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 is often implicated in disease processes, particularly cancer. We focus on the tumour suppressors MLL3 and MLL4, which are mutated in a high percentage of cancers, but very little is known about the underlying transcriptional and epigenetic changes that contribute to malignancy.

Results

Here we make use of the highly accessible planarian model system to uncover a role for MLL3/4 orthologs in controlling stem cell differentiation and proliferation, suggesting conservation of tumour suppressor function over a large evolutionary timescale for these epigenetic regulators. 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 us to investigate the early-stage epigenetic and transcriptional changes in cells that will go on to form tumours after loss of MLL3/4 function, identifying genome wide alterations that occur early in the development of the pathology. This reveals mis-regulation of both conserved and hypothesised 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.*

52

53

54 **Background**

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

70 Many conserved signaling pathways are known to be responsible for
71 regulating growth, proliferation and other stem cell functions. In disease states
72 changes in the activity of these pathways can be due to effects on expression
73 levels, rather than mutations that change proteins, and this can be mediated
74 epigenetically through chromatin modifications. It is not surprising, therefore,
75 that mutations in chromatin modifying enzymes, like members of the
76 Polycomb and Trithorax complexes, are implicated in cancer[21-26]. The
77 genome-wide effects of chromatin modifying enzymes make understanding
78 how they contribute to cancer phenotypes very challenging. Complexity in the
79 form of tissue and cell heterogeneity, life history stage and stage of pathology
80 make resolution of epigenetic regulatory cause and effect relationships *in vivo*
81 very challenging. From this perspective, planarians and their accessible and
82 relatively homogenous stem cell population may be a very useful model
83 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**).

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

141

142 **Results**

143

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

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

149 binding domain, suggesting that it has chromatin-binding properties[49] (**Figure**
150 **1a**). There are two planarian genes homologous to *Drosophila* Trr and the C-
151 terminus of mammalian *Mll3/4* – *Smed-trr-1* (KC262345) and *Smed-trr-2*
152 (DN309269, HO004937), both previously described[36]. Both SMED-TRR-1 and
153 SMED-TRR-2 contain a PHD-like zinc-binding domain, a FYRN (FY-rich N-
154 terminal domain), FYRC (FY-rich C-terminal domain) and a catalytic SET
155 domain. SMED-TRR-1 (TRR-1) contains only a single NR (Nuclear Receptor)
156 box at a non-conserved position and SMED-TRR-2 (TRR-2) has no NR boxes
157 (**Figure 1a**). This could indicate that the planarian members of the trithorax-
158 related family are not broadly involved in recognition of nuclear receptors like
159 their arthropod and mammalian counterparts [33,35,50,51]. It is also possible that
160 some functional divergence exists between TRR-1 and TRR-2, where only
161 TRR-1 is capable of interacting with nuclear receptors.

162 We performed wholemount *in situ* hybridization (WISH) and found that *LPT*,
163 *trr-1* and *trr-2* are broadly expressed across many tissues and organs.
164 Gamma irradiation, used to remove all cycling cells in *S. mediterranea* within
165 24 hours, revealed that the three transcripts are also likely to be expressed in
166 stem cells (**Figure 1b**). This was supported by using an alternative method for
167 stem cell depletion – *H2B*(RNAi)[7] (**Additional File 2c**). The genes also
168 showed clear expression in the brain, pharynx and other post-mitotic
169 differentiated tissues (**Figure 1b**). During regeneration *Mll3/4* gene orthologs
170 are expressed in structures like the brain and pharynx as those are being
171 reformed (**Additional File 2b**).

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

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

190

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

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

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

213 formation of tissue outgrowths were temporally distinct and could be studied
214 consecutively.
215 A more thorough study of the differentiation properties of *LPT*(RNAi) animals
216 following amputation showed that the triclad gut structure failed to regenerate
217 secondary and tertiary branches and to extend major anterior and posterior
218 branches (**Figure 2e**). Cephalic ganglia (CG) regenerated as smaller
219 structures, the two CG lobes did not join in their anterior ends in *LPT*(RNAi)
220 animals (**Figure 2f**) and the optic chiasma and optic cups were mis-patterned
221 and markedly reduced (**Figure 2g-h**). We found that 80% of *LPT*(RNAi)
222 animals did not regenerate any new pharyngeal tissue (**Figure 2i**). We
223 interpreted these regenerative defects as being indicative of either a broad
224 failure in stem cell maintenance or in differentiation. The number of NBs
225 (*H2B*+ve) or early stem cell progeny cells across all lineages (SMEDWI-
226 1+ve/*H2B*-ve) [52,53] was not affected by perturbation of *LPT* function
227 (**Additional File 6**). Therefore, we infer that defects in later stem cell progeny
228 formation and terminal differentiation likely underpin failure in tissue
229 regeneration. Alternatively, different early stem cell progeny lineages could be
230 affected in opposite directions depending on their lineage, leading to the
231 observation of no overall change in the bulk number of early stem cell
232 progeny (SMEDWI-1+ve/*H2B*-ve).

233

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

236 One of the structures most severely affected following loss of *Mll3/4* function
237 was the brain. To investigate this further, we looked at the regeneration of
238 different neuronal subtypes. *LPT*(RNAi) animals had reduced numbers of
239 GABAergic (**Figure 3a**), dopaminergic (**Figure 3b**), acetylcholinergic (**Figure**
240 **3c**) and serotonergic (**Figure 3d**) neurons. As expected, the brain defects
241 were milder following knockdown of *trr-1* or *trr-2* (**Additional File 7a-d**).

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

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

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

251 One of the other observable defects in *LPT*(RNAi) animals was abnormal
252 locomotion, with nearly all worms displaying muscular inch worming rather
253 than smooth cilia-mediated locomotion. Given differentiation of cilia was
254 unaffected, it seemed likely this effect was a result of neuronal differentiation
255 defects, specifically serotonin-dependent control of beating cilia [54]. In
256 agreement with this interpretation, ectopic serotonin hydrochloride treatment
257 improved the gliding movement of *Mll3/4* knockdown animals (**Additional File**
258 **9**).

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

262

263 ***Mll3/4* limits normal stem cell proliferation and tissue growth**

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

270 Planarian regeneration is characterized by an early burst of increased NB
271 proliferation, 6-12 hours after wounding, and a second peak of proliferation,
272 48 hours after amputation[53]. Following *LPT*(RNAi), we observed significant
273 increases in proliferation at both of these peaks and at 8 days post-
274 amputation, as proliferation fails to return to normal homeostatic levels
275 (**Figure 4a**). *Trr-1*(RNAi) and *trr-2*(RNAi) animals also show elevated
276 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

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

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

322

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

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

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

341 in X1 cells (**Figure 7a**), affecting genes normally expressed in many different
342 planarian cell compartments (**Figure 7b**). These findings confirm that
343 *LPT*(RNAi) has a complex effect, influencing different gene classes, including
344 47 transcription factors, in stem cells.

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

352

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

355 Previous studies tie MLL3/4/LPT-Trr function directly to mono- and tri-
356 methylation of H3K4 [29,31-34] and indirectly to trimethylation of H3K27, because
357 the H3K27me3 demethylase UTX is present in the same protein complex [60].
358 In order to understand potential epigenetic causes of the transcriptional
359 changes following *LPT*(RNAi), we also performed ChIP-seq on X1 cells. The
360 profile of H3K4me3, H3K4me1 and H3K27me3 in control X1 cells showed that
361 genes enriched in X1 stem cells have the highest H3K4me3 and the lowest
362 H3K27me3 signal at predicted transcriptional start sites (TSSs), consistent
363 with active transcription of these genes (**Figure 8a, Additional File 14**). We
364 observed the reverse pattern for genes enriched in differentiated cells,
365 consistent with repressed transcription. Furthermore, the peak in H3K4me1
366 signal is shifted downstream and away from the TSS for genes with enriched
367 expression in X1 NBs, allowing for active transcription. Conversely, the peak
368 in H3K4me1 is positioned across the TSS for genes with enriched expression
369 in differentiated cells and no or relatively low expression in NBs, indicative of
370 repressed transcription in NBs. These data are in agreement with previous
371 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**

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

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

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

436 failure of serotonergic neuron regeneration is not clear, but MLL3/4/LPT-Trr
437 activity does control the maturation of serotonergic neuronal precursors into
438 serotonergic neurons. The *LPT*(RNAi)-dependent up-regulation of *pitx* might
439 also be a conserved feature of MLL3/4 mis-regulation in some cancer types
440 [63].

