indicative of either a broad failure in stem cell maintenance or in differentiation. The number of NBs (*H2B*+ve) or early stem cell progeny cells across all lineages (SMEDWI-1+ve/*H2B*-ve) [52,53] was not affected by perturbation of *LPT* function (**Additional File 6**). Therefore, we infer that defects in later stem cell progeny formation and terminal differentiation likely underpin failure in tissue regeneration. Alternatively, different early stem cell progeny lineages could be affected in opposite directions depending on their lineage, leading to the observation of no overall change in the bulk number of early stem cell progeny (SMEDWI-1+ve/*H2B*-ve).

Mll3/4 function is required for correct differentiation of epidermal and neural lineages

One of the structures most severely affected following loss of *Mll3/4* function was the brain. To investigate this further, we looked at the regeneration of different neuronal subtypes. *LPT*(RNAi) animals had reduced numbers of GABAergic (**Figure 3a**), dopaminergic (**Figure 3b**), acetylcholinergic (**Figure 3c**) and serotonergic (**Figure 3d**) neurons. As expected, the brain defects were milder following knockdown of *trr-1* or *trr-2* (**Additional File 7a-d**).

Among other tissues, the epidermis was also affected. Both early (*NB.21.11e*+ve cells) and late (*AGAT-1*+ve cells) epidermal progeny cells were significantly decreased, but not entirely absent, in *LPT*(RNAi)

regenerating animals (**Figure 3e**). No such defect was seen in *trr-1* and *trr-2* knockdown animals (**Additional File 7e**).

While we observed defects in the pharynx, neurons, gut and epidermal progeny, not all lineages were affected by *LPT*(RNAi). Some cell lineages and organs were correctly regenerated, including protonephridia and ventral and dorsal cilia (**Additional File 8**).

One of the other observable defects in *LPT*(RNAi) animals was abnormal locomotion, with nearly all worms displaying muscular inch worming rather than smooth cilia-mediated locomotion. Given differentiation of cilia was unaffected, it seemed likely this effect was a result of neuronal differentiation defects, specifically serotonin-dependent control of beating cilia [54]. In agreement with this interpretation, ectopic serotonin hydrochloride treatment improved the gliding movement of *Mll3/4* knockdown animals (**Additional File 9**).

Overall, our data demonstrate that regenerative defects caused by the abrogation of *Mll3/4* function are associated with broad failures in stem cell differentiation to produce some but not all lineages.

Mll3/4 limits normal stem cell proliferation and tissue growth

Aside from impairment of regeneration following *LPT*(RNAi), the other major phenotype we observed were outgrowths of tissue that appeared in unpredictable positions in regenerating pieces. Only two previously reported planarian RNAi phenotypes have similar pervasive outgrowths of this nature, and these were caused by hyperplastic stem cell proliferation after knockdown of other tumor suppressors [55,56].

Planarian regeneration is characterized by an early burst of increased NB proliferation, 6-12 hours after wounding, and a second peak of proliferation, 48 hours after amputation[53]. Following *LPT*(RNAi), we observed significant increases in proliferation at both of these peaks and at 8 days post-amputation, as proliferation fails to return to normal homeostatic levels (**Figure 4a**). *Trr-1*(RNAi) and *trr-2*(RNAi) animals also show elevated proliferation in response to amputation (**Additional File 10a**) and similar

277 increases in cell division are seen in knockdown animals that are left
278 unwounded (**Additional File 10b**).

279 In 8 day-regenerating *LPT*(RNAi) worms the observed over-proliferation is a
280 result of localized clusters of mitotic cells (**Figure 4b**). Since 8 days of
281 regeneration is the last stage before outgrowth formation commences, these
282 clusters likely correspond to sites of future outgrowths (**Figure 4c**). Similar
283 mitotic clusters are also seen at later stages of regeneration in animals that
284 are yet to develop outgrowths (**Figure 4d**). We looked specifically in
285 outgrowths and found mitotic cells (**Figure 4d-e**). NBs not in M-phase (*H2B*-
286 positive/anti-H3P-negative) are also found outside of their usual
287 morphological compartments in tissue outgrowths (**Additional File 11**).

288 In order to understand if ectopically cycling NBs represented the breadth of
289 known stem cell heterogeneity in planarians or only a subset of lineages, we
290 performed FISH for markers of the *sigma* (collectively pluripotent NBs), *zeta*
291 (NBs committed to the epidermal lineage) and *gamma* (NBs committed to the
292 gut lineage) cell populations [57]. We found that different NB populations are
293 represented in the outgrowths of *LPT*(RNAi) animals. (**Figure 5a, b, c**). Some
294 outgrowths contain *gamma*+/*Smedwi*- cells (**Figure 5b**), demonstrating that
295 *LPT*(RNAi) animals form outgrowths comprised of potential cell fates that
296 would not normally be part of the epidermis.

297 *Sigma*, *zeta* and *gamma* NBs are not significantly increased in pre-outgrowth
298 *LPT*(RNAi) animals (**Additional File 12**), suggesting that the presence of
299 these cells in outgrowths is not a secondary effect of increased cell number
300 and passive spread of these populations.

301 The epidermal progeny markers *NB.21.11e* and *AGAT-1* are concentrated in
302 the outgrowths of *LPT*(RNAi) animals, while being relatively sparsely
303 expressed in non-outgrowth tissue (**Additional File 13a**). The observed
304 disarray of *NB.21.11e*-positive and *AGAT-1*-positive cells in outgrowths could
305 relate to the perturbed patterning of the epidermal layer in *LPT*(RNAi) animals
306 (**Additional File 13b**). Epidermal cells appear to have lost polarity and to be
307 no longer capable of forming a smooth epidermal layer. Furthermore, the
308 average epidermal nuclear size is significantly increased compared to control

(**Additional File 13c**), an effect similar to the pathology seen following knockdown of the tumor suppressor *SMG-1*[56]. The epithelial layer in *LPT*(RNAi) animals also appears less well-defined than that in control animals, with a blurred distinction between epithelium and mesenchyme. Another feature of the *LPT*(RNAi) phenotype, encountered in a variety of malignancies[58], is changed nuclear shape (**Additional File 13d**).

In summary, LPT controls NB proliferation and restricts stem cells to pre-defined tissue compartments. Experiments described earlier showed that LPT, and the other two planarian partial MLL3/4 orthologs, are responsible for the successful differentiation of some, but not all lineages. Thus, taken together, our data demonstrate that disturbance of the function of planarian MLL3/4 COMPASS-like complex by *LPT*(RNAi) leads to development of both differentiation and proliferation defects with cancer-like features (**Figure 6**).

***LPT*(RNAi) results in transcriptional changes consistent with driving proliferation in stem cells**

A key insight missing from the literature for many tumor suppressors, including MLL3 and MLL4, is how they regulate the behavior of transformed stem cells at early stages of cancer. To tackle this question, we decided to focus on early regeneration when *LPT*(RNAi) animals do not yet exhibit any outgrowth phenotype. We performed RNA-seq on X1 (G2/M) fluorescence activated cell sorted (FACS) NBs from *LPT*(RNAi) and *GFP*(RNAi) planarians at 3 days of regeneration. Our analysis revealed that 540 transcripts are down-regulated (fold change <= -1.5, $p<0.05$) and 542 –up-regulated (fold change >= 1.5, $p<0.05$) in X1 stem cells from *LPT*(RNAi) animals when compared to controls (**Additional file 20**).

A recent meta-analysis of all available *S. mediterranea* RNA-seq data allowed classification of all expressed loci in the planarian genome by their relative expression in FACS sorted cell populations representing stem cells, stem cell progeny and differentiated cells[9]. Superimposing the differentially expressed genes following *LPT*(RNAi) onto a gene expression spectrum reflecting FACS compartments shows that *LPT*(RNAi) has a broad effect on gene expression

in X1 cells (**Figure 7a**), affecting genes normally expressed in many different planarian cell compartments (**Figure 7b**). These findings confirm that *LPT*(RNAi) has a complex effect, influencing different gene classes, including 47 transcription factors, in stem cells.

Analysis of Gene Ontology (GO) terms revealed a clear enrichment for cell cycle and cell division-associated terms in the list of up-regulated genes (**Figure 7c**), in agreement with the observed hyperproliferation in *LPT*(RNAi) phenotype. The list of down-regulated genes is also enriched for cell cycle-related terms, as well as cell differentiation and metabolism-related processes (**Figure 7c**). Genes associated with metabolic processes have been previously shown to be down-regulated following *Mll3/4* loss of function [50,59].

LPT(RNAi)-induced changes to promoter H3K4 methylation and transcription are correlated

Previous studies tie MLL3/4/LPT-Trr function directly to mono- and tri-methylation of H3K4 [29,31-34] and indirectly to trimethylation of H3K27, because the H3K27me3 demethylase UTX is present in the same protein complex [60]. In order to understand potential epigenetic causes of the transcriptional changes following *LPT*(RNAi), we also performed ChIP-seq on X1 cells. The profile of H3K4me3, H3K4me1 and H3K27me3 in control X1 cells showed that genes enriched in X1 stem cells have the highest H3K4me3 and the lowest H3K27me3 signal at predicted transcriptional start sites (TSSs), consistent with active transcription of these genes (**Figure 8a, Additional File 14**). We observed the reverse pattern for genes enriched in differentiated cells, consistent with repressed transcription. Furthermore, the peak in H3K4me1 signal is shifted downstream and away from the TSS for genes with enriched expression in X1 NBs, allowing for active transcription. Conversely, the peak in H3K4me1 is positioned across the TSS for genes with enriched expression in differentiated cells and no or relatively low expression in NBs, indicative of repressed transcription in NBs. These data are in agreement with previous reports in planarians and mammalian cells[9,34,37].

372 *LPT*(RNAi) led to a broad decrease in the level of both H3K4me3 and
373 H3K4me1 from just upstream and across the TSSs throughout the genome,
374 consistent with an active role for MLL3/4 in deposition of these histone
375 modifications. This was particularly true for the H3K4me1 mark at the TSSs of
376 genes whose expression is normally enriched in differentiated cells (**Figure**
377 **8a**, **Additional File 15a**). Concomitant with this, we also observed an
378 increase in H3K4me1 signal upstream of the predicted TSS (**Figure 8a**,
379 **Additional File 15a**). For the H3K27me3 mark, no clear pattern was
380 observed across the genome as result of *LPT*(RNAi) in any group of genes
381 subdivided by FACS compartment expression profiles.

382 We next looked more closely at the promoter histone methylation status of
383 those genes whose transcript levels were affected by *LPT*(RNAi) (**Additional**
384 **File 15b)**. Most notably, for genes enriched for X1 NB expression, we
385 observed an inverse relationship between expression following *LPT*(RNAi)
386 and amount of TSS-proximal H3K4me1. This suggests that *LPT*(RNAi) leads
387 to a reduction of this repressive mark at these loci and subsequent up-
388 regulation of expression in stem cells. For mis-regulated genes not enriched
389 in X1 NBs, we observed instead a positive correlation between changes in
390 transcriptional expression and changes in H3K4me3 levels (**Figure 8b**).
391 Overall, our data suggest that reductions in H3K4me1 following *LPT*(RNAi)
392 cause up-regulation of some of the stem cell genes implicated by our RNA-
393 seq data from *LPT*(RNAi) animals, while reductions in H3K4me3 are related
394 to down-regulation of non-NB enriched genes.

395 Our data demonstrate that key features of promoter-centric histone
396 modification-mediated control of transcription are conserved between
397 planarians and mammals, as previously shown[9,37]. Consistent with MLL3/4's
398 known role in H3K4 methylation, changes in gene expression following
399 *LPT*(RNAi) are correlated with the amount of H3K4me1 and H3K4me3.

400

401 **_LPT_(RNAi) leads to up-regulation of known and putative oncogenes and**
402 **down-regulation of tumor suppressors**

After observing the global changes in expression and histone modification patterns following *LPT*(RNAi), we wanted to identify individual mis-regulated genes that could potentially be major contributors to the differentiation and tumor-like phenotype, and assess which of these were potentially direct or indirect targets of MLL3/4 activity in stem cells.

The well-known tumor suppressor *p53*, where hypomorphic loss of function has been previously shown to cause dorsal outgrowths in planarians [61], was found to be significantly down-regulated in X1 stem cells after *LPT*(RNAi) (**Figure 9a**). Consistent with this, we observed a small decrease in H3K4me3 around the promoter region. Other cancer-related genes, like the tumor suppressor *PR domain zinc finger protein 1* (*PRDM1*, also known as *Blimp-1*) and the Polycomb gene *Suppressor of zeste 12* (*Su(z)12*) also had significantly altered expression following *LPT*(RNAi), both correlating with changes in H3K4me3 levels. While an increase in H3K4me3 on the *Su(z)12* promoter would not be predicted as an effect caused by *LPT*(RNAi), elevated levels of the H3K4me3 are consistent with up-regulation and may result from subtle effects on H3K4me1 levels.

The *pituitary homeobox* (*pitx*) gene was also significantly up-regulated in expression and had elevated levels of H3K4me3 on its promoter. *Pitx* is expressed in the serotonergic neuronal precursor cells[54,62] and thus, in planarians, it is not directly implicated in stem cell proliferation, but rather in differentiation. Nonetheless, the fact that *LPT*(RNAi) led to *pitx* up-regulation was of great interest for two reasons. Firstly, we knew that serotonergic neurons require PITX function or fail to regenerate in planarians [54,62] and, secondly, in human medullobalstomas down-regulation of *Mll3* and over-expression of *pitx2* are co-occurrences (Pomeroy Brain Oncomine dataset[63] ((www.oncomine.org)) (**Figure 9b**). To investigate the cellular basis for *pitx* overexpression, we performed FISH for this gene in *LPT*(RNAi) animals. We observed an accumulation of *pitx*-positive cells in *LPT*(RNAi) regenerates (**Figure 9c**). Given that production of terminally differentiated serotonergic neurons is decreased (**Figure 3d**), the increase of *pitx*-positive cells following *LPT*(RNAi) marks the accumulation of serotonergic neuronal precursors that fail to differentiate. Whether *pitx* up-regulation is causal or just a marker in the

failure of serotonergic neuron regeneration is not clear, but MLL3/4/LPT-Trr activity does control the maturation of serotonergic neuronal precursors into serotonergic neurons. The *LPT*(RNAi)-dependent up-regulation of *pitx* might also be a conserved feature of MLL3/4 mis-regulation in some cancer types [63].

While we observed an agreement between expression levels and changes in H3K4me3 around the TSS for many mis-regulated genes, this was not the case for all genes. One example where transcriptional expression is significantly up-regulated, but H3K4me3 levels are slightly down-regulated is *utx* (**Figure 10a**). This finding suggests that for some mis-regulated genes there is no direct relationship between LPT activity at their promoter regions and gene expression. The effect on *utx* expression is particularly significant as UTX itself may interact with MLL3/4/LPT-Trr and regulate gene expression across the genome.

We identified two planarian orthologs of the serine/threonine kinase oncogene *pim-1* (*Smed-pim-1* and *Smed-pim-1-like* or *pim-1* and *pim-1-like* in short) with increased levels of expression in stem cells following *LPT*(RNAi). Like *utx*, *pim-1* (KY849969) did not show an increase in H3K4me3 levels on its promoter. For *pim-1-like* (KY849970), whose expression is enriched in NBs, promoter-proximal H3K4me1 levels were examined instead, based on previous correlation analysis (**Figure 8b**) establishing H3K4me1 as the most predictive mark of transcriptional expression for X1 enriched genes. We observed that H3K4me1 was decreased at the *pim-1-like* TSS, consistent with increased transcriptional levels. These data suggest *pim-1-like* is a direct MLL3/4 target. Observation of these two orthologs of the pim kinase oncogene and *utx* suggests that up-regulated genes identified in the RNA-seq dataset include those with and without correlated changes in histone modification patterns at promoters. Other genes associated with cancer and development were also mis-regulated following *LPT*(RNAi) with patterns of H3K4 methylation that were both consistent and inconsistent with changes in transcript levels (**Additional File 16**).

The up-regulation of *pim-1* has been associated with genome instability [64] and onset of malignancy [65,66], while the up-regulation of the MLL3/4 partner and

H3K27me3 demethylase, *utx,* has been implicated in increased proliferation and tumor invasiveness [67]. Based on this, the overexpression of *utx, pim-1* and *pim-1-like* together represented some of the best candidates for major effects amongst those with significant up-regulation in expression. In order to test whether the up-regulated expression of *pim-1*, *pim-1-like* or *utx* is potentially key to the *LPT*(RNAi) cancer-like phenotype, we attempted *LPT*(RNAi) rescue experiments in the form of double RNAi knockdowns (**Additional File 17)**. At 48 hours post-amputation, *LPT*(RNAi) regenerates have a significantly increased stem cell proliferation (**Figure 4a**, **b**) and so do *GFP/LPT*(RNAi) double knockdown animals (**Figure 10b**). Whereas *LPT/pim-1*(RNAi) regenerates still have elevated NB proliferation, both *LPT/pim-1-like*(RNAi) and *LPT/utx*(RNAi) regenerates have a significantly decreased NB proliferation compared to *GFP/LPT*(RNAi), and half as many animals in these two conditions went on to form outgrowths (**Figure 10c**). These findings suggest that the up-regulation of both *pim-1-like* and *utx* are involved in driving the *LPT*(RNAi) animals' cancer-like phenotype. *Pim-1-like*'s up-regulation in *LPT*(RNAi) planarians may be directly connected to changing histone modifications at its promoter, while *utx*'s up-regulation is likely not due to a direct effect on a promoter region (**Figure 10a**).

Our study not only shows a conserved physiological role in controlling cell proliferation for the conserved MLL3/4 gene family, but also has allowed us to identify novel gene targets of LPT and MLL3/4-mediated transcriptional control in stem cells and begin to elucidate the mechanisms behind *Mll3/4* loss of function phenotypes. Some of these mechanisms are also likely to be conserved in mammals.

Discussion

Mll3 and *Mll4* have been implicated in different malignancy landscapes [24], with clear evidence for tumor suppressor roles in mammalian systems [26,32,68]. Our study demonstrates that loss of function of the planarian *Mll3/4* ortholog also results in the emergence of a cancer-like phenotype characterized by differentiation and proliferation defects. Our work shows that LPT, TRR-1 and

TRR-2 control differentiation to form many (gut, eyes, brain, pharynx), but not all lineages (cilia, protonephridia), suggesting that the MLL3/4 COMPASS-like complex is not a universal and unilateral regulator of differentiation. This conclusion is supported by the opposing effects of *LPT*(RNAi) on different lineages of stem cell progeny production. For example, while epidermal NB progeny (*NB.21.11e*- and *AGAT-1*-positive cells) were decreased, the number of serotonergic neuronal precursors (*pitx*-positive cells) was increased. Such differential effects might be related to the diverse molecular function of MLL3/4 proteins, associated with both positive and negative regulation of transcription via control of both enhancer and promoter activity [29,31,33-35]. Future work in planarians will allow closer investigation of these and other epigenetic effects on stem cell function. Study of enhancers, in particular, will benefit from further improvements in planarian genome assemblies, to allow both epigenetic and comparative genomic methods for enhancer detection.

We found that clusters of mitotic cells preceded the appearance of outgrowths in *LPT*(RNAi) regenerating animals, likely pre-empting where the outgrowths would subsequently form. The observation of clusters of cells and the formation of outgrowths in some but not all RNAi animals suggests a heterogeneity in stem cell responses to *LPT*(RNAi). This probably reflects the stochastic nature of the broad genome wide epigenetic changes that will have some variability between cells, such that only some cells cycle out of control and cause outgrowths. We also observed that outgrowth tissue contained different classes of stem cells. Among these stem cells, the presence of sigma NBs (thought to include truly pluripotent stem cells [57] is of particular significance. When mis-regulated, these cells could share fundamental similarities with cancer stem cells (CSCs) often found in human malignancies [69]. CSCs have been described as one of the main factors in cancer aggressiveness and resistance to treatment [70]. Studying such cells in a simple *in vivo* stem cell model provided by the planarian system should bring further insight into important control mechanisms that are mis-regulated in different cancers. Our work here provides a useful example of this approach.

Our data suggest that LPT regulates expression of genes across cell types, including some genes with enriched expression in stem cells. Genes with

significant expression differences following *LPT* knockdown were mostly associated with cell proliferation, differentiation and metabolic processes. A subset of genes where RNA-seq and ChIP-seq data correlate are likely a direct consequence of *LPT*(RNAi) affecting promoter histone methylation status. Genes with altered expression where there is no such correlation, may represent indirect (secondary) changes or, alternatively, may have enhancers that have altered histone modifications as a result of *LPT*(RNAi). Future work will develop the use of planarians as a model of epigenetic gene regulation and allow further investigation.

One of the most famous and well-studied tumor suppressors – *p53*, was significantly down-regulated following *LPT*(RNAi). *P53* acts as a cell cycle checkpoint guardian and has been reported to undergo mutations in more than 40% of all cancers[24]. *P53*(RNAi) planarians exhibit hyper-proliferation and outgrowth formation (dependent on the dose), suggesting some conservation of function [61]. Studies in mice have postulated that *Mll3*'s role in cancer is (at least partially) dependent on *p53* function [26,32] and this may also be the case in planarians.

Many of the genes overexpressed as a result of *LPT*(RNAi) may have roles in driving inappropriate stem cell activity, and some of these may be directly regulated by MLL3/4. For example, the expression of the H3K27me3 demethylase*, utx,* was significantly increased in stem cells following *LPT*(RNAi). UTX is itself a part of the MLL3/4/Trithorax-related protein complex[60,71] and UTX protein and mRNA overexpression has been linked to increased cell proliferation and invasiveness in breast cancer [67]. Our RNA-seq results also identified two planarian homologs of the oncogene *pim-1*, called *Smed-pim-1* and *Smed-pim-1-like*, that were overexpressed in stem cells following *LPT*(RNAi). Amongst overexpressed genes, these represented likely candidates for contributing to the *LPT*(RNAi). Overexpressed target genes can potentially be validated as having a role in *Mll3/4* loss of function pathology in planarians by double RNAi experiments. We found that double RNAi with either *utx* or *pim-1-like*, was sufficient to rescue *Mll3/4* loss of function over-proliferation and outgrowth phenotypes induced by *LPT*(RNAi). This provides strong support for the hypothesis that the over-expression of

these two genes was significant in driving stem cell hyperplasia. These experiments demonstrate the value of our approach to identify potential downstream targets and implicate novel regulatory interactions driving the *Mll3/4* loss of function phenotype. These targets can now be tested for conservation in mammalian experimental systems.

Conclusion

In conclusion, our study confirms conservation of function between mammalian *Mll3* and *Mll4* genes and their planarian orthologs. We identified candidates that are mis-regulated by *LPT*(RNAi) that may be conserved targets of MLL3/4 and may help explain how *Mll3/4* loss of function mutations contribute to human cancers. These findings demonstrate the strength of the planarian system for understanding fundamental stem cell mechanisms and its potential for in-depth investigation of epigenetic mis-regulation in cancer-causing stem cells.

Methods

Animal husbandry

Asexual freshwater planarians of the species *S. mediterranea* were used. The culture was maintained in 1x Montjuic salts water [72]. Planarians were fed organic calf liver once a week. After every feeding, the water was changed. Planarians were starved for 7 days prior to each experiment. They were also starved throughout the duration of each experiment.

RNAi

Double-stranded RNA (dsRNA) was synthesized from DNA fragments cloned in pCRII (Invitrogen) or pGEM-T Easy (Promega) vectors. T7 (Roche) and SP6 (NEB) RNA polymerases were used for transcription of each strand. The two transcription reactions were combined upon ethanol precipitation. RNA was denatured at 68 ºC and re-annealed at 37 ºC. Quantification was performed on a 1% agarose gel and Nanodrop spectrophotometer.

For single RNAi experiments a working concentration of 2 µg/µl was used. For double RNAi, each gene's RNA was at a concentration 4 µg/µl, resulting in solution concentration of 2 µg/µl.

DsRNA was delivered via microinjection using 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). Each animal received around 100 nl dsRNA each day. *H2B*(RNAi) was performed for three consecutive days, as per Solana et al.'s (2012) protocol. For single and double *LPT*, *trr-1* and *trr-2* knockdown, a course of 7 days of microinjections was performed (3 consecutive days + 2 days rest + 4 consecutive days). S*et1*(RNAi) and *utx*(RNAi) were performed for 4 consecutive days.

Primers used for amplification of DNA for dsRNA synthesis can be found in **Supplementary Table 2**.

In situ hybridization

RNA probes labeled with digoxigenin and fluorescein were generated via anti-sense transcription of DNA cloned in PCRII (Invitrogen) or PGemTEasy (Promega) vector. *In situ* hybridization was performed as described in King and Newmark's (2013) protocol for most fluorescent experiments. For *LPT*, *trr-1*, *trr-2*, *sigma*, *zeta* and *gamma* fluorescent *in situ* procedures, a pooled probes method was used, as described in van Wolfswinkel et al.[57]. Colorimetric *in situ* hybridization procedures were performed as described in Gonzalez-Estevez et al.[73]. Primers used for amplification of DNA for RNA probe synthesis can be found in (**Additional File 21)**.

Immunohistochemistry

Immunohistochemistry was performed as described in Cebria and Newmark[74]. Antibodies used were: anti-H3P (phosphorylated serine 10 on histone H3; Millipore; 09-797; 1:1000 dilution), anti-VC1 (kindly provided by Prof. Hidefumi Orii (check title); 1:10000 dilution), anti-SMEDWI-1 (kindly provided by Prof.

Jochen Rink; 1:500 dilution), anti-SYNORF-1 (3C11; Developmental Studies Hybridoma Bank; 1:50 dilution), anti-acetylated tubulin (Developmental Studies Hybridoma Bank; 1:200 dilution).

Imaging and image analysis

Colorimetric images were taken on Zeiss Discovery V8 (Carl Zeiss) microscope with a Canon EOS 600D or Canon EOS 1200D camera. Fluorescent images were taken on either Inverted Olympus FV1000 or FV1200 Confocal microscope. Cells were counted via Adobe Photoshop CS6 or FIJI software and the count was normalized to imaged area in mm^2.

Flow cytometry

A modified version of Romero et al.'s[75] planarian FACS protocol was used, as described in Kao et al.[9]. A FACS Aria III machine equipped with a violet laser was used for the sort. BD FACSDiva and FlowJo software was used for analysis and gate-setting.

Western blot

2xLaemmli buffer (Sigma Aldrich), 1M DTT and cOmplete protease inhibitors (Roche) were used for protein extraction from 10-15 animals per condition. Protein extract was quantified with Qubit Protein Assay kit (Thermo Fisher Scientific). NuPAGE Novex 4-12% Bis-Tris protein gels (Thermo Fisher Scienitific) were used, followed by a wet transfer in a Mini Trans-Blot Electrophoretic Transfer Cell machine. Ponceau S (Sigma Aldrich) whole-protein stain was used prior to antibody incubation. The antibodies used were: anti-H3 (unmodified histone H3; rabbit polyclonal; Abcam; ab1791; 1:10000 dilution), anti-H3K4me3 (rabbit polyclonal; Abcam; ab8580; 1:1000 dilution), anti-H3K4me1 (rabbit polyclonal; Abcam; ab8895; 1:1000 dilution), anti-H3K27me3 (mouse monoclonal; Abcam; ab6002; 1:1000 dilution), anti-mouse IgG HRP-linked antibody (Cell Signalling; 7076P2), anti-rabbit IgG HRP-linked

antibody (Cell Signalling; 7074P2). The experiments were done to validate the specificity of the histone modification antibodies (**Additional File 18**).

ChIP-seq

600000-700000 planarian x1 cells were FACS-sorted (using 3-day knockdown regenerates) in PBS and pelleted at 4 ºC. During the pelleting, S2 cells were added (corresponding to roughly 15% of the number of planarian x1 cells) for the purpose of downstream data normalisation. Samples were then processed as described in Kao et al. (2017). The process is summarized in **Additional File 19**. The libraries were sequenced on an Illumina NextSeq machine. Three biological replicates were prepared. The raw reads are available in the Short Read Archive (PRJNA338116).

RNA-seq

300000 x1 NBs were FACS-sorted in RNALater (Ambion) from knockdown animals at 3 days of regeneration. Cells were pelleted at 4 ºC and Trizol-based total RNA extraction was performed. The amount of total RNA used for each library preparation was 0.8-1 µg. Illumina TruSeq Stranded mRNA LT kit was used for library preparation. The kit instructions were followed. Libraries were quantified with Qubit, Agilent Bioanalyzer and KAPA Library Quantification qPCR kit. Samples were sequenced on an Illumina NextSeq machine. Two biological replicates were prepared. The raw reads are available in the Short Read Archive (PRJNA338115).

ChIP-seq data analysis

ChIP-seq reads were trimmed with Trimmomatic 0.32[76] and aligned to the *S. mediterranea* SmedGD asexual genome 1.1[77] and *D.melanogaster* genome r6.10[78] with BWA mem 0.7.12. Picard tools 1.115 was used to remove read duplicates after mapping. Python scripts were used to filter and separate out read pairs belonging to either genome. ChIP-seq coverage tracks were then

generated and normalized according to Orlando et al.[79]. For more in-depth methods, including code, refer to the **Supplementary Python Notebook**.

RNA-seq data analysis

Raw reads were trimmed with Trimmomatic 0.32[76] and pseudo-aligned to a set of asexual genome annotations described in Kao et al. (2017) with Kallisto 0.42[80]. Differential expression was subsequently performed with Sleuth 0.28.1[81]. For more in-depth methods, including code, refer to the **Supplementary Python Notebook**.

Statistical methods

Wherever cell number was compared between experimental condition and control, a 2-tailed ttest assuming unequal variance was used. Each legend states the number of specimens per condition, where relevant. Bar graphs show the mean average and the error bars are always Standard Error of the Mean.