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

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

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

469 H3K27me3 demethylase, *utx,* has been implicated in increased proliferation
470 and tumor invasiveness [67]. Based on this, the overexpression of *utx, pim-1*
471 and *pim-1-like* together represented some of the best candidates for major
472 effects amongst those with significant up-regulation in expression. In order to
473 test whether the up-regulated expression of *pim-1*, *pim-1-like* or *utx* is
474 potentially key to the *LPT*(RNAi) cancer-like phenotype, we attempted
475 *LPT*(RNAi) rescue experiments in the form of double RNAi knockdowns
476 (**Additional File 17)**. At 48 hours post-amputation, *LPT*(RNAi) regenerates
477 have a significantly increased stem cell proliferation (**Figure 4a**, **b**) and so do
478 *GFP/LPT*(RNAi) double knockdown animals (**Figure 10b**). Whereas *LPT/pim-*
479 *1*(RNAi) regenerates still have elevated NB proliferation, both *LPT/pim-1-*
480 *like*(RNAi) and *LPT/utx*(RNAi) regenerates have a significantly decreased NB
481 proliferation compared to *GFP/LPT*(RNAi), and half as many animals in these
482 two conditions went on to form outgrowths (**Figure 10c**). These findings
483 suggest that the up-regulation of both *pim-1-like* and *utx* are involved in
484 driving the *LPT*(RNAi) animals' cancer-like phenotype. *Pim-1-like'*s up-
485 regulation in *LPT*(RNAi) planarians may be directly connected to changing
486 histone modifications at its promoter, while *utx*'s up-regulation is likely not due
487 to a direct effect on a promoter region (**Figure 10a**).

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

494

495 **Discussion**

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

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

515 We found that clusters of mitotic cells preceded the appearance of outgrowths
516 in *LPT*(RNAi) regenerating animals, likely pre-empting where the outgrowths
517 would subsequently form. The observation of clusters of cells and the
518 formation of outgrowths in some but not all RNAi animals suggests a
519 heterogeneity in stem cell responses to *LPT*(RNAi). This probably reflects the
520 stochastic nature of the broad genome wide epigenetic changes that will have
521 some variability between cells, such that only some cells cycle out of control
522 and cause outgrowths. We also observed that outgrowth tissue contained
523 different classes of stem cells. Among these stem cells, the presence of
524 sigma NBs (thought to include truly pluripotent stem cells [57] is of particular
525 significance. When mis-regulated, these cells could share fundamental
526 similarities with cancer stem cells (CSCs) often found in human malignancies
527 [69]. CSCs have been described as one of the main factors in cancer
528 aggressiveness and resistance to treatment [70]. Studying such cells in a simple
529 *in vivo* stem cell model provided by the planarian system should bring further
530 insight into important control mechanisms that are mis-regulated in different
531 cancers. Our work here provides a useful example of this approach.

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

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

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

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

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

572

573 **Conclusion**

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

582

583 **Methods**

584 *Animal husbandry*

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

590

591 *RNAi*

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

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

601 DsRNA was delivered via microinjection using Nanoject II apparatus
602 (Drummond Scientific) with 3.5'' Drummond Scientific (Harvard Apparatus)
603 glass capillaries pulled into fine needles on a Flaming/Brown Micropipette
604 Puller (Patterson Scientific). Each animal received around 100 nl dsRNA each
605 day. *H2B*(RNAi) was performed for three consecutive days, as per Solana et
606 al.'s (2012) protocol. For single and double *LPT*, *trr-1* and *trr-2* knockdown, a
607 course of 7 days of microinjections was performed (3 consecutive days + 2
608 days rest + 4 consecutive days). S*et1*(RNAi) and *utx*(RNAi) were performed
609 for 4 consecutive days.

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

612

613 *In situ hybridization*

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

623

624 *Immunohistochemistry*

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

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

632

633 *Imaging and image analysis*

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

639

640 *Flow cytometry*

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

645

646 *Western blot*

647 2xLaemmli buffer (Sigma Aldrich), 1M DTT and cOmplete protease inhibitors
648 (Roche) were used for protein extraction from 10-15 animals per condition.
649 Protein extract was quantified with Qubit Protein Assay kit (Thermo Fisher
650 Scientific). NuPAGE Novex 4-12% Bis-Tris protein gels (Thermo Fisher
651 Scienitific) were used, followed by a wet transfer in a Mini Trans-Blot
652 Electrophoretic Transfer Cell machine. Ponceau S (Sigma Aldrich) whole-
653 protein stain was used prior to antibody incubation. The antibodies used were:
654 anti-H3 (unmodified histone H3; rabbit polyclonal; Abcam; ab1791; 1:10000
655 dilution), anti-H3K4me3 (rabbit polyclonal; Abcam; ab8580; 1:1000 dilution),
656 anti-H3K4me1 (rabbit polyclonal; Abcam; ab8895; 1:1000 dilution), anti-
657 H3K27me3 (mouse monoclonal; Abcam; ab6002; 1:1000 dilution), anti-mouse
658 IgG HRP-linked antibody (Cell Signalling; 7076P2), anti-rabbit IgG HRP-linked

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

661

662 *ChIP-seq*

663 600000-700000 planarian x1 cells were FACS-sorted (using 3-day knockdown
664 regenerates) in PBS and pelleted at 4 ºC. During the pelleting, S2 cells were
665 added (corresponding to roughly 15% of the number of planarian x1 cells) for
666 the purpose of downstream data normalisation. Samples were then processed
667 as described in Kao et al. (2017). The process is summarized in **Additional**
668 **File 19**. The libraries were sequenced on an Illumina NextSeq machine.
669 Three biological replicates were prepared. The raw reads are available in the
670 Short Read Archive (PRJNA338116).

671

672 *RNA-seq*

673 300000 x1 NBs were FACS-sorted in RNALater (Ambion) from knockdown
674 animals at 3 days of regeneration. Cells were pelleted at 4 ºC and Trizol-
675 based total RNA extraction was performed. The amount of total RNA used for
676 each library preparation was 0.8-1 µg. Illumina TruSeq Stranded mRNA LT kit
677 was used for library preparation. The kit instructions were followed. Libraries
678 were quantified with Qubit, Agilent Bioanalyzer and KAPA Library
679 Quantification qPCR kit. Samples were sequenced on an Illumina NextSeq
680 machine. Two biological replicates were prepared. The raw reads are
681 available in the Short Read Archive (PRJNA338115).

682

683 *ChIP-seq data analysis*

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

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

691

692 *RNA-seq data analysis*

693 Raw reads were trimmed with Trimmomatic 0.32[76] and pseudo-aligned to a
694 set of asexual genome annotations described in Kao et al. (2017) with Kallisto
695 0.42[80]. Differential expression was subsequently performed with Sleuth
696 0.28.1[81]. For more in-depth methods, including code, refer to the
697 **Supplementary Python Notebook**.

698

699 *Statistical methods*

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

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

710

711 *Data availability*

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

752 5. Labbé, R. M. *et al.* A Comparative Transcriptomic Analysis Reveals
753 Conserved Features of Stem Cell Pluripotency in Planarians and
754 Mammals. *STEM CELLS* **30,** 1734–1745 (2012).
755 6. Solana, J. *et al.* Conserved functional antagonism of CELF and MBNL
756 proteins controls stem cell-specific alternative splicing in planarians.
757 *eLife* **5,** 1193 (2016).
758 7. Solana, J. *et al.* Defining the molecular profile of planarian pluripotent
759 stem cells using a combinatorial RNA-seq, RNA interference and
760 irradiation approach. *Genome Biol.* **13,** R19 (2012).
761 8. Alié, A. *et al.* The ancestral gene repertoire of animal stem cells. *Proc*
762 *Natl Acad Sci USA* 201514789–8 (2015).
763 doi:10.1073/pnas.1514789112
764 9. Kao, D., Mihaylova, Y., Hughes, S., Lai, A. & Aboobaker, A. Epigenetic
765 analyses of the planarian genome reveals conservation of bivalent
766 promoters in animal stem cells. *bioRxiv* 122135 (2017).
767 doi:10.1101/122135
768 10. Shibata, N. *et al.* Inheritance of a Nuclear PIWI from Pluripotent Stem
769 Cells by Somatic Descendants Ensures Differentiation by Silencing
770 Transposons in Planarian. *Developmental Cell* **37,** 226–237 (2016).
771 11. Salvetti, A. DjPum, a homologue of Drosophila Pumilio, is essential to
772 planarian stem cell maintenance. *Development* **132,** 1863–1874 (2005).
773 12. Reddien, P. W. Specialized progenitors and regeneration. *Development*
774 **140,** 951–957 (2013).
775 13. Juliano, C. E., Swartz, S. Z. & Wessel, G. M. A conserved germline
776 multipotency program. *Development* **137,** 4113–4126 (2010).
777 14. Jaber-Hijazi, F. *et al.* Planarian MBD2/3 is required for adult stem cell
778 pluripotency independently of DNA methylation. *Developmental Biology*
779 **384,** 141–153 (2013).
780 15. Scimone, M. L., Meisel, J. & Reddien, P. W. The Mi-2-like Smed-CHD4
781 gene is required for stem cell differentiation in the planarian Schmidtea
782 mediterranea. *Development* **137,** 1231–1241 (2010).
783 16. Zhu, S. J., Hallows, S. E., Currie, K. W., Xu, C. & Pearson, B. J. A mex3
784 homolog is required for differentiation during planarian stem cell lineage
785 development. *eLife* **4,** 304 (2015).
786 17. Cowles, M. W., Omuro, K. C., Stanley, B. N., Quintanilla, C. G. & Zayas,
787 R. M. COE Loss-of-Function Analysis Reveals a Genetic Program
788 Underlying Maintenance and Regeneration of the Nervous System in
789 Planarians. *PLoS Genet* **10,** e1004746–12 (2014).
790 18. Barberan, S., Fraguas, S. & Cebrià, F. The EGFR signaling pathway
791 controls gut progenitor differentiation during planarian regeneration and
792 homeostasis. *Development* **143,** 2089–2102 (2016).
793 19. Guedelhoefer, O. C. & Alvarado, A. S. Amputation induces stem cell
794 mobilization to sites of injury during planarian regeneration.
795 *Development* **139,** 3510–3520 (2012).
796 20. Abnave, P. *et al.* A shielded irradiation assay to investigate mechanisms
797 of in vivo stem cell migration in planarians. 1–42 (2016).
798 doi:10.1101/080853
799 21. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in
800 progression of prostate cancer. *Nature* **419,** 624–629 (2002).
801 22. Villa, R. *et al.* Role of the Polycomb Repressive Complex 2 in Acute

802 Promyelocytic Leukemia. *Cancer Cell* **11,** 513–525 (2007).
803 23. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer
804 medulloblastoma. *Science* **331,** 435–439 (2011).
805 24. Kandoth, C. *et al.* Mutational landscape and significance across 12
806 major cancer types. *Nature* **502,** 333–339 (2013).
807 25. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in
808 transitional cell carcinoma of the bladder. *Nature Publishing Group* **43,**
809 875–878 (2011).
810 26. Chen, C. *et al.* MLL3 Is a Haploinsufficient 7q Tumor Suppressor in
811 Acute Myeloid Leukemia. *Cancer Cell* **25,** 652–665 (2014).
812 27. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases:
813 Mechanisms of Regulation in Development and Disease Pathogenesis.
814 *Annu. Rev. Biochem.* **81,** 65–95 (2012).
815 28. Wu, M. *et al.* Molecular Regulation of H3K4 Trimethylation by Wdr82, a
816 Component of Human Set1/COMPASS. *Molecular and Cellular Biology*
817 **28,** 7337–7344 (2008).
818 29. Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by
819 Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4.
820 *Genes & Development* **26,** 2604–2620 (2012).
821 30. Wang, P. *et al.* Global Analysis of H3K4 Methylation Defines MLL
822 Family Member Targets and Points to a Role for MLL1-Mediated H3K4
823 Methylation in the Regulation of Transcriptional Initiation by RNA
824 Polymerase II. *Molecular and Cellular Biology* **29,** 6074–6085 (2009).
825 31. Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family
826 Function as Major Histone H3K4 Monomethylases at Enhancers.
827 *Molecular and Cellular Biology* **33,** 4745–4754 (2013).
828 32. Lee, J. *et al.* A tumor suppressive coactivator complex of p53 containing
829 ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue
830 MLL4. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 8513–8518 (2009).
831 33. Sedkov, Y. *et al.* Methylation at lysine 4 of histone H3 in ecdysone-
832 dependent development of Drosophila. *Nature* **426,** 78–83 (2003).
833 34. Cheng, J. *et al.* A Role for H3K4 Monomethylation in Gene Repression
834 and Partitioning of Chromatin Readers. *Molecular Cell* **53,** 979–992
835 (2014).
836 35. Chauhan, C., Zraly, C. B., Parilla, M., Diaz, M. O. & Dingwall, A. K.
837 Histone recognition and nuclear receptor co-activator functions of
838 Drosophila Cara Mitad, a homolog of the N-terminal portion of
839 mammalian MLL2 and MLL3. *Development* **139,** 1997–2008 (2012).
840 36. Hubert, A. *et al.* Epigenetic regulation of planarian stem cells by the
841 SET1/MLL family of histone methyltransferases. *Epigenetics* **8,** 79–91
842 (2013).
843 37. Duncan, E. M., Chitsazan, A. D., Seidel, C. W. & Alvarado, A. S. Set1
844 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci
845 In Vivo. *CellReports* **13,** 2741–2755 (2015).
846 38. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required
847 for enhancer activation during cell differentiation. *eLife* **2,** 2817–25
848 (2013).
849 39. Mohan, M. *et al.* The COMPASS Family of H3K4 Methylases in
850 Drosophila. *Molecular and Cellular Biology* **31,** 4310–4318 (2011).
851 40. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4