For analysis of RNA-seq data, Wald's test (as part of the Sleuth[82] software) was used for assessing differential expression. Spearman's rank correlation was used for assessing the correlation between RNA-seq and ChIP-seq data. Hypergeometric tests were used for assessing enrichment in the RNA-seq data.

Data availability

The ChIP-seq and RNA-seq datasets are deposited in the Short Read Archive with accession numbers: PRJNA338116 and PRJNA338115 respectively). The 'Pomeroy Brain' dataset[63] from the oncomine database (https://www.oncomine.com) was used for assessing expression level of *pitx2* and *Mll3* in human medulloblastoma versus normal cerebellum. All other data availability is either within the article (and its supplementary information) or available upon request.

Declarations

Competing interests

The authors declare they have no competing interests.

Funding

This work was funded by grants from the Medical Research Council (grant number MR/M000133/1) and the Biotechnology and Biological Sciences Research Council (grant number BB/K007564/1) to AA. A.G.L. is funded by a Human Frontier Science Program fellowship.

Authors' contributions

AA and YM conceived and designed the study. YM performed the experiments. DK performed the bioinformatic analyses. SH participated in the optimization of the ChIP-seq protocol. AGL participated in the optimization of the RNA-seq protocol. FJH performed initial work on the project, including generating the first *LPT*(RNAi) results. NK and PA helped with sigma, zeta and gamma *in situ* hybridization experiments. YM and AA wrote the manuscript.

Acknowledgements

We thank past and present members of the AA lab for comments on the manuscript.

References

1. Aboobaker, A. A. Planarian stem cells: a simple paradigm for regeneration. *Trends in Cell Biology* **21,** 304–311 (2011).
2. Rink, J. C. Stem cell systems and regeneration in planaria. *Dev Genes Evol* **223,** 67–84 (2012).
3. Onal, P. *et al.* Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *The EMBO Journal* **31,** 2755–2769 (2012).
4. Adamidi, C. *et al.* De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Research* **21,** 1193–1200 (2011).

5. Labbé, R. M. *et al.* A Comparative Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians and Mammals. *STEM CELLS* **30,** 1734–1745 (2012).
6. Solana, J. *et al.* Conserved functional antagonism of CELF and MBNL proteins controls stem cell-specific alternative splicing in planarians. *eLife* **5,** 1193 (2016).
7. Solana, J. *et al.* Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. *Genome Biol.* **13,** R19 (2012).
8. Alié, A. *et al.* The ancestral gene repertoire of animal stem cells. *Proc Natl Acad Sci USA* 201514789–8 (2015). doi:10.1073/pnas.1514789112
9. Kao, D., Mihaylova, Y., Hughes, S., Lai, A. & Aboobaker, A. Epigenetic analyses of the planarian genome reveals conservation of bivalent promoters in animal stem cells. *bioRxiv* 122135 (2017). doi:10.1101/122135
10. Shibata, N. *et al.* Inheritance of a Nuclear PIWI from Pluripotent Stem Cells by Somatic Descendants Ensures Differentiation by Silencing Transposons in Planarian. *Developmental Cell* **37,** 226–237 (2016).
11. Salvetti, A. DjPum, a homologue of Drosophila Pumilio, is essential to planarian stem cell maintenance. *Development* **132,** 1863–1874 (2005).
12. Reddien, P. W. Specialized progenitors and regeneration. *Development* **140,** 951–957 (2013).
13. Juliano, C. E., Swartz, S. Z. & Wessel, G. M. A conserved germline multipotency program. *Development* **137,** 4113–4126 (2010).
14. Jaber-Hijazi, F. *et al.* Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Developmental Biology* **384,** 141–153 (2013).
15. Scimone, M. L., Meisel, J. & Reddien, P. W. The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian Schmidtea mediterranea. *Development* **137,** 1231–1241 (2010).
16. Zhu, S. J., Hallows, S. E., Currie, K. W., Xu, C. & Pearson, B. J. A mex3 homolog is required for differentiation during planarian stem cell lineage development. *eLife* **4,** 304 (2015).
17. Cowles, M. W., Omuro, K. C., Stanley, B. N., Quintanilla, C. G. & Zayas, R. M. COE Loss-of-Function Analysis Reveals a Genetic Program Underlying Maintenance and Regeneration of the Nervous System in Planarians. *PLoS Genet* **10,** e1004746–12 (2014).
18. Barberan, S., Fraguas, S. & Cebrià, F. The EGFR signaling pathway controls gut progenitor differentiation during planarian regeneration and homeostasis. *Development* **143,** 2089–2102 (2016).
19. Guedelhoefer, O. C. & Alvarado, A. S. Amputation induces stem cell mobilization to sites of injury during planarian regeneration. *Development* **139,** 3510–3520 (2012).
20. Abnave, P. *et al.* A shielded irradiation assay to investigate mechanisms of in vivo stem cell migration in planarians. 1–42 (2016). doi:10.1101/080853
21. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419,** 624–629 (2002).
22. Villa, R. *et al.* Role of the Polycomb Repressive Complex 2 in Acute

Promyelocytic Leukemia. *Cancer Cell* **11,** 513–525 (2007).

23. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
24. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
25. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature Publishing Group* **43,** 875–878 (2011).
26. Chen, C. *et al.* MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. *Cancer Cell* **25,** 652–665 (2014).
27. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81,** 65–95 (2012).
28. Wu, M. *et al.* Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Molecular and Cellular Biology* **28,** 7337–7344 (2008).
29. Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & Development* **26,** 2604–2620 (2012).
30. Wang, P. *et al.* 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. *Molecular and Cellular Biology* **29,** 6074–6085 (2009).
31. Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology* **33,** 4745–4754 (2013).
32. Lee, J. *et al.* 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 (2009).
33. Sedkov, Y. *et al.* Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. *Nature* **426,** 78–83 (2003).
34. Cheng, J. *et al.* A Role for H3K4 Monomethylation in Gene Repression and Partitioning of Chromatin Readers. *Molecular Cell* **53,** 979–992 (2014).
35. Chauhan, C., Zraly, C. B., Parilla, M., Diaz, M. O. & Dingwall, A. K. 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 (2012).
36. Hubert, A. *et al.* Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone methyltransferases. *Epigenetics* **8,** 79–91 (2013).
37. Duncan, E. M., Chitsazan, A. D., Seidel, C. W. & Alvarado, A. S. Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *CellReports* **13,** 2741–2755 (2015).
38. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2,** 2817–25 (2013).
39. Mohan, M. *et al.* The COMPASS Family of H3K4 Methylases in Drosophila. *Molecular and Cellular Biology* **31,** 4310–4318 (2011).
40. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4

controls cell fate transition. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 11871–11876 (2016).

41. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141,** 526–537 (2014).
42. Hsieh, J. J.-D., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S. J. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Molecular and Cellular Biology* **23,** 186–194 (2003).
43. Chen, W. *et al.* Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. *Cancer Cell* **13,** 432–440 (2008).
44. Corral, J. *et al.* An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell* **85,** 853–861 (1996).
45. Thirman, M. J. *et al.* Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N. Engl. J. Med.* **329,** 909–914 (1993).
46. Sobulo, O. M. *et al.* MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc Natl Acad Sci USA* **94,** 8732–8737 (1997).
47. Goo, Y. H. *et al.* Activating Signal Cointegrator 2 Belongs to a Novel Steady-State Complex That Contains a Subset of Trithorax Group Proteins. *Molecular and Cellular Biology* **23,** 140–149 (2003).
48. Lee, S. *et al.* Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc Natl Acad Sci USA* **103,** 15392–15397 (2006).
49. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends in Biochemical Sciences* **31,** 35–40 (2006).
50. Lee, S., Lee, J., Lee, S.-K. & Lee, J. W. 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 (2008).
51. Ansari, K. I., Hussain, I., Kasiri, S. & Mandal, S. S. HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4. *Journal of Molecular Endocrinology* **48,** 61–75 (2012).
52. Guo, T., Peters, A. H. F. M. & Newmark, P. A. A bruno-like Gene Is Required for Stem Cell Maintenance in Planarians. *Developmental Cell* **11,** 159–169 (2006).
53. Wenemoser, D. & Reddien, P. W. Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Developmental Biology* **344,** 979–991 (2010).
54. Currie, K. W. & Pearson, B. J. Transcription factors lhx1/5-1 and pitx are required for the maintenance and regeneration of serotonergic neurons in planarians. *Development* **140,** 3577–3588 (2013).
55. Oviedo, N. J., Pearson, B. J., Levin, M. & Sanchez Alvarado, A. Planarian PTEN homologs regulate stem cells and regeneration through TOR signaling. *Disease Models and Mechanisms* **1,** 131–143 (2008).
56. Gonzalez-Estevez, C. *et al.* SMG-1 and mTORC1 Act Antagonistically

to Regulate Response to Injury and Growth in Planarians. *PLoS Genet* **8,** e1002619–17 (2012).

57. van Wolfswinkel, J. C., Wagner, D. E. & Reddien, P. W. Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. *Stem Cell* **15,** 326–339 (2014).
58. Zink, D., Fische, A. H. & Nickerson, J. A. Nuclear structure in cancer cells. *Nat Rev Cancer* **4,** 677–687 (2004).
59. Kim, D.-H., Kim, J. & Lee, J. W. Requirement for MLL3 in p53 Regulation of Hepatic Expression of Small Heterodimer Partner and Bile Acid Homeostasis. *Molecular Endocrinology* **25,** 2076–2083 (2011).
60. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318,** 447–450 (2007).
61. Pearson, B. J. & Alvarado, A. S. A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* **137,** 213–221 (2009).
62. März, M., Seebeck, F. & Bartscherer, K. A Pitx transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. *Development* **140,** 4499–4509 (2013).
63. Pomeroy, S. L. *et al.* Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415,** 436–442 (2002).
64. Roh, M. *et al.* Overexpression of the oncogenic kinase Pim-1 leads to genomic instability. *Cancer Res* **63,** 8079–8084 (2003).
65. Valdman, A., Fang, X., Pang, S.-T., Ekman, P. & Egevad, L. Pim-1 expression in prostatic intraepithelial neoplasia and human prostate cancer. *Prostate* **60,** 367–371 (2004).
66. Shirogane, T. *et al.* Synergistic roles for Pim-1 and c-Myc in STAT3-mediated cell cycle progression and antiapoptosis. *Immunity* **11,** 709–719 (1999).
67. Kim, J.-H. *et al.* UTX and MLL4 Coordinately Regulate Transcriptional Programs for Cell Proliferation and Invasiveness in Breast Cancer Cells. *Cancer Res* **74,** 1705–1717 (2014).
68. Zhang, Z. *et al.* Mammary-Stem-Cell-Based Somatic Mouse Models Reveal Breast Cancer Drivers Causing Cell Fate Dysregulation. *CellReports* **16,** 3146–3156 (2016).
69. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414,** 105–111 (2001).
70. Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nat Rev Cancer* **5,** 275–284 (2005).
71. Cho, Y. W. *et al.* PTIP Associates with MLL3- and MLL4-containing Histone H3 Lysine 4 Methyltransferase Complex. *Journal of Biological Chemistry* **282,** 20395–20406 (2007).
72. Cebria, F. 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 (2005).
73. Gonzalez-Estevez, C., Arseni, V., Thambyrajah, R. S., Felix, D. A. & Aboobaker, A. A. Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int. J. Dev. Biol.* **53,** 493–505 (2009).

74. Cebria, F. & Newmark, P. A. Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. *Development* **134,** 833–837 (2007).
75. Romero, B. T., Evans, D. J. & Aboobaker, A. A. in *Progenitor Cells* **916,** 167–179 (Humana Press, 2012).
76. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
77. Robb, S. M. C., Gotting, K., Ross, E. & Sánchez Alvarado, A. SmedGD 2.0: The Schmidtea mediterranea genome database. *Genesis* **53,** 535–546 (2015).
78. Attrill, H. *et al.* FlyBase: establishing a Gene Group resource for Drosophila melanogaster. *Nucleic Acids Research* **44,** D786–92 (2016).
79. Orlando, D. A. *et al.* Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. *CellReports* **9,** 1163–1170 (2014).
80. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotech* **34,** 525–527 (2016).
81. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-Seq incorporating quantification uncertainty. (2016). doi:10.1101/058164
82. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-Seq incorporating quantification uncertainty. (2016). doi:10.1101/058164

Figure legends

Figure 1. *S. mediterranea* has three partial *Mll3/4* orthologs expressed in stem cells. (**a**) A schematic depicting the structure and domain composition of MLL3/MLL4 proteins in *D. melanogaster*, *H. sapiens* and *S. mediterranea*. (**b**) *Mll3/4* genes' expression pattern in wildtype (WT) and two days following a lethal dose (60 Gy) of gamma irradiation (PI = post-irradiation). *Porcupine-1* (expressed in the irradiation-insensitive cells of the differentiated gut) and *H2B* (expressed in the irradiation-sensitive neoblasts) are used as a negative and positive control respectively. Ten worms per condition were used. (**c**) White arrows point to examples of cells double-positive for *Mll3/4* transcripts and *H2B* transcripts. The schematic shows the body area imaged. (**d**) Graph showing the raw cell counts used for percentage estimates in (**c**). Green colour represents all counted *H2B*-positive cells, yellow represents *H2B*-positive cells also expressing a *Mll3/4* ortholog. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition were used. (**e**)

Expression profiles of *Mll3/4* genes according to RNA-seq data from FACS-sorted X1 (stem cells in G2/M phase), X2 (stem cells in G1 and stem cell progeny) and X ins (differentiated cells) planarian cell populations.

Figure 2. *LPT*(RNAi) results in differentiation defects and outgrowth formation during regeneration. (**a**) A schematic showing the amputation of RNAi worms into head (H), middle (M) and tail (T) pieces in order to observe regeneration of different structures. The time-course of all the experiments on *Mll3/4* knockdown animals is depicted underneath the worm schematic. A total of 9 days of dsRNA microinjection-mediated RNAi was followed by amputation on the 10th day and subsequent observation of regeneration. (**b**) Head, middle and tail pieces following *LPT*(RNAi) or control *GFP*(RNAi) at day 8 of regeneration. Yellow arrows point towards the smaller blastema and the eye formation defects. (**c**) Head, middle and tail pieces following *LPT*(RNAi) or control *GFP*(RNAi) at day 10 of regeneration. Red arrows point towards outgrowths. (**d**) Head, middle and tail pieces following *LPT*(RNAi) or control *GFP*(RNAi) at day 14 of regeneration. Red arrows point towards outgrowths. (**e**) Gut regeneration and maintenance in middle pieces following *LPT*(RNAi), as illustrated by RNA probe for the gene *porcupine-1* at 8 days of regeneration. (**f**) Brain regeneration in middle pieces at 8 days post-amputation following *LPT*(RNAi), as illustrated by anti-SYNORF-1 antibody labeling the central nervous system (CNS). (**g**) Optic chiasma recovery in tail pieces at 8 days of regeneration following *LPT*(RNAi), as shown by anti-VC-1 antibody. (**h**) Recovery of optic cups and organized trail of optic cup precursor cells in tail pieces at 8 days of regeneration following *LPT*(RNAi), as demonstrated by RNA probe for *SP6-9*. (**i**) Pharynx recovery in head pieces at 8 days of regeneration following *LPT*(RNAi), as illustrated by RNA probe for *laminin*. Images in (**e**,**f**, **i**) are representative of two separate experiments using 10 animals per condition each. Images in (**g**, **h**) were obtained from one experiment each, using 10 animals per condition. Numbers at the top right of each regenerating piece represent number of animals in the condition showing the same phenotypic features as the animal in the panel.