852 controls cell fate transition. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 11871–
853 11876 (2016).
854 41. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent
855 promoters in embryonic stem cells, whereas Mll1 is redundant.
856 *Development* **141,** 526–537 (2014).
857 42. Hsieh, J. J.-D., Ernst, P., Erdjument-Bromage, H., Tempst, P. &
858 Korsmeyer, S. J. Proteolytic cleavage of MLL generates a complex of
859 N- and C-terminal fragments that confers protein stability and
860 subnuclear localization. *Molecular and Cellular Biology* **23,** 186–194
861 (2003).
862 43. Chen, W. *et al.* Malignant Transformation Initiated by Mll-AF9: Gene
863 Dosage and Critical Target Cells. *Cancer Cell* **13,** 432–440 (2008).
864 44. Corral, J. *et al.* An Mll-AF9 fusion gene made by homologous
865 recombination causes acute leukemia in chimeric mice: a method to
866 create fusion oncogenes. *Cell* **85,** 853–861 (1996).
867 45. Thirman, M. J. *et al.* Rearrangement of the MLL gene in acute
868 lymphoblastic and acute myeloid leukemias with 11q23 chromosomal
869 translocations. *N. Engl. J. Med.* **329,** 909–914 (1993).
870 46. Sobulo, O. M. *et al.* MLL is fused to CBP, a histone acetyltransferase, in
871 therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc*
872 *Natl Acad Sci USA* **94,** 8732–8737 (1997).
873 47. Goo, Y. H. *et al.* Activating Signal Cointegrator 2 Belongs to a Novel
874 Steady-State Complex That Contains a Subset of Trithorax Group
875 Proteins. *Molecular and Cellular Biology* **23,** 140–149 (2003).
876 48. Lee, S. *et al.* Coactivator as a target gene specificity determinant for
877 histone H3 lysine 4 methyltransferases. *Proc Natl Acad Sci USA* **103,**
878 15392–15397 (2006).
879 49. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends*
880 *in Biochemical Sciences* **31,** 35–40 (2006).
881 50. Lee, S., Lee, J., Lee, S.-K. & Lee, J. W. Activating Signal Cointegrator-2
882 Is an Essential Adaptor to Recruit Histone H3 Lysine 4
883 Methyltransferases MLL3 and MLL4 to the Liver X Receptors. *Molecular*
884 *Endocrinology* **22,** 1312–1319 (2008).
885 51. Ansari, K. I., Hussain, I., Kasiri, S. & Mandal, S. S. HOXC10 is
886 overexpressed in breast cancer and transcriptionally regulated by
887 estrogen via involvement of histone methylases MLL3 and MLL4.
888 *Journal of Molecular Endocrinology* **48,** 61–75 (2012).
889 52. Guo, T., Peters, A. H. F. M. & Newmark, P. A. A bruno-like Gene Is
890 Required for Stem Cell Maintenance in Planarians. *Developmental Cell*
891 **11,** 159–169 (2006).
892 53. Wenemoser, D. & Reddien, P. W. Planarian regeneration involves
893 distinct stem cell responses to wounds and tissue absence.
894 *Developmental Biology* **344,** 979–991 (2010).
895 54. Currie, K. W. & Pearson, B. J. Transcription factors lhx1/5-1 and pitx are
896 required for the maintenance and regeneration of serotonergic neurons
897 in planarians. *Development* **140,** 3577–3588 (2013).
898 55. Oviedo, N. J., Pearson, B. J., Levin, M. & Sanchez Alvarado, A.
899 Planarian PTEN homologs regulate stem cells and regeneration through
900 TOR signaling. *Disease Models and Mechanisms* **1,** 131–143 (2008).
901 56. Gonzalez-Estevez, C. *et al.* SMG-1 and mTORC1 Act Antagonistically

902 to Regulate Response to Injury and Growth in Planarians. *PLoS Genet*
903 **8,** e1002619–17 (2012).
904 57. van Wolfswinkel, J. C., Wagner, D. E. & Reddien, P. W. Single-Cell
905 Analysis Reveals Functionally Distinct Classes within the Planarian
906 Stem Cell Compartment. *Stem Cell* **15,** 326–339 (2014).
907 58. Zink, D., Fische, A. H. & Nickerson, J. A. Nuclear structure in cancer
908 cells. *Nat Rev Cancer* **4,** 677–687 (2004).
909 59. Kim, D.-H., Kim, J. & Lee, J. W. Requirement for MLL3 in p53
910 Regulation of Hepatic Expression of Small Heterodimer Partner and Bile
911 Acid Homeostasis. *Molecular Endocrinology* **25,** 2076–2083 (2011).
912 60. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb
913 recruitment and H2A ubiquitination. *Science* **318,** 447–450 (2007).
914 61. Pearson, B. J. & Alvarado, A. S. A planarian p53 homolog regulates
915 proliferation and self-renewal in adult stem cell lineages. *Development*
916 **137,** 213–221 (2009).
917 62. März, M., Seebeck, F. & Bartscherer, K. A Pitx transcription factor
918 controls the establishment and maintenance of the serotonergic lineage
919 in planarians. *Development* **140,** 4499–4509 (2013).
920 63. Pomeroy, S. L. *et al.* Prediction of central nervous system embryonal
921 tumour outcome based on gene expression. *Nature* **415,** 436–442
922 (2002).
923 64. Roh, M. *et al.* Overexpression of the oncogenic kinase Pim-1 leads to
924 genomic instability. *Cancer Res* **63,** 8079–8084 (2003).
925 65. Valdman, A., Fang, X., Pang, S.-T., Ekman, P. & Egevad, L. Pim-1
926 expression in prostatic intraepithelial neoplasia and human prostate
927 cancer. *Prostate* **60,** 367–371 (2004).
928 66. Shirogane, T. *et al.* Synergistic roles for Pim-1 and c-Myc in STAT3-
929 mediated cell cycle progression and antiapoptosis. *Immunity* **11,** 709–
930 719 (1999).
931 67. Kim, J.-H. *et al.* UTX and MLL4 Coordinately Regulate Transcriptional
932 Programs for Cell Proliferation and Invasiveness in Breast Cancer Cells.
933 *Cancer Res* **74,** 1705–1717 (2014).
934 68. Zhang, Z. *et al.* Mammary-Stem-Cell-Based Somatic Mouse Models
935 Reveal Breast Cancer Drivers Causing Cell Fate Dysregulation.
936 *CellReports* **16,** 3146–3156 (2016).
937 69. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells,
938 cancer, and cancer stem cells. *Nature* **414,** 105–111 (2001).
939 70. Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance.
940 *Nat Rev Cancer* **5,** 275–284 (2005).
941 71. Cho, Y. W. *et al.* PTIP Associates with MLL3- and MLL4-containing
942 Histone H3 Lysine 4 Methyltransferase Complex. *Journal of Biological*
943 *Chemistry* **282,** 20395–20406 (2007).
944 72. Cebria, F. Planarian homologs of netrin and netrin receptor are required
945 for proper regeneration of the central nervous system and the
946 maintenance of nervous system architecture. *Development* **132,** 3691–
947 3703 (2005).
948 73. Gonzalez-Estevez, C., Arseni, V., Thambyrajah, R. S., Felix, D. A. &
949 Aboobaker, A. A. Diverse miRNA spatial expression patterns suggest
950 important roles in homeostasis and regeneration in planarians. *Int. J.*
951 *Dev. Biol.* **53,** 493–505 (2009).

952 74. Cebria, F. & Newmark, P. A. Morphogenesis defects are associated
953 with abnormal nervous system regeneration following roboA RNAi in
954 planarians. *Development* **134,** 833–837 (2007).
955 75. Romero, B. T., Evans, D. J. & Aboobaker, A. A. in *Progenitor Cells* **916,**
956 167–179 (Humana Press, 2012).
957 76. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer
958 for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
959 77. Robb, S. M. C., Gotting, K., Ross, E. & Sánchez Alvarado, A. SmedGD
960 2.0: The Schmidtea mediterranea genome database. *Genesis* **53,** 535–
961 546 (2015).
962 78. Attrill, H. *et al.* FlyBase: establishing a Gene Group resource for
963 Drosophila melanogaster. *Nucleic Acids Research* **44,** D786–92 (2016).
964 79. Orlando, D. A. *et al.* Quantitative ChIP-Seq Normalization Reveals
965 Global Modulation of the Epigenome. *CellReports* **9,** 1163–1170 (2014).
966 80. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal
967 probabilistic RNA-seq quantification. *Nat Biotech* **34,** 525–527 (2016).
968 81. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L.
969 Differential analysis of RNA-Seq incorporating quantification
970 uncertainty. (2016). doi:10.1101/058164
971 82. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L.
972 Differential analysis of RNA-Seq incorporating quantification
973 uncertainty. (2016). doi:10.1101/058164
974

975

976

977 **Figure legends**

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

992 Expression profiles of *Mll3/4* genes according to RNA-seq data from FACS-
993 sorted X1 (stem cells in G2/M phase), X2 (stem cells in G1 and stem cell
994 progeny) and X ins (differentiated cells) planarian cell populations.
995
996
997 **Figure 2. *LPT*(RNAi) results in differentiation defects and outgrowth**
998 **formation during regeneration.** (**a**) A schematic showing the amputation of
999 RNAi worms into head (H), middle (M) and tail (T) pieces in order to observe
1000 regeneration of different structures. The time-course of all the experiments on
1001 *Mll3/4* knockdown animals is depicted underneath the worm schematic. A
1002 total of 9 days of dsRNA microinjection-mediated RNAi was followed by
1003 amputation on the 10th day and subsequent observation of regeneration. (**b**)
1004 Head, middle and tail pieces following *LPT*(RNAi) or control *GFP*(RNAi) at
1005 day 8 of regeneration. Yellow arrows point towards the smaller blastema and
1006 the eye formation defects. (**c**) Head, middle and tail pieces following
1007 *LPT*(RNAi) or control *GFP*(RNAi) at day 10 of regeneration. Red arrows point
1008 towards outgrowths. (**d**) Head, middle and tail pieces following *LPT*(RNAi) or
1009 control *GFP*(RNAi) at day 14 of regeneration. Red arrows point towards
1010 outgrowths. (**e**) Gut regeneration and maintenance in middle pieces following
1011 *LPT*(RNAi), as illustrated by RNA probe for the gene *porcupine-1* at 8 days of
1012 regeneration. (**f**) Brain regeneration in middle pieces at 8 days post-
1013 amputation following *LPT*(RNAi), as illustrated by anti-SYNORF-1 antibody
1014 labeling the central nervous system (CNS). (**g**) Optic chiasma recovery in tail
1015 pieces at 8 days of regeneration following *LPT*(RNAi), as shown by anti-VC-1
1016 antibody. (**h**) Recovery of optic cups and organized trail of optic cup precursor
1017 cells in tail pieces at 8 days of regeneration following *LPT*(RNAi), as
1018 demonstrated by RNA probe for *SP6-9*. (**i**) Pharynx recovery in head pieces at
1019 8 days of regeneration following *LPT*(RNAi), as illustrated by RNA probe for
1020 *laminin*. Images in (**e**,**f**, **i**) are representative of two separate experiments
1021 using 10 animals per condition each. Images in (**g**, **h**) were obtained from one
1022 experiment each, using 10 animals per condition. Numbers at the top right of
1023 each regenerating piece represent number of animals in the condition
1024 showing the same phenotypic features as the animal in the panel.
1025

1026

1027 **Figure 3. LPT controls differentiation across neuronal and epidermal**
1028 **lineages.** Quantification of the number of GABAergic neurons (labeled by
1029 *GAD*) (**a**), dopaminergic neurons (labeled by *TH*) (**b**), acetylcholinergic
1030 neurons (labeled by *chat*) (**c**), serotonergic neurons (labeled by *TPH*) (**d**) and
1031 early (labeled by *NB.21.11e*) and late (labeled by *AGAT-1*) epidermal stem
1032 cell progeny (**e**) at 8 days of regeneration of tail or middle pieces following
1033 *LPT*(RNAi). For each of the comparisons in this figure a 2-tailed ttest
1034 assuming unequal variance was used; a single asterisk indicates $p<0.05$,
1035 while three asterisks indicate $p<0.001$. Error bars represent Standard Error of
1036 the Mean (SEM). Ten animals per condition per experiment were assessed
1037 over the course of two separate experiments.

1038

1039

1040 **Figure 4. Over-proliferation and mitotic cell clustering precedes and**
1041 **accompanies the emergence of outgrowths in *LPT*(RNAi) regenerating**
1042 **animals.** (**a**) Quantification of mitotic cell numbers (mitotic cells labeled by
1043 anti-H3P antibody) at different post-amputation timepoints following
1044 *LPT*(RNAi). The figure is a representative of three repeats of the same
1045 experiment, each using 10 animals per timepoint. The statistical test used was
1046 a 2-tailed ttest assuming unequal variance. The asterisks indicate $p<0.05$.
1047 Error bars represent Standard Error of the Mean (SEM). (**b**) Examples of
1048 middle pieces at the timepoints post-amputation showing significant difference
1049 in mitotic cell counts according to (**a**). 'ph' indicates where the pharynx is in
1050 each piece. The red arrows point towards clusters of mitotic cells in late stage
1051 regenerates (192 hrs/8 days). (**c**) Brightfield examples of middle pieces
1052 forming outgrowths at timepoints after the observation of mitotic clusters in
1053 (**b**). Red arrows point towards outgrowths. (**d**) *LPT*(RNAi) head pieces that do
1054 not contain an outgrowth, but show mitotic cell clusters (indicated by yellow
1055 arrows). One piece shows staining in the eye region, which is an artifact of the
1056 procedure. The lower panel shows a mitotic cell at the border of an outgrowth
1057 and the body in a head piece at 10 days of regeneration following *LPT*(RNAi).
1058 Red arrow points towards the outgrowth and yellow arrow points towards the
1059 mitotic cell. (**e**) An example of a tail piece at 12 days of regeneration having a

1060 mitotic cell-rich cephalic outgrowth following *LPT*(RNAi). Yellow arrows show
1061 the mitotic cells in the outgrowth.
1062
1063
1064 **Figure 5. Stem cells at different stages of commitment are found in**
1065 **outgrowths of *LPT*(RNAi) regenerating animals.** (**a**) A head piece at 18
1066 days of regeneration following *LPT*(RNAi) showing *sigma* stem cells in its
1067 posterior outgrowth. *Sigma* stem cells are double positive for *Smedwi-1* and
1068 the '*sigma pool*' of RNA probes (*Soxp1*, *Soxp2*). Red arrows in the brightfield
1069 images panel point towards the outgrowths and white arrows in the zoomed-in
1070 panel show a double-positive cell and a *Smedwi-1* single-positive cell. (**b**) A
1071 head piece at 10 days of regeneration following *LPT*(RNAi) showing *zeta*
1072 stem cells in its posterior outgrowth. *Zeta* stem cells are double positive for
1073 *Smedwi-1* and the '*zeta pool*' of RNA probes (*zfp-1*, *Soxp3, egr-1*). The red
1074 arrow in the brightfield images panel points towards the outgrowth and the
1075 white arrows in the zoomed-in panel show double-positive cells. (**c**) A middle
1076 piece at 11 days of regeneration following *LPT*(RNAi) showing *gamma* stem
1077 cells in its lateral outgrowth. *Gamma* stem cells are double positive for
1078 *Smedwi-1* and the '*gamma pool*' of RNA probes (*gata4/5/6*, *hnf4*). The red
1079 arrow in the brightfield images panel points towards the outgrowth and the
1080 white arrows in the zoomed-in panel show a double-positive cell, a *Smedwi-1*
1081 single-positive cell and a *gamma pool* single-positive cell. In (**a**), (**b**) and (**c**)
1082 the normal tissue margin is shown via white dashed lines.
1083
1084
1085 **Figure 6. *LPT*(RNAi) results in a cancer-like phenotype.** A summary of the
1086 differentiation and neoblast proliferation data presented, together with a
1087 simplified flowchart illustrating the tested lineages' development under
1088 knockdown conditions. A red cross sign indicates where the defect in a
1089 lineage is detected following *LPT*(RNAi).
1090
1091
1092 **Figure 7. RNA-seq of G2/M stem cells following *LPT*(RNAi) reveals**
1093 **effects on genes enriched in different cell populations.** (**a**) Genes were