Figure 3. LPT controls differentiation across neuronal and epidermal lineages. Quantification of the number of GABAergic neurons (labeled by *GAD*) (**a**), dopaminergic neurons (labeled by *TH*) (**b**), acetylcholinergic neurons (labeled by *chat*) (**c**), serotonergic neurons (labeled by *TPH*) (**d**) and early (labeled by *NB.21.11e*) and late (labeled by *AGAT-1*) epidermal stem cell progeny (**e**) at 8 days of regeneration of tail or middle pieces following *LPT*(RNAi). For each of the comparisons in this figure a 2-tailed ttest assuming unequal variance was used; a single asterisk indicates $p<0.05$, while three asterisks indicate $p<0.001$. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition per experiment were assessed over the course of two separate experiments.

Figure 4. Over-proliferation and mitotic cell clustering precedes and accompanies the emergence of outgrowths in *LPT*(RNAi) regenerating animals. (**a**) Quantification of mitotic cell numbers (mitotic cells labeled by anti-H3P antibody) at different post-amputation timepoints following *LPT*(RNAi). The figure is a representative of three repeats of the same experiment, each using 10 animals per timepoint. The statistical test used was a 2-tailed ttest assuming unequal variance. The asterisks indicate $p<0.05$. Error bars represent Standard Error of the Mean (SEM). (**b**) Examples of middle pieces at the timepoints post-amputation showing significant difference in mitotic cell counts according to (**a**). 'ph' indicates where the pharynx is in each piece. The red arrows point towards clusters of mitotic cells in late stage regenerates (192 hrs/8 days). (**c**) Brightfield examples of middle pieces forming outgrowths at timepoints after the observation of mitotic clusters in (**b**). Red arrows point towards outgrowths. (**d**) *LPT*(RNAi) head pieces that do not contain an outgrowth, but show mitotic cell clusters (indicated by yellow arrows). One piece shows staining in the eye region, which is an artifact of the procedure. The lower panel shows a mitotic cell at the border of an outgrowth and the body in a head piece at 10 days of regeneration following *LPT*(RNAi). Red arrow points towards the outgrowth and yellow arrow points towards the mitotic cell. (**e**) An example of a tail piece at 12 days of regeneration having a

mitotic cell-rich cephalic outgrowth following *LPT*(RNAi). Yellow arrows show the mitotic cells in the outgrowth.

Figure 5. Stem cells at different stages of commitment are found in outgrowths of *LPT*(RNAi) regenerating animals. (**a**) A head piece at 18 days of regeneration following *LPT*(RNAi) showing *sigma* stem cells in its posterior outgrowth. *Sigma* stem cells are double positive for *Smedwi-1* and the '*sigma pool*' of RNA probes (*Soxp1*, *Soxp2*). Red arrows in the brightfield images panel point towards the outgrowths and white arrows in the zoomed-in panel show a double-positive cell and a *Smedwi-1* single-positive cell. (**b**) A head piece at 10 days of regeneration following *LPT*(RNAi) showing *zeta* stem cells in its posterior outgrowth. *Zeta* stem cells are double positive for *Smedwi-1* and the '*zeta pool*' of RNA probes (*zfp-1*, *Soxp3, egr-1*). The red arrow in the brightfield images panel points towards the outgrowth and the white arrows in the zoomed-in panel show double-positive cells. (**c**) A middle piece at 11 days of regeneration following *LPT*(RNAi) showing *gamma* stem cells in its lateral outgrowth. *Gamma* stem cells are double positive for *Smedwi-1* and the '*gamma pool*' of RNA probes (*gata4/5/6*, *hnf4*). The red arrow in the brightfield images panel points towards the outgrowth and the white arrows in the zoomed-in panel show a double-positive cell, a *Smedwi-1* single-positive cell and a *gamma pool* single-positive cell. In (**a**), (**b**) and (**c**) the normal tissue margin is shown via white dashed lines.

Figure 6. *LPT*(RNAi) results in a cancer-like phenotype. A summary of the differentiation and neoblast proliferation data presented, together with a simplified flowchart illustrating the tested lineages' development under knockdown conditions. A red cross sign indicates where the defect in a lineage is detected following *LPT*(RNAi).

Figure 7. RNA-seq of G2/M stem cells following *LPT*(RNAi) reveals effects on genes enriched in different cell populations. (**a**) Genes were

classified according to their proportional expression in the X1 (G2/M stem cells; dark blue), X2 (G1 stem cells and stem cell progeny; light blue) and X ins (differentiated cells; orange) FACS populations of cells. Genes were defined as enriched in certain population(s) if more that 50% of their expression is observed in that population in wildtype or more than 75% of their expression is observed across two cell populations. Genes not enriched in either population were classified as 'not enriched'. Each vertical line represents a gene. Under the population expression enrichment track is a track with all the significantly up- and down-regulated genes in G2/M stem cells following *LPT*(RNAi). The genes with fold change >1.5 ($p<0.05$) are shown in red following a log2 fold change transformation. The genes with fold change <-1.5 ($p<0.05$) are shown in blue following a log2 fold change transformation. The Wald's test (as part of the Sleuth software) was used for assessing differential expression. (**b**) Enrichment for genes in each of the six classes was calculated for the up- and down-regulated genes' list (red and blue respectively). Enrichment for Transcription Factors (TFs) was also performed. The number of genes in each group is indicated in brackets under the group's name. Numbers in white represent significant enrichment ($p<0.01$) according to a hypergeometric enrichment test. (**c**) Gene Ontology (GO) enrichment analysis on the genes significantly up-regulated (red) and down-regulated (blue) in G2/M stem cells following *LPT*(RNAi). Categories are sorted by average Log2 fold change of the up- or down-regulated genes falling in each category. In bold are shown terms that relate to the described *Mll3/4* loss of function phenotype.

Figure 8. *LPT*(RNAi) is mainly manifested in changes in H3K4me1 and H3K4me3 around the TSS in G2/M stem cells. (**a**) Graphs presenting the average read coverage across the genome for H3K4me3, H3K4me1 and H3K27me3 (centered on the TSS, showing 2 kb upstream and downstream) normalised to *Drosophila* S2 signal spike-in. The input coverage is subtracted. Log2 fold change graphs are also shown for each histone modification, where signal above zero shows increase following *LPT*(RNAi) and signal below zero represents a decrease. Three colours are used for different gene classes –

dark blue (genes enriched in G2/M stem cells (X1)), light blue (genes enriched in G1 stem cells and stem cell progeny (X2)), orange (genes enriched in differentiated cells (X ins)). Standard deviation is shown by a faded colour around each line. (**b**) Spearman's rank correlation between changes in RNA-seq signal and H3K4me1 or H3K4me3 ChIP-seq signal for the region around the TSS of genes from different enrichment classes (only examples where a significant correlation exists are shown). The blue line represents a correlation where no filter for fold change in the RNA-seq data was applied. The green line shows a correlation where RNA-seq fold change data was filtered for Log2 fold changes =<-1 and >= +1. Faded areas of the lines represent results not significant at p<0.001, while darker colours represent results significant at p<0.001.

Figure 9. *LPT* regulates the expression of known and putative oncogenes and tumor suppressors. (**a**) Examples of genes significantly (p<0.05) mis-regulated in G2/M stem cells following *LPT*(RNAi). RNA-seq fold change is shown in red (up-regulation) and blue (down-regulation). The genes' enrichment class is also shown. The ChIP-seq profile for H3K4me3 in the 2 kb region around the TSS of each gene is presented. Purple colour represents normalised signal following *LPT*(RNAi) and green colour is used to show the normalised signal following *GFP*(RNAi). 'TF' stands for 'transcription factor'. (**b**) *in silico* analysis (www.oncomine.org; ttest, p<0.0001) of *Mll3* and *pitx2* expression in normal tissue (cerebellum) and cancer tissue (medulloblastoma). (**c**) *pitx* and *Smedwi-1 in situ* hybridization at 8 days of regeneration of middle pieces following *LPT*(RNAi). White arrows show double-positive cells. Cell counts are compared using a 2-tailed ttest assuming unequal variance. The asterisk indicates p<0.05. Ten animals per condition were used.

Figure 10. Double knockdown with *utx* or *pim-1-like* alleviates the *LPT*(RNAi) over-proliferation and outgrowth phenotype. (**a**) More examples of genes significantly (p<0.05) mis-regulated in G2/M stem cells

following *LPT*(RNAi). RNA-seq fold change is shown in red (up-regulation) and blue (down-regulation). The genes' enrichment class is also shown. The ChIP-seq profile for a histone modification in the 2 kb region around the TSS of each gene is presented. Purple colour represents normalised signal following *LPT*(RNAi) and green colour is used to show the normalised signal following *GFP*(RNAi). Depending on the gene enrichment class, H3K4me1 or H3K4me3 ChIP-seq signal is presented for each gene (based on previous Spearman's rank correlation analyses in **Figure 8**). Bold font of a gene name illustrates an example where there is a correlation between ChIP-seq and RNA-seq data. (**b**) Mean average mitotic cell (labeled by anti-H3P antibody) counts at 48 hours post-amputation following double knockdown experiments. Ten animals per condition were used. The statistical test used was a 2-tailed ttest assuming unequal variance. The asterisks indicate significant differences at $p<0.05$. Error bars represent Standard Error of the Mean (SEM). Representative tail piece examples are shown for each condition significantly different from the control *GFP/GFP*(RNAi) animals or from the *GFP/LPT*(RNAi) condition. (**c**) Percentage quantification of double knockdown regenerates developing outgrowths.

Additional File Legends

Additional File 1 (PDF) Structure and function of COMPASS and COMPASS-like core proteins. (**a**) Schematics of the core subunits of the COMPASS and the two COMPASS-like complexes in mammals are presented with coloured boxes corresponding to different protein domains – RRM1 (RNA-recognition motif), N-SET, SET, CXXC (zinc finger), PHD (Plant Homeodomain fingers), zf (PHD-like zinc finger), FYRN (Phenylalanine/Tyrosine rich N-terminus domain), FYRC (Phenylalanine/Tyrosine rich C-terminus domain), purple stars signifying nuclear receptor recognition motifs. Dashed vertical line represents proteolytic cleavage. (**b**) As in (**a**), but in fruitfly. (**c**) Proposed mechanisms of action of each core complex subunit. COMPASS complex – 1) performing H3K4 trimethylation on TSS of most actively transcribed genes and 2) depositing

H3K4me2 on the gene bodies of actively transcribed genes. MLL1/2/Trithorax COMPASS-like complex – 1) a role in transcriptional activation of Hox genes via trimethylating H3K4 on TSS of their promoters and 2) MLL2 is involved in trimethylation of H3K4 on TSS of bivalent promoters. MLL3/4/LPT/Trr – 1) role in hormone-dependent transcription – when the Nuclear Receptor protein (NR) is bound to the DNA Hormone Response Element (HRE) upon Hormone Ligand (HL) detection, MLL3/4/LPT/Trr complex binds the nuclear receptor and serves as its co-activator via trimethylating H3K4 and promoting active transcription on selected loci; 2) a switch between inactive and active enhancer states where MLL3/4/LPT/Trr complex deposits H3K4me1 on both active and inactive enhancers; upon UTX recruitment, it demethylates H3K27me3 and allows for CBP/p300 to acetylate H3K27 and activate the enhancer; 3) a switch between active and inactive promoters - MLL3/4/LPT/Trr complex bound to TSS deposits H3K4me1 on the TSS and around it, leads to repressed transcription of the gene; when H3K4me1 is depleted from the TSS and another complex performs trimethylation of H3K4 on TSS, this is correlated with activated transcription. (**d**) Schematic representation of planarian COMPASS and COMPASS-like core subunits. SMED-LPT (in red) is characterized in the present study. (**e**) Planarian COMPASS and COMPASS-like core subunits' expression in the three populations of cells sortable by fluorescence-activated cell sorting (FACS) (X1=G2/M stem cells, X2=G1 stem cells and stem cell progeny, X ins=differentiated cells) according to RNA-seq data. (**f**) Known defects after RNAi-mediated knockdown of core COMPASS and COMPASS-like subunits in planarians.

Additional File 2 (PDF) Planarian *Mll3/4* genes are expressed in neoblasts and neoblast progeny and colocalise with each other. (**a**) Protein alignment of conserved regions of COMPASS-like families' core proteins. Asterisks indicate complete conservation in all sequences, while black boxes are drawn around areas of conservation specific to the MLL3/4/Trithorax-related family. Colours represent similarity of amino acids. The image was produced using MEGA.5.2 software. (**b**) *LPT*, *trr-1* and *trr-2*

expression in wildtype and irradiated 3 day-regenerating head and middle pieces. Arrows in head pieces point towards expression in the forming pharynx, while arrows in middle pieces point towards expression in the forming brain. (**c**) *LPT*, *trr-1* and *trr-2* expression in intact animals following *GFP*(RNAi) or following *H2B*(RNAi). *Porcupine-1* and *Smedwi-2* were used as a negative and positive control respectively. (**d**) *LPT*, *trr-1* and *trr-2* co-expression in the head region (as shown by the schematics). White arrows point towards cells showing colocalisation.

Additional File 3 (PDF) Phenotype scoring of *Mll3/4* knockdown planarians during regeneration and homeostasis. (**a**) Proportion of head, middle and tail regenerates exhibiting particular phenotypic characteristics following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**b**) Survival curves for head, middle and tail pieces following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**c**) Proportion of intact (homeostatic) animals with particular phenotypic characteristics following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**d**) Survival curves for homeostatic animals following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi).

Additional File 4 (PDF) *Trr-1*(RNAi) and *trr-2*(RNAi) lead to mild differentiation defects during regeneration. (**a**) A schematic showing the amputation of RNAi worms into head (H), middle (M) and tail (T) pieces in order to observe regeneration of different structures. The time-course of the experiments on *Mll3/4* knockdown animals is depicted underneath the worm schematic. A total of 9 days of dsRNA microinjection-mediated RNAi was followed by amputation on the 10th day and observation of regeneration. (**b**) Head, middle and tail pieces following *trr-1*(RNAi), *trr-2*(RNAi) or control *GFP*(RNAi) at day 8 of regeneration. Yellow arrows point towards the regenerative defects – smaller blastema, delayed eye formation or posterior bloating. (**c**) Head, middle and tail pieces following *trr-1*(RNAi), *trr-2*(RNAi) or control *GFP*(RNAi) at day 14 of regeneration. (**d**) Central nervous system (CNS) maintenance and recovery at 8 days of middle piece regeneration, as

labeled by CNS-specific anti-SYNORF-1 antibody, following *trr-1*(RNAi) or *trr-2*(RNAi). (**e**) Gut maintenance and recovery at 8 days of middle piece regeneration, as labeled by *porcupine-1*, following *trr-1*(RNAi) or *trr-2*(RNAi). (**f**) Pharynx recovery at 8 days of head piece regeneration, as labeled by *laminin*, following *trr-1*(RNAi) or *trr-2*(RNAi). Numbers at the top of each piece represent number of animals in that condition showing the same phenotypic features as the animal in the panel.