1094 classified according to their proportional expression in the X1 (G2/M stem
1095 cells; dark blue), X2 (G1 stem cells and stem cell progeny; light blue) and X
1096 ins (differentiated cells; orange) FACS populations of cells. Genes were
1097 defined as enriched in certain population(s) if more that 50% of their
1098 expression is observed in that population in wildtype or more than 75% of
1099 their expression is observed across two cell populations. Genes not enriched
1100 in either population were classified as 'not enriched'. Each vertical line
1101 represents a gene. Under the population expression enrichment track is a
1102 track with all the significantly up- and down-regulated genes in G2/M stem
1103 cells following *LPT*(RNAi). The genes with fold change >1.5 ($p<0.05$) are
1104 shown in red following a log2 fold change transformation. The genes with fold
1105 change <-1.5 ($p<0.05$) are shown in blue following a log2 fold change
1106 transformation. The Wald's test (as part of the Sleuth software) was used for
1107 assessing differential expression. (**b**) Enrichment for genes in each of the six
1108 classes was calculated for the up- and down-regulated genes' list (red and
1109 blue respectively). Enrichment for Transcription Factors (TFs) was also
1110 performed. The number of genes in each group is indicated in brackets under
1111 the group's name. Numbers in white represent significant enrichment ($p<0.01$)
1112 according to a hypergeometric enrichment test. (**c**) Gene Ontology (GO)
1113 enrichment analysis on the genes significantly up-regulated (red) and down-
1114 regulated (blue) in G2/M stem cells following *LPT*(RNAi). Categories are
1115 sorted by average Log2 fold change of the up- or down-regulated genes
1116 falling in each category. In bold are shown terms that relate to the described
1117 *Mll3/4* loss of function phenotype.
1118
1119
1120 **Figure 8. *LPT*(RNAi) is mainly manifested in changes in H3K4me1 and**
1121 **H3K4me3 around the TSS in G2/M stem cells.** (**a**) Graphs presenting the
1122 average read coverage across the genome for H3K4me3, H3K4me1 and
1123 H3K27me3 (centered on the TSS, showing 2 kb upstream and downstream)
1124 normalised to *Drosophila* S2 signal spike-in. The input coverage is subtracted.
1125 Log2 fold change graphs are also shown for each histone modification, where
1126 signal above zero shows increase following *LPT*(RNAi) and signal below zero
1127 represents a decrease. Three colours are used for different gene classes –

1128 dark blue (genes enriched in G2/M stem cells (X1)), light blue (genes enriched
1129 in G1 stem cells and stem cell progeny (X2)), orange (genes enriched in
1130 differentiated cells (X ins)). Standard deviation is shown by a faded colour
1131 around each line. (**b**) Spearman's rank correlation between changes in RNA-
1132 seq signal and H3K4me1 or H3K4me3 ChIP-seq signal for the region around
1133 the TSS of genes from different enrichment classes (only examples where a
1134 significant correlation exists are shown). The blue line represents a correlation
1135 where no filter for fold change in the RNA-seq data was applied. The green
1136 line shows a correlation where RNA-seq fold change data was filtered for
1137 Log2 fold changes =<-1 and >= +1. Faded areas of the lines represent results
1138 not significant at p<0.001, while darker colours represent results significant at
1139 p<0.001.
1140
1141
1142 **Figure 9. *LPT* regulates the expression of known and putative**
1143 **oncogenes and tumor suppressors.** (**a**) Examples of genes significantly
1144 (p<0.05) mis-regulated in G2/M stem cells following *LPT*(RNAi). RNA-seq fold
1145 change is shown in red (up-regulation) and blue (down-regulation). The
1146 genes' enrichment class is also shown. The ChIP-seq profile for H3K4me3 in
1147 the 2 kb region around the TSS of each gene is presented. Purple colour
1148 represents normalised signal following *LPT*(RNAi) and green colour is used to
1149 show the normalised signal following *GFP*(RNAi). 'TF' stands for 'transcription
1150 factor'. (**b**) *in silico* analysis (www.oncomine.org; ttest, p<0.0001) of *Mll3* and
1151 *pitx2* expression in normal tissue (cerebellum) and cancer tissue
1152 (medulloblastoma). (**c**) *pitx* and *Smedwi-1 in situ* hybridization at 8 days of
1153 regeneration of middle pieces following *LPT*(RNAi). White arrows show
1154 double-positive cells. Cell counts are compared using a 2-tailed ttest
1155 assuming unequal variance. The asterisk indicates p<0.05. Ten animals per
1156 condition were used.
1157
1158
1159 **Figure 10. Double knockdown with *utx* or *pim-1-like* alleviates the**
1160 ***LPT*(RNAi) over-proliferation and outgrowth phenotype.** (**a**) More
1161 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

1195 H3K4me2 on the gene bodies of actively transcribed genes. MLL1/2/Trithorax
1196 COMPASS-like complex – 1) a role in transcriptional activation of Hox genes
1197 via trimethylating H3K4 on TSS of their promoters and 2) MLL2 is involved in
1198 trimethylation of H3K4 on TSS of bivalent promoters. MLL3/4/LPT/Trr – 1) role
1199 in hormone-dependent transcription – when the Nuclear Receptor protein
1200 (NR) is bound to the DNA Hormone Response Element (HRE) upon Hormone
1201 Ligand (HL) detection, MLL3/4/LPT/Trr complex binds the nuclear receptor
1202 and serves as its co-activator via trimethylating H3K4 and promoting active
1203 transcription on selected loci; 2) a switch between inactive and active
1204 enhancer states where MLL3/4/LPT/Trr complex deposits H3K4me1 on both
1205 active and inactive enhancers; upon UTX recruitment, it demethylates
1206 H3K27me3 and allows for CBP/p300 to acetylate H3K27 and activate the
1207 enhancer; 3) a switch between active and inactive promoters -
1208 MLL3/4/LPT/Trr complex bound to TSS deposits H3K4me1 on the TSS and
1209 around it, leads to repressed transcription of the gene; when H3K4me1 is
1210 depleted from the TSS and another complex performs trimethylation of H3K4
1211 on TSS, this is correlated with activated transcription. (**d**) Schematic
1212 representation of planarian COMPASS and COMPASS-like core subunits.
1213 SMED-LPT (in red) is characterized in the present study. (**e**) Planarian
1214 COMPASS and COMPASS-like core subunits' expression in the three
1215 populations of cells sortable by fluorescence-activated cell sorting (FACS)
1216 (X1=G2/M stem cells, X2=G1 stem cells and stem cell progeny, X
1217 ins=differentiated cells) according to RNA-seq data. (**f**) Known defects after
1218 RNAi-mediated knockdown of core COMPASS and COMPASS-like subunits
1219 in planarians.
1220
1221
1222 **Additional File 2 (PDF) Planarian *Mll3/4* genes are expressed in**
1223 **neoblasts and neoblast progeny and colocalise with each other.** (**a**)
1224 Protein alignment of conserved regions of COMPASS-like families' core
1225 proteins. Asterisks indicate complete conservation in all sequences, while
1226 black boxes are drawn around areas of conservation specific to the
1227 MLL3/4/Trithorax-related family. Colours represent similarity of amino acids.
1228 The image was produced using MEGA.5.2 software. (**b**) *LPT*, *trr-1* and *trr-2*