Additional File 5 (PDF) *Trr-1/ trr-2* double knockdown results in more prevalent and accelerated outgrowth formation compared to *LPT*(RNAi). (**a**) Head, middle and tail pieces at 3 days of regeneration following *GFP/LPT*(RNAi), *GFP/trr-1*(RNAi), *GFP/trr-2*(RNAi), *trr-1/trr-2*(RNAi) and *GFP/GFP*(RNAi). Red arrows point towards outgrowths. (**b**) Percentage of head, middle and tail regenerating pieces developing outgrowths throughout their life-time. (**c**) Survival curves of head, middle and tail regenerating pieces. The *GFP/GFP*(RNAi) line overlaps with *GFP/trr-1*(RNAi) and *GFP/ trr-2*(RNAi). Ten animals per condition were used.

Additional File 6 (PDF) Number if stem cells and early stem cell progeny is unchanged following knockdown of *LPT*. The pre-pharyngeal area of middle pieces at 8 days of regeneration was used for this experiment. Stem cells are labeled with *H2B* and early neoblast progeny cells are *H2B*-negative/anti-SMEDWI-1 antibody-positive cells. Numbers of stem cells and progeny cells between *LPT*(RNAi) animals and controls were not significantly different (ns). A 2-tailed ttest assuming unequal variance was used. Ten animals per condition were processed.

Additional File 7 (PDF) *Trr-2*(RNAi) regenerating animals produce less GABAergic and dopaminergic neurons. Quantification of the number of GABAergic neurons (labeled by *GAD*) (**a**), dopaminergic neurons (labeled by *TH*) (**b**), serotonergic neurons (labeled by *TPH*) (**c**), acetylcholinergic neurons

(labeled by *chat*) (**d**) and early (labeled by *NB.21.11e*) and late (labeled by *AGAT-1*) epidermal stem cell progeny (**e**) at 8 days of regeneration of tail or middle pieces following *trr-1*(RNAi) or *trr-2*(RNAi). For each of the comparisons in this figure a 2-tailed ttest assuming unequal variance was used; a single asterisk indicates $p<0.05$. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition per experiment were assessed over the course of two separate experiments.

Additional File 8. (PDF) Tubule-associated protonephridia and cilia cell regeneration is not affected by *Mll3/4* knockdown. (**a**) Recovery of dorsal and ventral cilia in middle pieces at 8 days of regeneration following *LPT*(RNAi), *trr-1*(RNAi) or *trr-2*(RNAi), as labeled by anti-acetylated tubulin antibody. (**b**) Recovery and maintenance of tubule-associated protonephridia cells (labeled by *CAVII-1*) in middle pieces at 8 days of regeneration following *LPT*(RNAi), *trr-1*(RNAi) or *trr-2*(RNAi). The white dashed line indicates where the anterior regenerated body parts following amputation should be. The graph compares numbers of *CAVII-1*-positive cells in the newly recovered regions, as well as in the whole body, between *Mll3/4* genes knockdown conditions and controls. For each of the comparisons in this figure a 2-tailed ttest assuming unequal variance was used; 'ns' stands for 'not significant'. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition were used.

Additional File 9. (PDF) 'Inchworming' following *Mll3/4* knockdown is due to serotonin deficiency. Locomotive defects in *Mll3/4* knockdown animals are shown. The number of regenerating head pieces per condition exhibiting the respective locomotive defect before and after 45-minute-long serotonin hydrochloride treatment is shown in a table. Movie still shots are shown for worms in each treatment. There was one-second interval between the chosen still shots. The dashed line indicates the locomotive progress that the worm had achieved in 3 seconds (from first to last shot). Steeper line indicates faster movement. White numbers at the bottom right corners

represent number of animals per condition showing the illustrated behaviour. Ten animals per condition were used.

Additional File 10. (PDF) *Mll3/4* knockdown leads to changes in mitotic activity during regeneration and homeostasis. (**a**) Mitotic cell number fluctuations during regeneration following *trr-1*(RNAi), *trr-2*(RNAi) and *GFP*(RNAi). (**b**) Mitotic cell number fluctuations during homeostatic observations following *LPT*(RNAi), *trr-1*(RNAi), *trr-2*(RNAi) and *GFP*(RNAi). For each of the comparisons in this figure a 2-tailed ttest assuming unequal variance was used; a single asterisk indicates $p<0.05$. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition were assessed.

Additional File 11. (PDF) Non-mitotic stem cells are present in outgrowths of animals following *LPT*(RNAi). A head piece (containing outgrowths) at 10 days of regeneration following *LPT*(RNAi) stained with *H2B* RNA probe and anti-H3P mitotic cell antibody. Stem cells found outside the usual stem cell compartment are indicated via white arrows. The border of the usual neoblast compartment is indicated with a white dashed line.

Additional File 12. (PDF) *Sigma*, *zeta* and *gamma* neoblast numbers are unchanged following *LPT*(RNAi). (**a, d**) Cells in pre-pharyngeal regions of middle pieces at 8 days of regeneration following *LPT*(RNAi) labeled by the *sigma pool* of RNA probes (*Soxp1*, *Soxp2*) and *Smedwi-1*. White arrows point towards *sigma* neoblasts (double-positive for *sigma pool* and *Smedwi-1*). (**b**, **e**) Cells in pre-pharyngeal regions of middle pieces at 8 days of regeneration following *LPT*(RNAi) labeled by the *zeta pool* of RNA probes (*zfp-1*, *Soxp3, egr-1*) and *Smedwi-1*. White arrows point towards *zeta* neoblasts (double-positive for *zeta pool* and *Smedwi-1*). (**c**, **f**) Cells in pre-pharyngeal regions of middle pieces at 8 days of regeneration following *LPT*(RNAi) labeled by the *gamma pool* of RNA probes (*gata4/5/6*, *hnf4*) and *Smedwi-1*. White arrows

point towards *gamma* neoblasts (double-positive for *gamma pool* and *Smedwi-1*). (**g**) Overall number of *Smedwi-1*-positive cells (regardless of colocalisation with other markers) in the pre-pharyngeal region of middle pieces at 8 days of regeneration following *LPT*(RNAi). The statistical comparisons in this figure were performed via 2-tailed ttest assuming unequal variance. 'ns' stands for 'not significant'. Ten worms per condition were processed.

Additional File 13. (PDF) *LPT*(RNAi) results in disorganized outgrowth-focused expression of epidermal precursor markers, epithelial disarray and hypertrophy and changes of nuclear morphology. (**a**) Anterior part (containing an outgrowth) of a tail piece at 18 days of regeneration following *LPT*(RNAi) labeled with *NB.21.11e* and *AGAT-1* epidermal precursor markers. 'CG' stands for 'cephalic ganglia'. (**b**) The epidermal layer (stained with Hoechst 33342) of a tail piece at 10 days of regeneration following *LPT*(RNAi) compared to control (**c**) The nuclear area of 20 epithelial cells per experimental and control condition was compared via 2-tailed ttest assuming unequal variance. Triple asterisk indicates $p<0.001$. (**d**) Nuclear morphology comparison between a 10-day *LPT*(RNAi) and control regenerate. Samples were stained with Hoechst 33342. Yellow arrows point towards misshapen nuclei.

Additional File 14. (PDF) Histone modification ChIP-seq profiles at promoter-proximal regions of different classes of genes. (**a**), (**b**) and (**c**) show histone modification patterns for H3K4me3, H3K4me1 and H3K27me3 respectively. ChIP-seq signal is shown in black. Six groups of genes are presented – enriched >50% in X1 (G2/M stem cells) shown by dark blue, enriched >50% in X2 (G1 stem cells and stem cell progeny) shown in light blue, enriched >50% in X ins (differentiated cells) shown in orange, genes enriched >75% in X1/X2, >75% in X2/X ins and 'not enriched'. Histone modification graphs are centered on the Transcriptional Start Site (TSS) with

2.5 kb shown upstream and downstream.

Additional File 15. (PDF) The expression of stem cell-enriched genes mis-regulated following *LPT*(RNAi) is inversely correlated with H3K4me1 TSS-proximal levels. (**a**) Graphs presenting the average read coverage across the genome for H3K4me3, H3K4me1 and H3K27me3 (centered on the TSS, showing 2 kb upstream and downstream) normalised to *Drosophila* S2 signal spike-in. The input coverage is subtracted. Log2 fold change graphs are also shown for each histone modification, where signal above zero shows increase following *LPT*(RNAi) and signal below zero represents a decrease. Three colours are used for different gene classes – green (genes enriched in X1/X2 cells), red (genes enriched in X2/X ins cells), black (genes not enriched in any population of cells). Standard deviation is shown as a faded colour around each line. (**b**) Log2 fold change of signal around the TSS across different histone marks and gene classes following *LPT*(RNAi). Blue represents genes down-regulated following *LPT*(RNAi) and red – up-regulated. Standard deviation is shown by a faded colour around each line.

Additional File 16. (PDF) *LPT* regulates the expression of cancer- and development-associated genes. Examples of genes significantly ($p<0.05$) mis-regulated in G2/M stem cells following *LPT*(RNAi). RNA-seq fold change is shown in red (up-regulation) and blue (down-regulation). The genes' enrichment class is also shown. The ChIP-seq profile for a histone modification in the 2 kb region around the TSS of each gene is presented. Purple colour represents normalised signal following *LPT*(RNAi) and green colour is used to show the normalised signal following *GFP*(RNAi). Depending on the gene enrichment class, H3K4me1 (**a**) or H3K4me3 (**b**) ChIP-seq signal is presented for each gene (based on previous Spearman's rank correlation analyses in **Figure 8**). Bold font of a gene name illustrates an example where there is a correlation between ChIP-seq and RNA-seq data. 'TF' stands for 'transcription factor'.

Additional File 17. (PDF) Phenotype scoring of double knockdown (*GFP/LPT, GFP/pim-1, GFP/pim-1-like, GFP/utx, pim-1/pim-1-like, LPT/pim-1, LPT/pim-1-like, LPT/utx* and *GFP/GFP*) planarians during regeneration. Proportion of head, middle and tail regenerates exhibiting particular phenotypic characteristics following a double knockdown. The proportion of animals forming outgrowths is given regardless of regenerating piece identity.

Additional File 18. (PDF) The histone modifications antibodies used for ChIP-seq experiments are specific. (**a**) Western blot with loading control anti-H3 (unmodified histone H3) and anti-H3K4me1 on whole animal protein lysate from *GFP*(RNAi) and *LPT*(RNAi) samples. (**b**) Western blot with loading control anti-H3 (unmodified histone H3) and anti-H3K4me3 on whole animal protein lysate from *GFP*(RNAi) and *set1*(RNAi) samples. (**c**) Western blot with loading control anti-H3 (unmodified histone H3) and anti-H3K27me3 on whole animal protein lysate from *GFP*(RNAi) and *utx(RNAi)* samples.

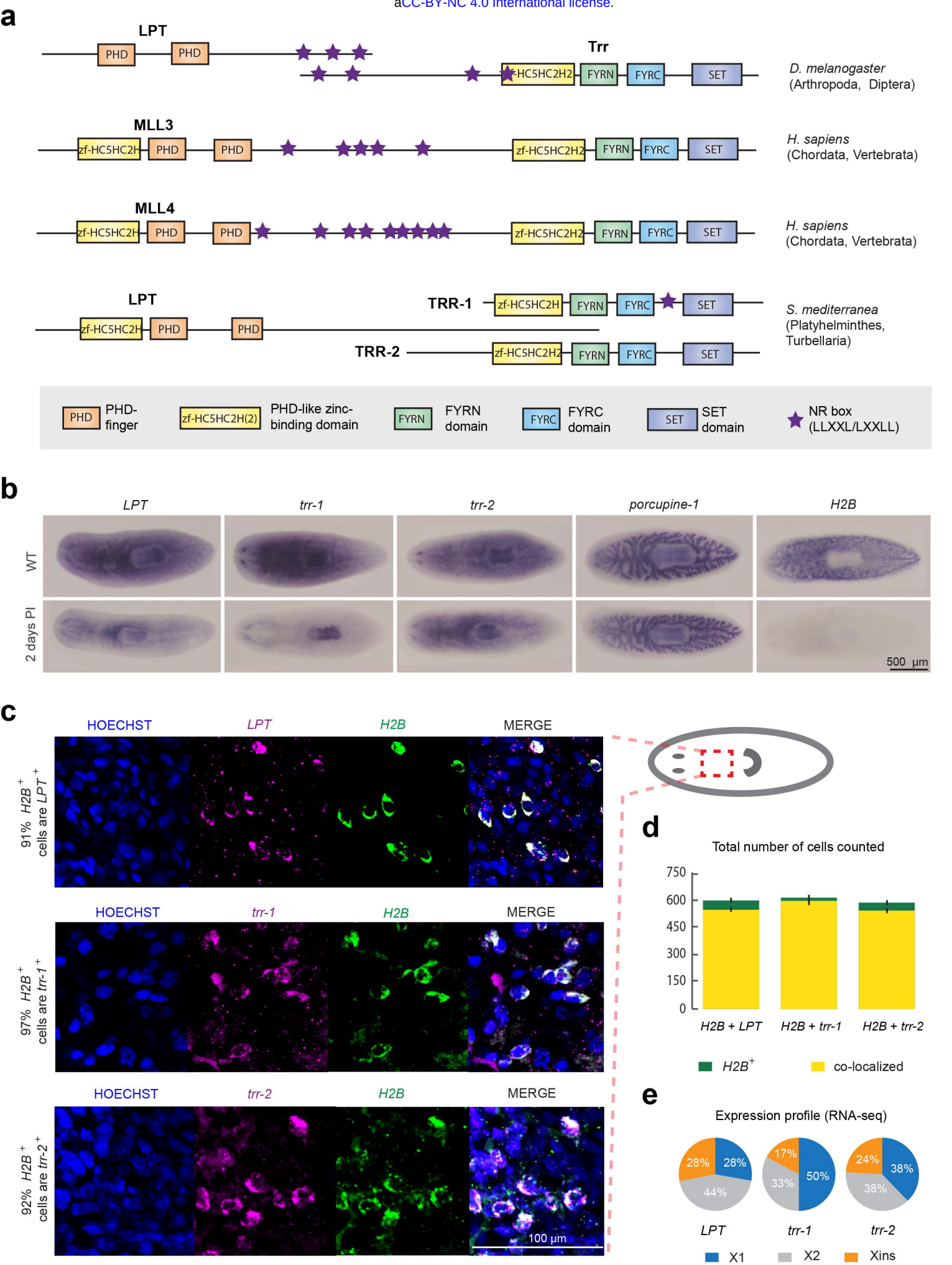
Additional file19. (PDF) Summary of planarian ChIP-seq procedure. Three day-regenerating planarians were dissociated into single cells. Cells were stained with Hoechst 34580 and Calcein AM in order to visualize cell populations according to nuclear size and cytoplasmic complexity. The X1 (G2/M) stem cells (magenta) were sorted and mixed with 4% *Drosophila* S2 cells. Cells were crosslinked with 1% Formaldehyde and sonicated. Immunoprecipitation with anti-H3K4me3, anti-H3K4me1 and anti-H3K27me3 antibodies followed. Samples were reverse-crosslinked and libraries were prepared using NEBNext Ultra II library preparation kit.

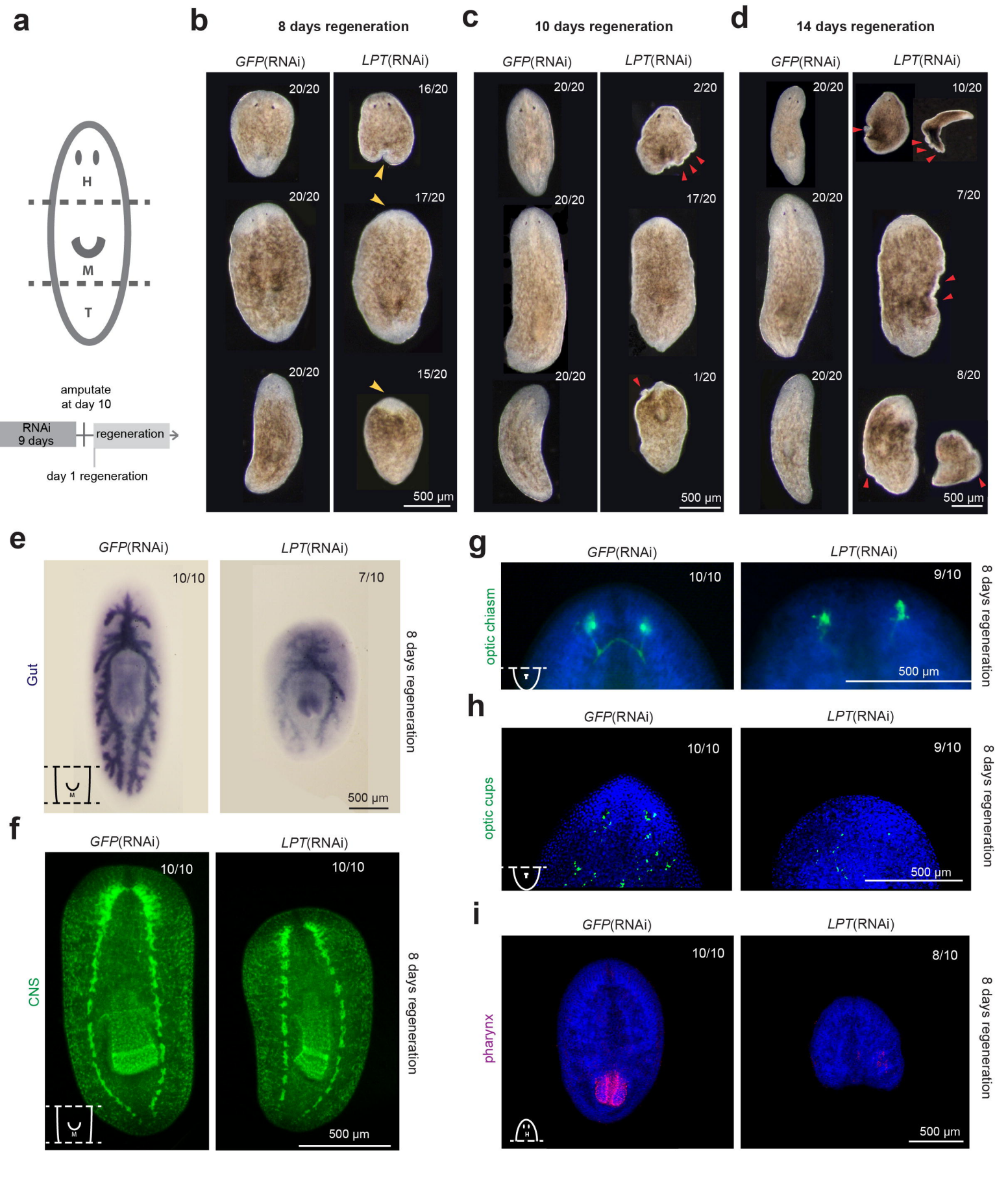
Additional File 20. (.xlsx) **Differentially expressed loci following**

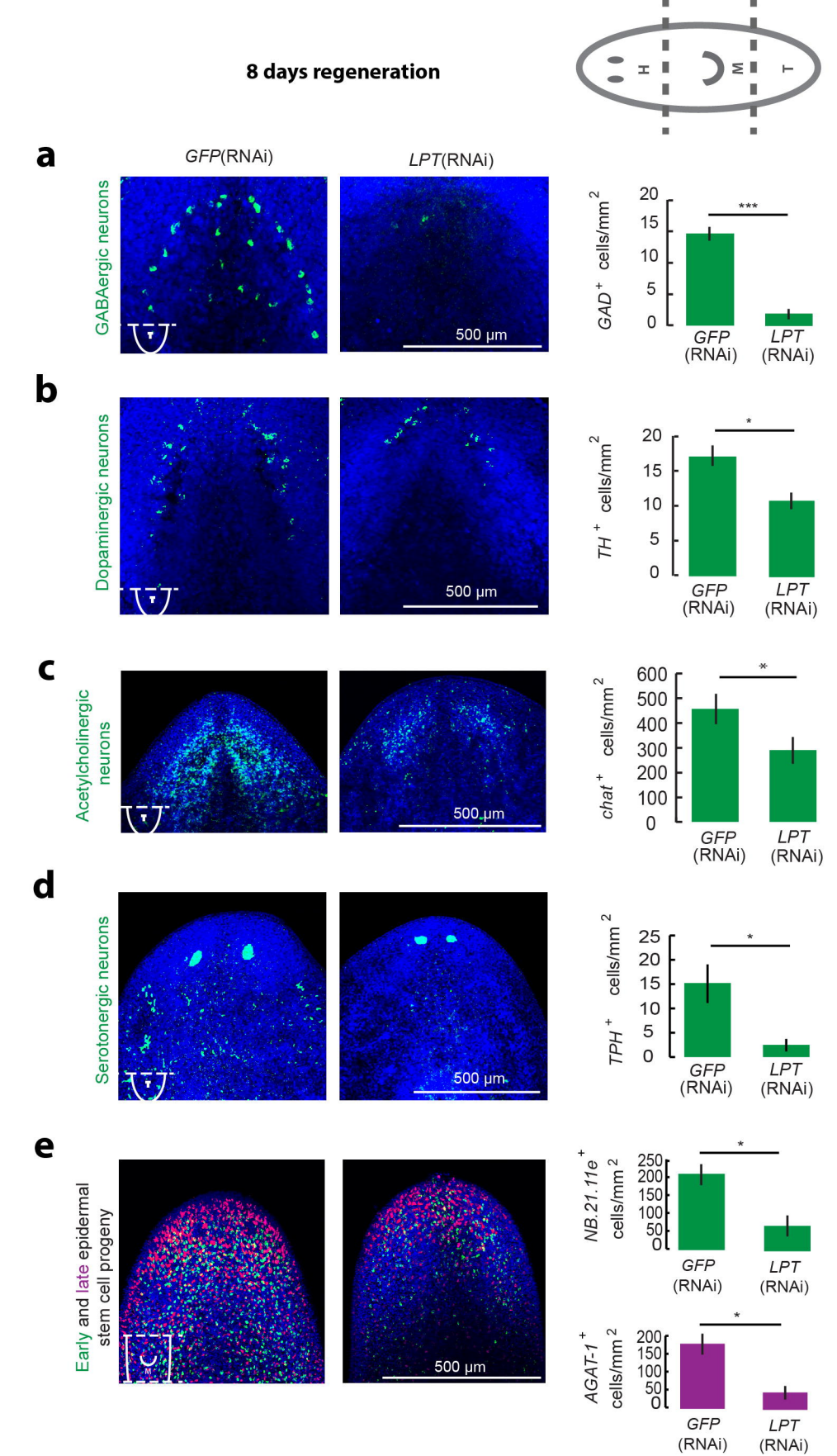
***LPT*(RNAi).** Each row represents one locus that was differentially expressed with a p-value less than 0.05 and fold change <-1.5 or >1.5. The Wald's test (as part of the Sleuth software) was used for assessing differential expression. The top BLAST hit (with e-value) and the common model organism top BLAST hit is also provided for each locus.

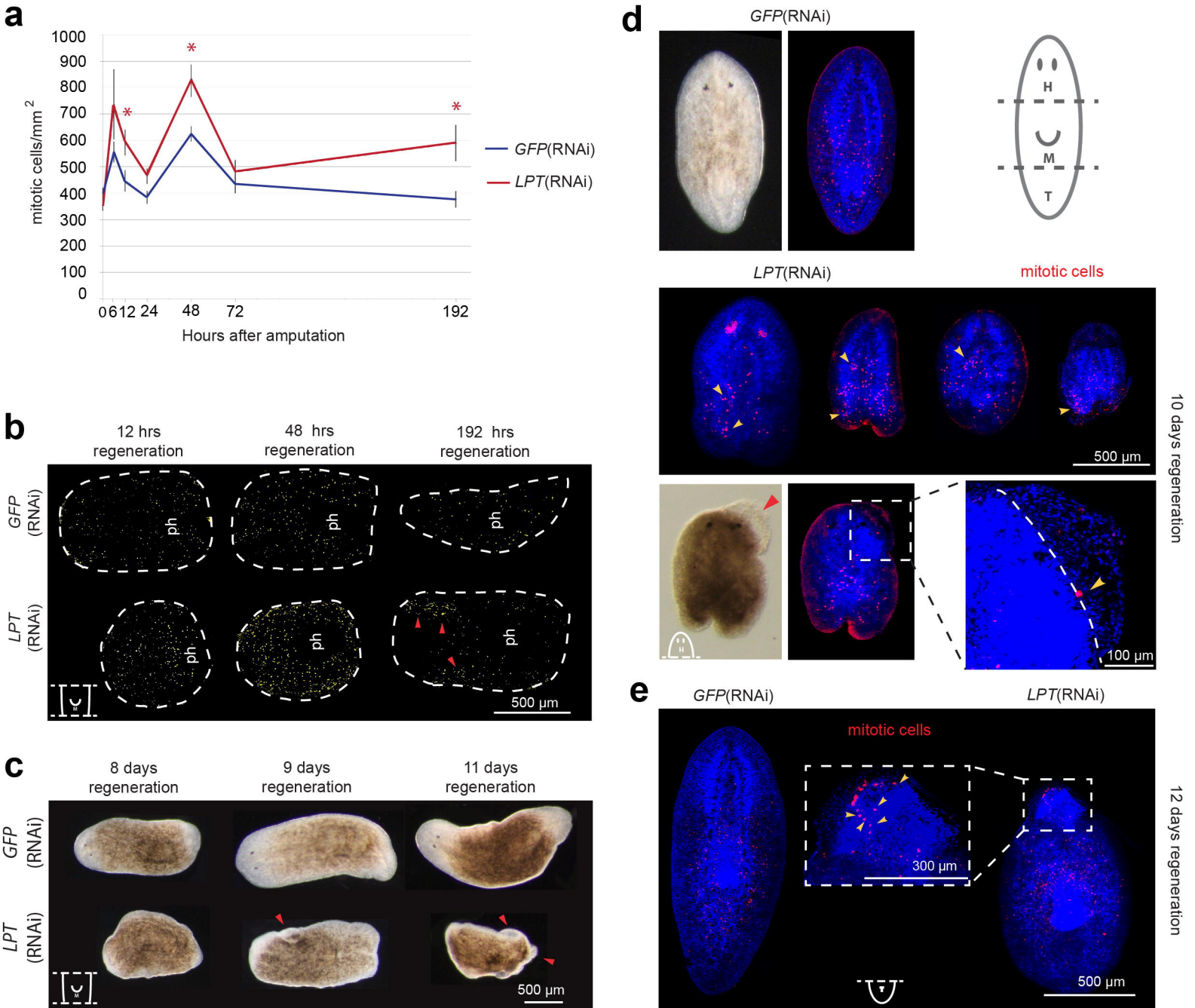
Additional File 21. (.xlsx) **Primer sequences.** All primers are given in 5'->3' orientation. 'F' and 'R' stand for 'forward' and 'reverse' primer respectively.

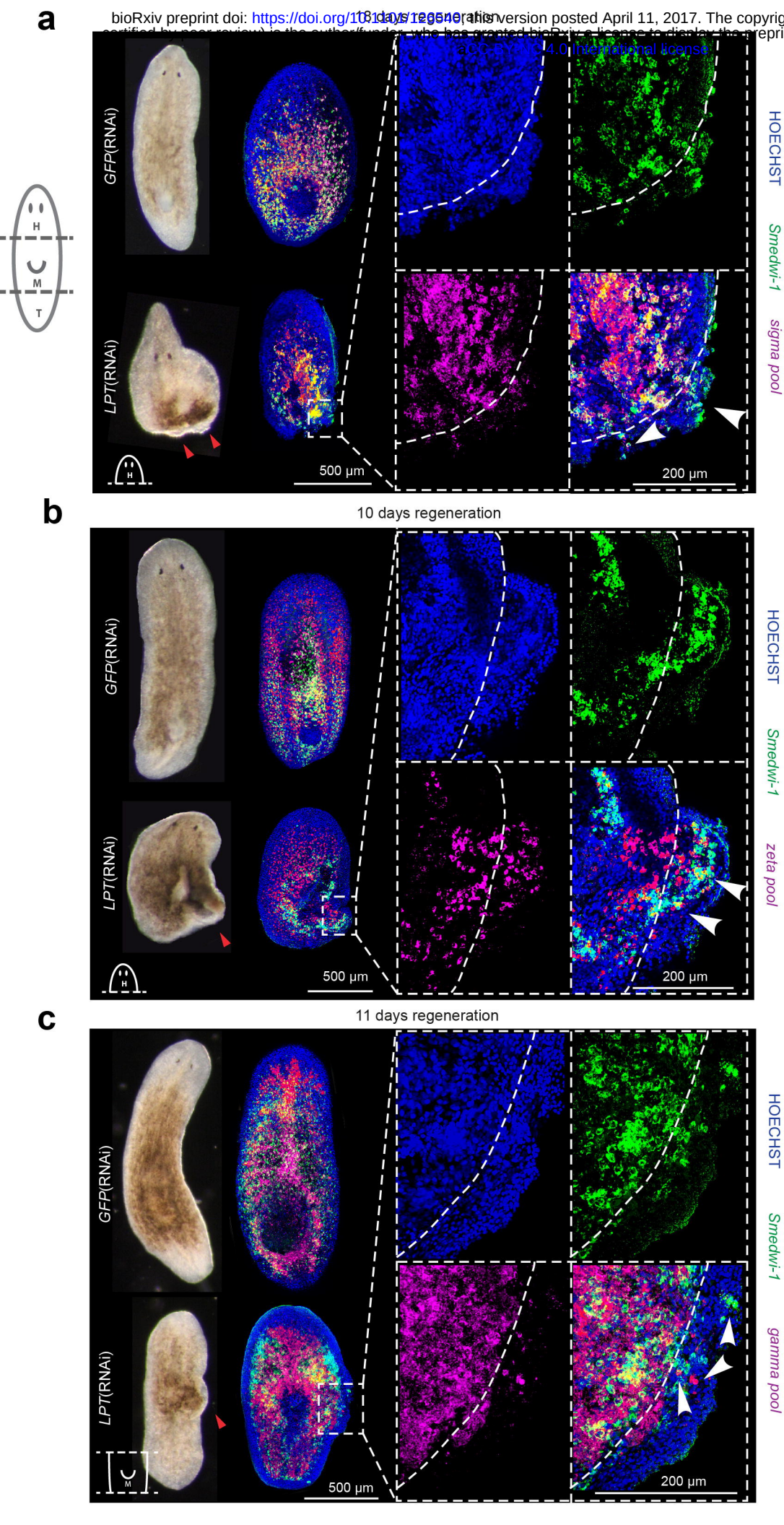
Additional File 22 (html) Supplementary Python Notebook**.** Provides details on the ChIP-seq and RNA-seq bioinformatic analyses.