1229 expression in wildtype and irradiated 3 day-regenerating head and middle
1230 pieces. Arrows in head pieces point towards expression in the forming
1231 pharynx, while arrows in middle pieces point towards expression in the
1232 forming brain. (**c**) *LPT*, *trr-1* and *trr-2* expression in intact animals following
1233 *GFP*(RNAi) or following *H2B*(RNAi). *Porcupine-1* and *Smedwi-2* were used as
1234 a negative and positive control respectively. (**d**) *LPT*, *trr-1* and *trr-2* co-
1235 expression in the head region (as shown by the schematics). White arrows
1236 point towards cells showing colocalisation.
1237
1238
1239 **Additional File 3 (PDF) Phenotype scoring of *Mll3/4* knockdown**
1240 **planarians during regeneration and homeostasis.** (**a**) Proportion of head,
1241 middle and tail regenerates exhibiting particular phenotypic characteristics
1242 following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**b**) Survival curves for head,
1243 middle and tail pieces following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**c**)
1244 Proportion of intact (homeostatic) animals with particular phenotypic
1245 characteristics following *LPT*(RNAi), *trr-1*(RNAi) and *trr-2*(RNAi). (**d**) Survival
1246 curves for homeostatic animals following *LPT*(RNAi), *trr-1*(RNAi) and *trr-*
1247 *2*(RNAi).
1248
1249
1250 **Additional File 4 (PDF) *Trr-1*(RNAi) and *trr-2*(RNAi) lead to mild**
1251 **differentiation defects during regeneration.** (**a**) A schematic showing the
1252 amputation of RNAi worms into head (H), middle (M) and tail (T) pieces in
1253 order to observe regeneration of different structures. The time-course of the
1254 experiments on *Mll3/4* knockdown animals is depicted underneath the worm
1255 schematic. A total of 9 days of dsRNA microinjection-mediated RNAi was
1256 followed by amputation on the 10th day and observation of regeneration. (**b**)
1257 Head, middle and tail pieces following *trr-1*(RNAi), *trr-2*(RNAi) or control
1258 *GFP*(RNAi) at day 8 of regeneration. Yellow arrows point towards the
1259 regenerative defects – smaller blastema, delayed eye formation or posterior
1260 bloating. (**c**) Head, middle and tail pieces following *trr-1*(RNAi), *trr-2*(RNAi) or
1261 control *GFP*(RNAi) at day 14 of regeneration. (**d**) Central nervous system
1262 (CNS) maintenance and recovery at 8 days of middle piece regeneration, as

1263 labeled by CNS-specific anti-SYNORF-1 antibody, following *trr-1*(RNAi) or *trr-*
1264 *2*(RNAi). (**e**) Gut maintenance and recovery at 8 days of middle piece
1265 regeneration, as labeled by *porcupine-1*, following *trr-1*(RNAi) or *trr-2*(RNAi).
1266 (**f**) Pharynx recovery at 8 days of head piece regeneration, as labeled by
1267 *laminin*, following *trr-1*(RNAi) or *trr-2*(RNAi). Numbers at the top of each piece
1268 represent number of animals in that condition showing the same phenotypic
1269 features as the animal in the panel.
1270
1271
1272 **Additional File 5 (PDF) *Trr-1/ trr-2* double knockdown results in more**
1273 **prevalent and accelerated outgrowth formation compared to *LPT*(RNAi).**
1274 (**a**) Head, middle and tail pieces at 3 days of regeneration following
1275 *GFP/LPT*(RNAi), *GFP/trr-1*(RNAi), *GFP/trr-2*(RNAi), *trr-1/trr-2*(RNAi) and
1276 *GFP/GFP*(RNAi). Red arrows point towards outgrowths. (**b**) Percentage of
1277 head, middle and tail regenerating pieces developing outgrowths throughout
1278 their life-time. (**c**) Survival curves of head, middle and tail regenerating pieces.
1279 The *GFP/GFP*(RNAi) line overlaps with *GFP/trr-1*(RNAi) and *GFP/ trr-*
1280 *2*(RNAi). Ten animals per condition were used.
1281
1282
1283 **Additional File 6 (PDF) Number if stem cells and early stem cell progeny**
1284 **is unchanged following knockdown of *LPT*.** The pre-pharyngeal area of
1285 middle pieces at 8 days of regeneration was used for this experiment. Stem
1286 cells are labeled with *H2B* and early neoblast progeny cells are *H2B*-
1287 negative/anti-SMEDWI-1 antibody-positive cells. Numbers of stem cells and
1288 progeny cells between *LPT*(RNAi) animals and controls were not significantly
1289 different (ns). A 2-tailed ttest assuming unequal variance was used. Ten
1290 animals per condition were processed.
1291
1292
1293 **Additional File 7 (PDF) *Trr-2*(RNAi) regenerating animals produce less**
1294 **GABAergic and dopaminergic neurons.** Quantification of the number of
1295 GABAergic neurons (labeled by *GAD*) (**a**), dopaminergic neurons (labeled by
1296 *TH*) (**b**), serotonergic neurons (labeled by *TPH*) (**c**), acetylcholinergic neurons

1297 (labeled by *chat*) (**d**) and early (labeled by *NB.21.11e*) and late (labeled by
1298 *AGAT-1*) epidermal stem cell progeny (**e**) at 8 days of regeneration of tail or
1299 middle pieces following *trr-1*(RNAi) or *trr-2*(RNAi). For each of the
1300 comparisons in this figure a 2-tailed ttest assuming unequal variance was
1301 used; a single asterisk indicates $p<0.05$. Error bars represent Standard Error
1302 of the Mean (SEM). Ten animals per condition per experiment were assessed
1303 over the course of two separate experiments.
1304
1305
1306 **Additional File 8. (PDF) Tubule-associated protonephridia and cilia cell**
1307 **regeneration is not affected by *Mll3/4* knockdown.** (**a**) Recovery of dorsal
1308 and ventral cilia in middle pieces at 8 days of regeneration following
1309 *LPT*(RNAi), *trr-1*(RNAi) or *trr-2*(RNAi), as labeled by anti-acetylated tubulin
1310 antibody. (**b**) Recovery and maintenance of tubule-associated protonephridia
1311 cells (labeled by *CAVII-1*) in middle pieces at 8 days of regeneration following
1312 *LPT*(RNAi), *trr-1*(RNAi) or *trr-2*(RNAi). The white dashed line indicates where
1313 the anterior regenerated body parts following amputation should be. The
1314 graph compares numbers of *CAVII-1*-positive cells in the newly recovered
1315 regions, as well as in the whole body, between *Mll3/4* genes knockdown
1316 conditions and controls. For each of the comparisons in this figure a 2-tailed
1317 ttest assuming unequal variance was used; 'ns' stands for 'not significant'.
1318 Error bars represent Standard Error of the Mean (SEM). Ten animals per
1319 condition were used.
1320
1321
1322 **Additional File 9. (PDF) 'Inchworming' following *Mll3/4* knockdown is**
1323 **due to serotonin deficiency.** Locomotive defects in *Mll3/4* knockdown
1324 animals are shown. The number of regenerating head pieces per condition
1325 exhibiting the respective locomotive defect before and after 45-minute-long
1326 serotonin hydrochloride treatment is shown in a table. Movie still shots are
1327 shown for worms in each treatment. There was one-second interval between
1328 the chosen still shots. The dashed line indicates the locomotive progress that
1329 the worm had achieved in 3 seconds (from first to last shot). Steeper line
1330 indicates faster movement. White numbers at the bottom right corners

1331 represent number of animals per condition showing the illustrated behaviour.
1332 Ten animals per condition were used.

1333

1334 **Additional File 10. (PDF) *Mll3/4* knockdown leads to changes in mitotic**
1335 **activity during regeneration and homeostasis.** (**a**) Mitotic cell number
1336 fluctuations during regeneration following *trr-1*(RNAi), *trr-2*(RNAi) and
1337 *GFP*(RNAi). (**b**) Mitotic cell number fluctuations during homeostatic
1338 observations following *LPT*(RNAi), *trr-1*(RNAi), *trr-2*(RNAi) and *GFP*(RNAi).
1339 For each of the comparisons in this figure a 2-tailed ttest assuming unequal
1340 variance was used; a single asterisk indicates $p<0.05$. Error bars represent
1341 Standard Error of the Mean (SEM). Ten animals per condition were assessed.

1342

1343 **Additional File 11. (PDF) Non-mitotic stem cells are present in**
1344 **outgrowths of animals following *LPT*(RNAi).** A head piece (containing
1345 outgrowths) at 10 days of regeneration following *LPT*(RNAi) stained with *H2B*
1346 RNA probe and anti-H3P mitotic cell antibody. Stem cells found outside the
1347 usual stem cell compartment are indicated via white arrows. The border of the
1348 usual neoblast compartment is indicated with a white dashed line.