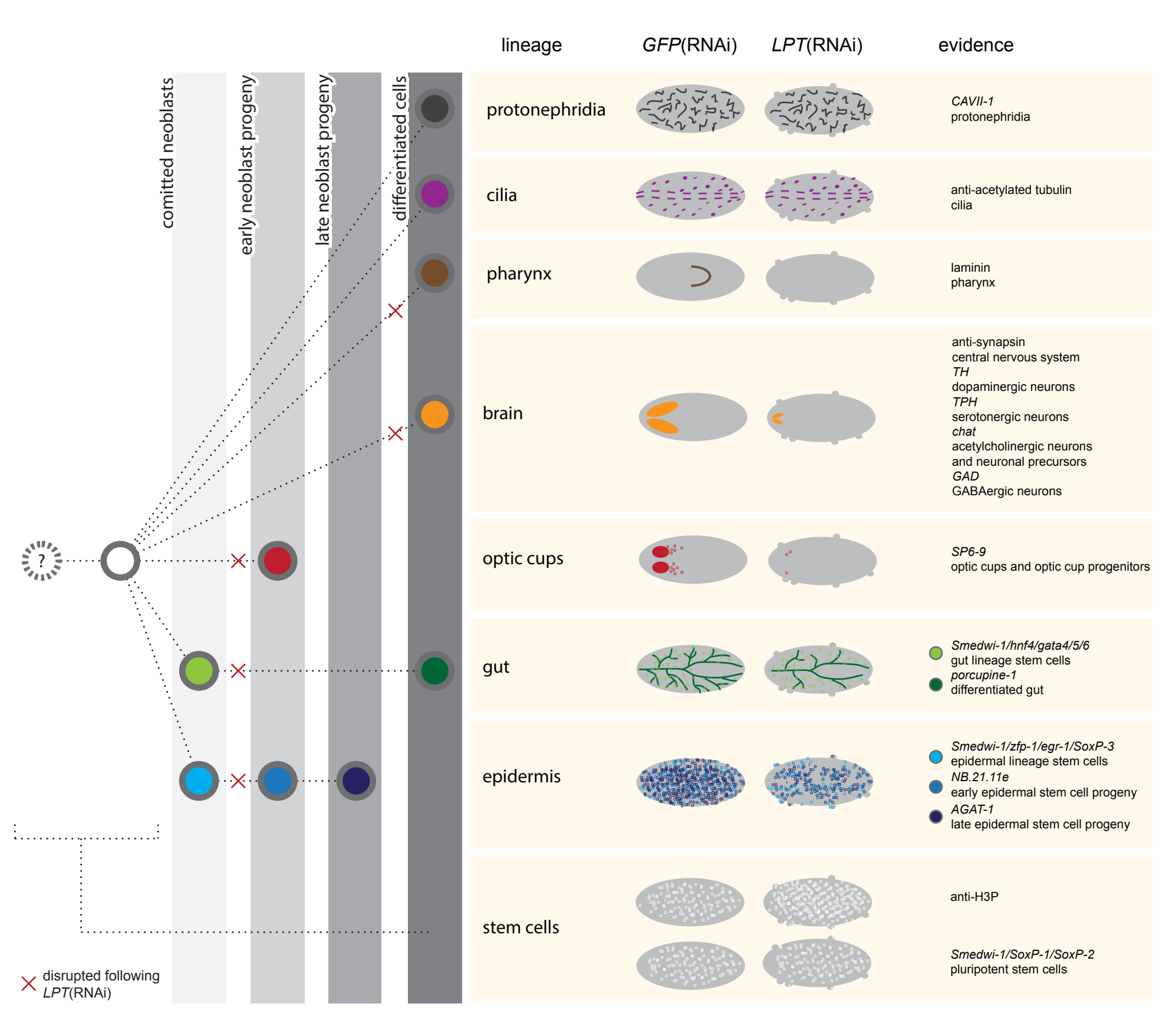



a
H
M
T
amputate
at day 10
RNAi
9 days
regeneration
day 1 regeneration
b
8 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
16/20
20/20
17/20
20/20
15/20
500 µm
c
10 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
2/20
20/20
17/20
20/20
1/20
500 µm
d
14 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
10/20
20/20
7/20
20/20
8/20
500 µm
e
GFP(RNAi)
LPT(RNAi)
10/10
7/10
Gut
8 days regeneration
500 µm
f
GFP(RNAi)
LPT(RNAi)
10/10
10/10
CNS
8 days regeneration
500 µm
g
GFP(RNAi)
LPT(RNAi)
10/10
9/10
optic chiasm
8 days regeneration
500 µm
h
GFP(RNAi)
LPT(RNAi)
10/10
9/10
optic cups
8 days regeneration
500 µm
i
GFP(RNAi)
LPT(RNAi)
10/10
8/10
pharynx
8 days regeneration
500 µm

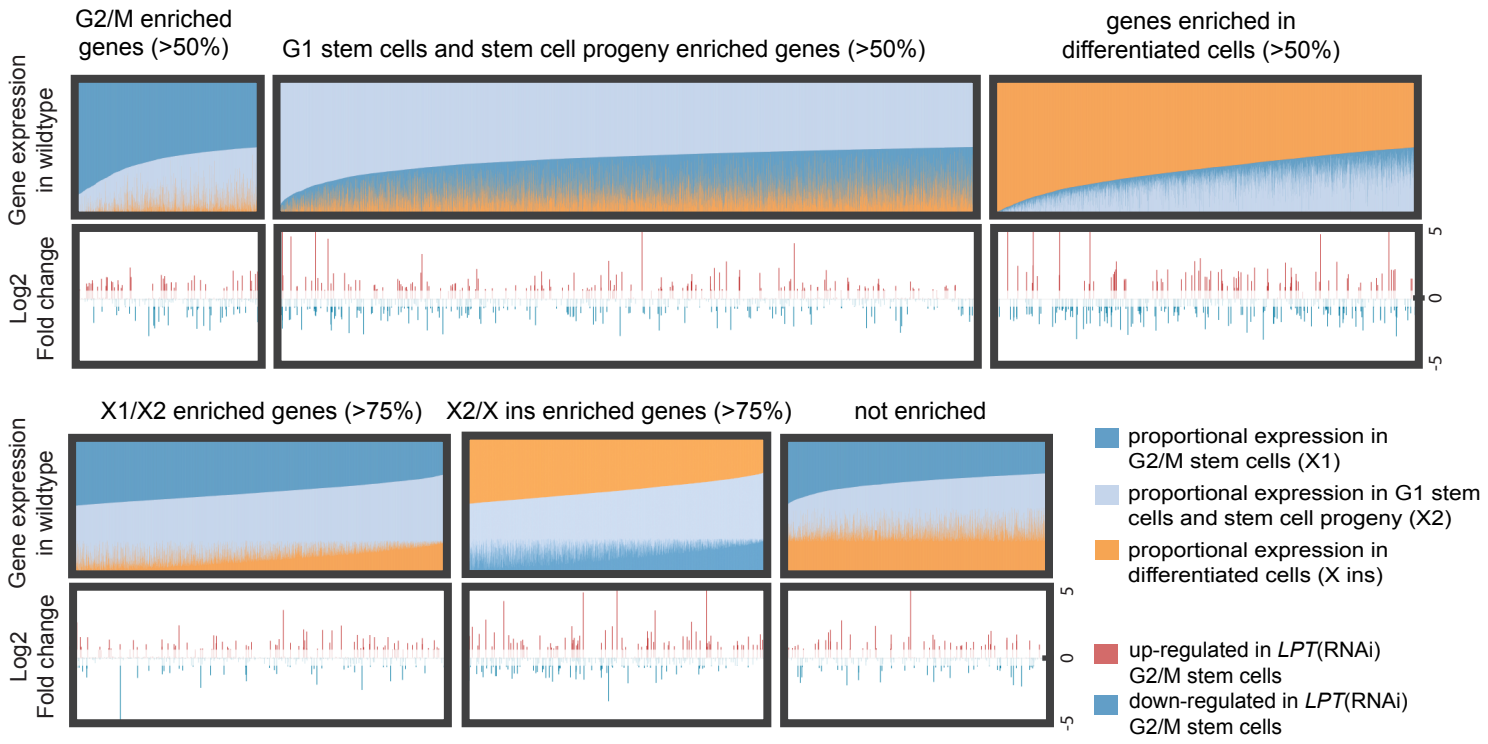

8 days regeneration
H
M
T
a
GFP(RNAi)
LPT(RNAi)
GABAergic neurons
500 µm
GAD+ cells/mm2
20
15
10
5
0
***
GFP (RNAi)
LPT (RNAi)
b
Dopaminergic neurons
500 µm
TH+ cells/mm2
20
15
10
5
0
*
GFP (RNAi)
LPT (RNAi)
c
Acetylcholinergic neurons
500 µm
chat+ cells/mm2
600
500
400
300
200
100
0
*
GFP (RNAi)
LPT (RNAi)
d
Serotonergic neurons
500 µm
TPH+ cells/mm2
25
20
15
10
5
0
*
GFP (RNAi)
LPT (RNAi)
e
Early and late epidermal stem cell progeny
500 µm
NB.21.11e+ cells/mm2
250
200
150
100
50
0
*
GFP (RNAi)
LPT (RNAi)
AGAT-1+ cells/mm2
200
150
100
50
0
*
GFP (RNAi)
LPT (RNAi)


a
1000
900
800
700
600
500
400
300
200
100
0
mitotic cells/mm 2
0 6 12 24 48 72 192
Hours after amputation
GFP(RNAi)
LPT(RNAi)
b
12 hrs regeneration
48 hrs regeneration
192 hrs regeneration
GFP (RNAi)
LPT (RNAi)
ph
500 µm
c
8 days regeneration
9 days regeneration
11 days regeneration
GFP (RNAi)
LPT (RNAi)
500 µm
d
GFP(RNAi)
H
M
T
LPT(RNAi)
mitotic cells
500 µm
10 days regeneration
100 µm
e
GFP(RNAi)
LPT(RNAi)
mitotic cells
300 µm
500 µm
12 days regeneration




lineage
GFP(RNAi)
LPT(RNAi)
evidence
comitted neoblasts
early neoblast progeny
late neoblast progeny
differentiated cells
?
protonephridia
CAVII-1
protonephridia
cilia
anti-acetylated tubulin
cilia
pharynx
laminin
pharynx
brain
anti-synapsin
central nervous system
TH
dopaminergic neurons
TPH
serotonergic neurons
chat
acetylcholinergic neurons
and neuronal precursors
GAD
GABAergic neurons
optic cups
SP6-9
optic cups and optic cup progenitors
gut
Smedwi-1/hnf4/gata4/5/6
gut lineage stem cells
porcupine-1
differentiated gut
epidermis
Smedwi-1/zfp-1/egr-1/SoxP-3
epidermal lineage stem cells
NB.21.11e
early epidermal stem cell progeny
AGAT-1
late epidermal stem cell progeny
stem cells
anti-H3P
Smedwi-1/SoxP-1/SoxP-2
pluripotent stem cells
disrupted following
LPT(RNAi)

a



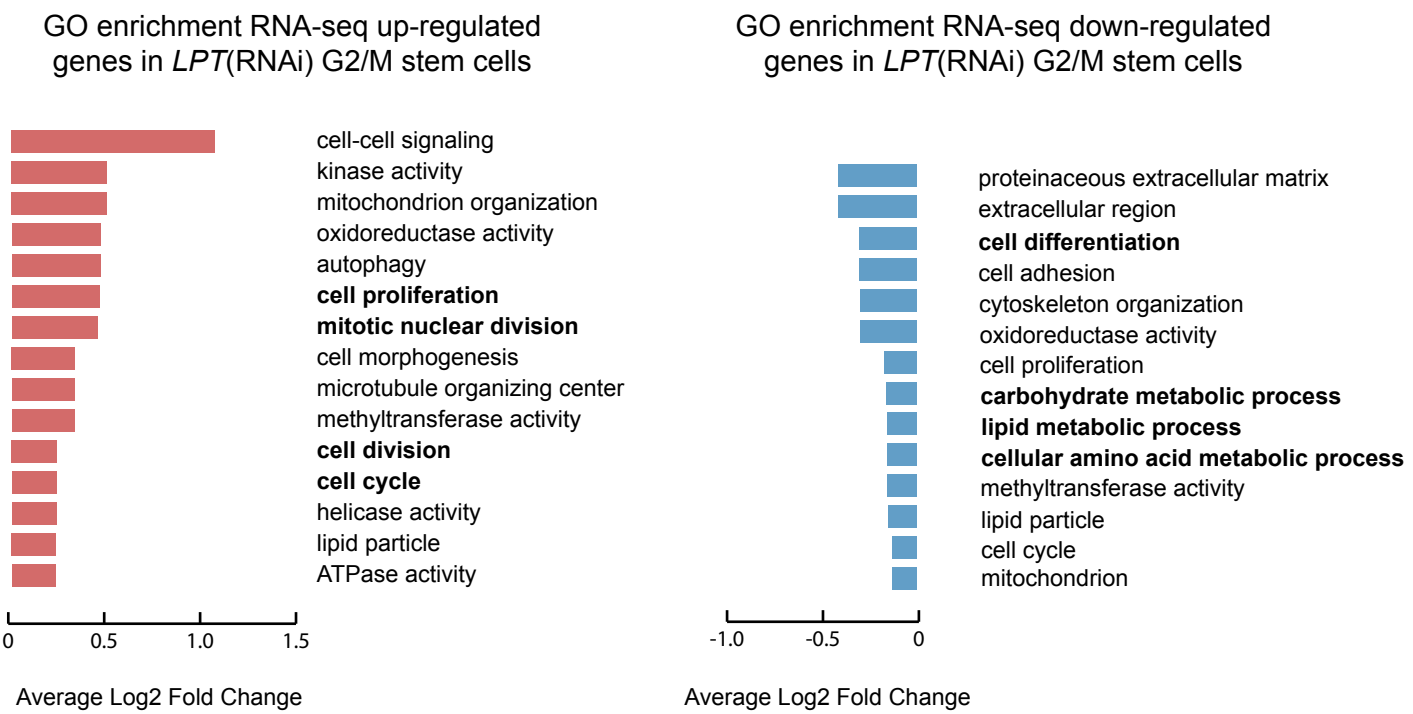
b

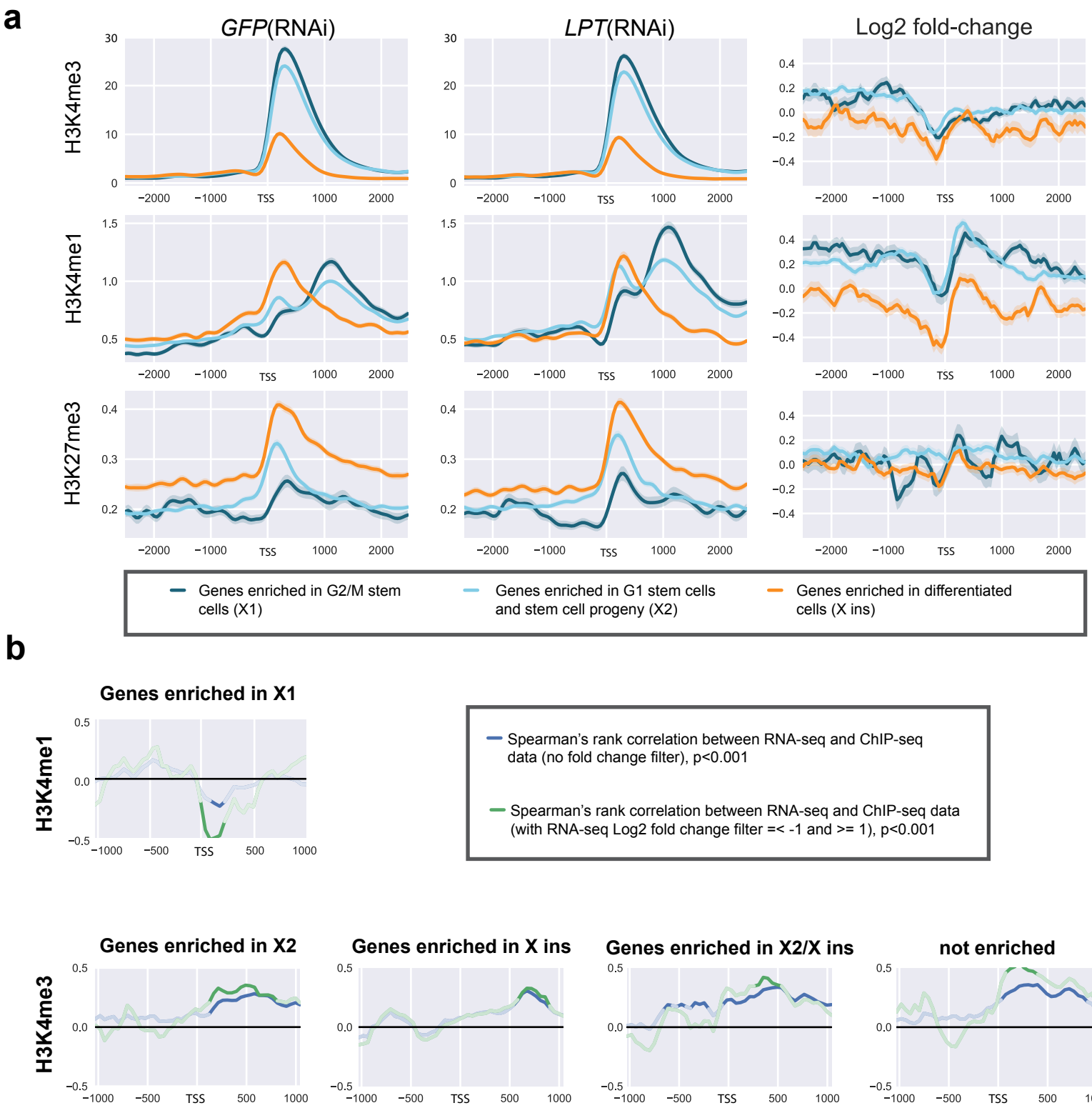
Number of genes up/down-regulated according to cell population and putative Transcription Factors

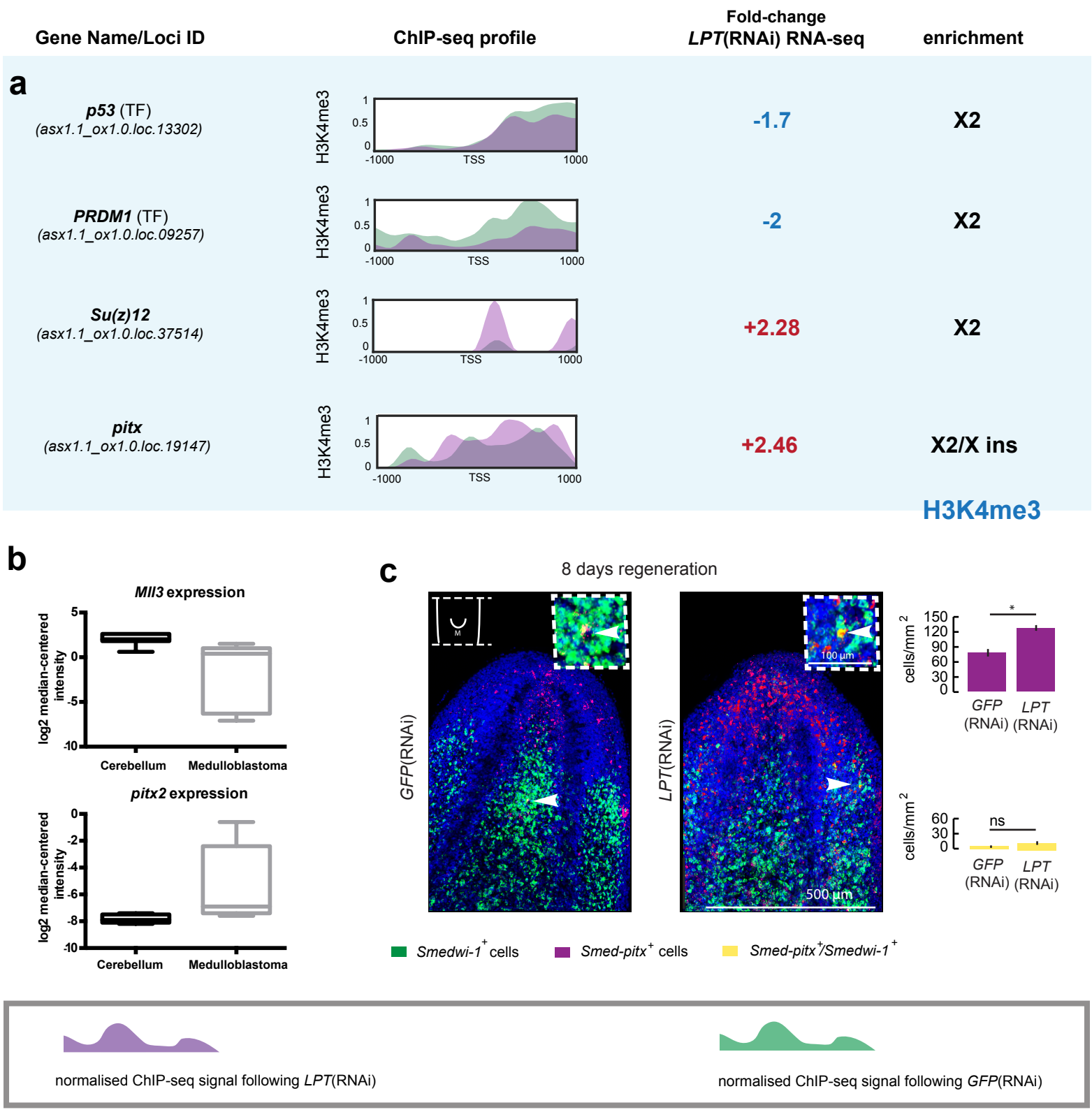
	X1 (2,253)	X2 (8,444)	X ins (5,119)	X1/X2 (4,538)	X2/X ins (3,652)	not enriched (3,200)	TFs (489)
up-regulated (542)	**49**	**160**	86	**83**	**79**	57	**18**
down-regulated (540)	37	**174**	**155**	44	**76**	42	**29**

*white indicates significant enrichment (p-value < 0.01)

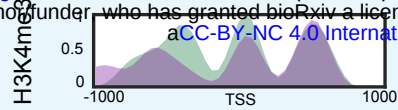
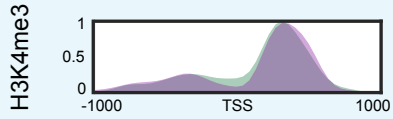
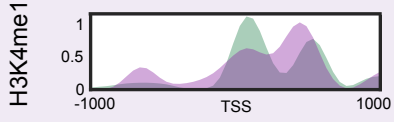
c




a
GFP(RNAi)
LPT(RNAi)
Log2 fold-change
H3K4me3
H3K4me1
H3K27me3
TSS
Genes enriched in G2/M stem cells (X1)
Genes enriched in G1 stem cells and stem cell progeny (X2)
Genes enriched in differentiated cells (X ins)
b
Genes enriched in X1
H3K4me1
Spearman's rank correlation between RNA-seq and ChIP-seq data (no fold change filter), p<0.001
Spearman's rank correlation between RNA-seq and ChIP-seq data (with RNA-seq Log2 fold change filter =< -1 and >= 1), p<0.001
Genes enriched in X2
Genes enriched in X ins
Genes enriched in X2/X ins
not enriched
H3K4me3


Gene Name/Loci ID
ChIP-seq profile
Fold-change
LPT(RNAi) RNA-seq
enrichment
a
p53 (TF)
(asx1.1_ox1.0.loc.13302)
PRDM1 (TF)
(asx1.1_ox1.0.loc.09257)
Su(z)12
(asx1.1_ox1.0.loc.37514)
pitx
(asx1.1_ox1.0.loc.19147)
H3K4me3
-1.7
-2
+2.28
+2.46
X2
X2
X2
X2/X ins
H3K4me3
-1000
TSS
1000
b
Mll3 expression
log2 median-centered intensity
Cerebellum
Medulloblastoma
pitx2 expression
c
8 days regeneration
GFP(RNAi)
LPT(RNAi)
100 µm
500 µm
cells/mm 2
GFP (RNAi)
LPT (RNAi)
ns
Smedwi-1+ cells
Smed-pitx+ cells
Smed-pitx+/Smedwi-1+
normalised ChIP-seq signal following LPT(RNAi)
normalised ChIP-seq signal following GFP(RNAi)

a

Gene Name/Loci ID	ChIP-seq profile	Fold-change *LPT*(RNAi) RNA-seq	enrichment
utx (*asx1.1_ox1.0.loc.29616*)		+2.2	X2
pim-1 (*asx1.1_ox1.0.loc.12499*)		+2.7	X2/X ins
			H3K4me3
pim-1-like (*asx1.1_ox1.0.loc.18774*)		+3	X1
			H3K4me1

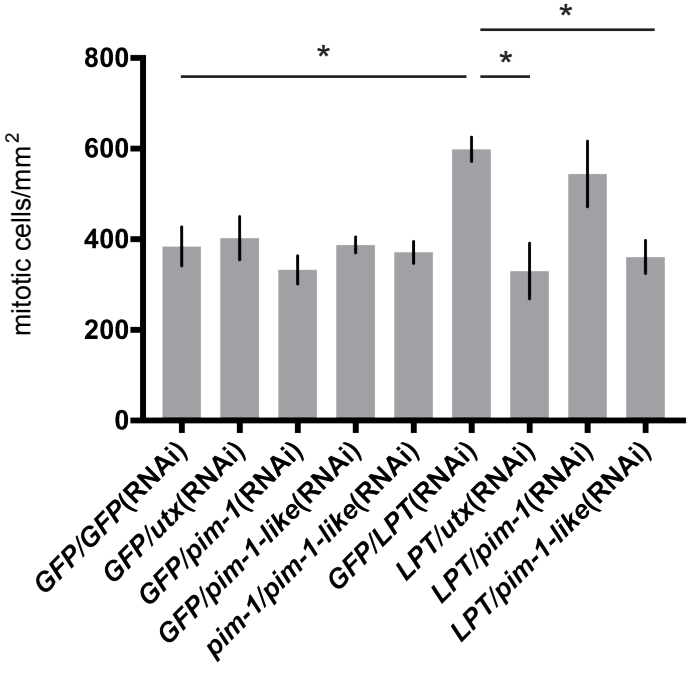
Bold font = correlation between RNA-seq and ChIP-seq data

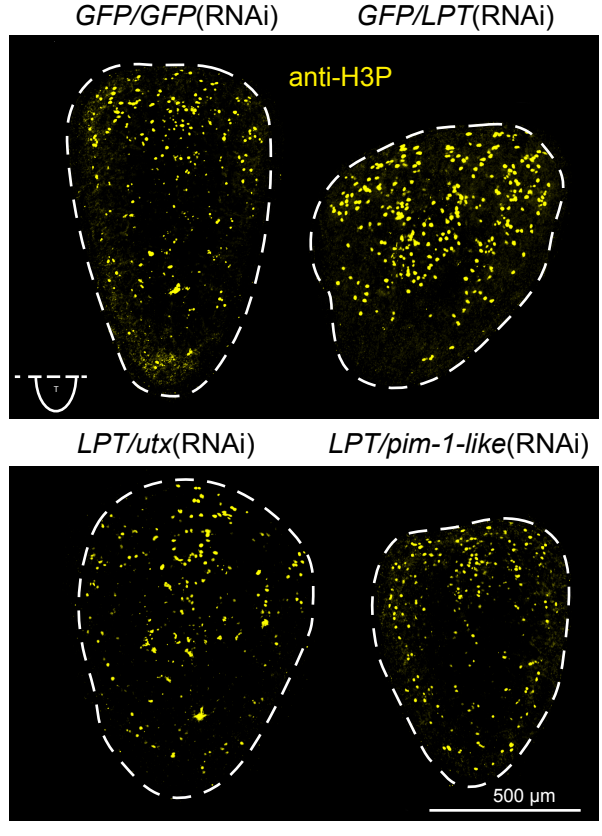

normalised ChIP-seq signal following *LPT*(RNAi)


normalised ChIP-seq signal following *GFP*(RNAi)

b

48 hours post-amputation





c