1349

1350 **Additional File 12. (PDF) *Sigma*, *zeta* and *gamma* neoblast numbers are**
1351 **unchanged following *LPT*(RNAi).** (**a, d**) Cells in pre-pharyngeal regions of
1352 middle pieces at 8 days of regeneration following *LPT*(RNAi) labeled by the
1353 *sigma pool* of RNA probes (*Soxp1*, *Soxp2*) and *Smedwi-1*. White arrows point
1354 towards *sigma* neoblasts (double-positive for *sigma pool* and *Smedwi-1*). (**b**,
1355 **e**) Cells in pre-pharyngeal regions of middle pieces at 8 days of regeneration
1356 following *LPT*(RNAi) labeled by the *zeta pool* of RNA probes (*zfp-1*, *Soxp3,*
1357 *egr-1*) and *Smedwi-1*. White arrows point towards *zeta* neoblasts (double-
1358 positive for *zeta pool* and *Smedwi-1*). (**c**, **f**) Cells in pre-pharyngeal regions of
1359 middle pieces at 8 days of regeneration following *LPT*(RNAi) labeled by the
1360 *gamma pool* of RNA probes (*gata4/5/6*, *hnf4*) and *Smedwi-1*. White arrows

1361 point towards *gamma* neoblasts (double-positive for *gamma pool* and
1362 *Smedwi-1*). (**g**) Overall number of *Smedwi-1*-positive cells (regardless of
1363 colocalisation with other markers) in the pre-pharyngeal region of middle
1364 pieces at 8 days of regeneration following *LPT*(RNAi). The statistical
1365 comparisons in this figure were performed via 2-tailed ttest assuming unequal
1366 variance. 'ns' stands for 'not significant'. Ten worms per condition were
1367 processed.

1368

1369 **Additional File 13. (PDF) *LPT*(RNAi) results in disorganized outgrowth-**
1370 **focused expression of epidermal precursor markers, epithelial disarray**
1371 **and hypertrophy and changes of nuclear morphology.** (**a**) Anterior part
1372 (containing an outgrowth) of a tail piece at 18 days of regeneration following
1373 *LPT*(RNAi) labeled with *NB.21.11e* and *AGAT-1* epidermal precursor markers.
1374 'CG' stands for 'cephalic ganglia'. (**b**) The epidermal layer (stained with
1375 Hoechst 33342) of a tail piece at 10 days of regeneration following *LPT*(RNAi)
1376 compared to control (**c**) The nuclear area of 20 epithelial cells per
1377 experimental and control condition was compared via 2-tailed ttest assuming
1378 unequal variance. Triple asterisk indicates $p<0.001$. (**d**) Nuclear morphology
1379 comparison between a 10-day *LPT*(RNAi) and control regenerate. Samples
1380 were stained with Hoechst 33342. Yellow arrows point towards misshapen
1381 nuclei.

1382

1383 **Additional File 14. (PDF) Histone modification ChIP-seq profiles at**
1384 **promoter-proximal regions of different classes of genes.** (**a**), (**b**) and (**c**)
1385 show histone modification patterns for H3K4me3, H3K4me1 and H3K27me3
1386 respectively. ChIP-seq signal is shown in black. Six groups of genes are
1387 presented – enriched >50% in X1 (G2/M stem cells) shown by dark blue,
1388 enriched >50% in X2 (G1 stem cells and stem cell progeny) shown in light
1389 blue, enriched >50% in X ins (differentiated cells) shown in orange, genes
1390 enriched >75% in X1/X2, >75% in X2/X ins and 'not enriched'. Histone
1391 modification graphs are centered on the Transcriptional Start Site (TSS) with

1392 2.5 kb shown upstream and downstream.

1393

1394 **Additional File 15. (PDF) The expression of stem cell-enriched genes**
1395 **mis-regulated following *LPT*(RNAi) is inversely correlated with H3K4me1**
1396 **TSS-proximal levels.** (**a**) Graphs presenting the average read coverage
1397 across the genome for H3K4me3, H3K4me1 and H3K27me3 (centered on the
1398 TSS, showing 2 kb upstream and downstream) normalised to *Drosophila* S2
1399 signal spike-in. The input coverage is subtracted. Log2 fold change graphs
1400 are also shown for each histone modification, where signal above zero shows
1401 increase following *LPT*(RNAi) and signal below zero represents a decrease.
1402 Three colours are used for different gene classes – green (genes enriched in
1403 X1/X2 cells), red (genes enriched in X2/X ins cells), black (genes not enriched
1404 in any population of cells). Standard deviation is shown as a faded colour
1405 around each line. (**b**) Log2 fold change of signal around the TSS across
1406 different histone marks and gene classes following *LPT*(RNAi). Blue
1407 represents genes down-regulated following *LPT*(RNAi) and red – up-
1408 regulated. Standard deviation is shown by a faded colour around each line.
1409

1410 **Additional File 16. (PDF) *LPT* regulates the expression of cancer- and**
1411 **development-associated genes.** Examples of genes significantly ($p<0.05$)
1412 mis-regulated in G2/M stem cells following *LPT*(RNAi). RNA-seq fold change
1413 is shown in red (up-regulation) and blue (down-regulation). The genes'
1414 enrichment class is also shown. The ChIP-seq profile for a histone
1415 modification in the 2 kb region around the TSS of each gene is presented.
1416 Purple colour represents normalised signal following *LPT*(RNAi) and green
1417 colour is used to show the normalised signal following *GFP*(RNAi). Depending
1418 on the gene enrichment class, H3K4me1 (**a**) or H3K4me3 (**b**) ChIP-seq signal
1419 is presented for each gene (based on previous Spearman's rank correlation
1420 analyses in **Figure 8**). Bold font of a gene name illustrates an example where
1421 there is a correlation between ChIP-seq and RNA-seq data. 'TF' stands for
1422 'transcription factor'.

1423

1424 **Additional File 17. (PDF) Phenotype scoring of double knockdown**
1425 **(*GFP/LPT, GFP/pim-1, GFP/pim-1-like, GFP/utx, pim-1/pim-1-like,***
1426 ***LPT/pim-1, LPT/pim-1-like, LPT/utx* and *GFP/GFP*) planarians during**
1427 **regeneration.** Proportion of head, middle and tail regenerates exhibiting
1428 particular phenotypic characteristics following a double knockdown. The
1429 proportion of animals forming outgrowths is given regardless of regenerating
1430 piece identity.

1431

1432 **Additional File 18. (PDF) The histone modifications antibodies used for**
1433 **ChIP-seq experiments are specific.** (**a**) Western blot with loading control
1434 anti-H3 (unmodified histone H3) and anti-H3K4me1 on whole animal protein
1435 lysate from *GFP*(RNAi) and *LPT*(RNAi) samples. (**b**) Western blot with loading
1436 control anti-H3 (unmodified histone H3) and anti-H3K4me3 on whole animal
1437 protein lysate from *GFP*(RNAi) and *set1*(RNAi) samples. (**c**) Western blot with
1438 loading control anti-H3 (unmodified histone H3) and anti-H3K27me3 on whole
1439 animal protein lysate from *GFP*(RNAi) and *utx(RNAi)* samples.

1440

1441 **Additional File 19. (PDF) Summary of planarian ChIP-seq procedure.**
1442 Three day-regenerating planarians were dissociated into single cells. Cells
1443 were stained with Hoechst 34580 and Calcein AM in order to visualize cell
1444 populations according to nuclear size and cytoplasmic complexity. The X1
1445 (G2/M) stem cells (magenta) were sorted and mixed with 4% *Drosophila* S2
1446 cells. Cells were crosslinked with 1% Formaldehyde and sonicated.
1447 Immunoprecipitation with anti-H3K4me3, anti-H3K4me1 and anti-H3K27me3
1448 antibodies followed. Samples were reverse-crosslinked and libraries were
1449 prepared using NEBNext Ultra II library preparation kit.

1450

1451 **Additional File 20.** (.xlsx) **Differentially expressed loci following**

1452 ***LPT*(RNAi).** Each row represents one locus that was differentially expressed
1453 with a p-value less than 0.05 and fold change <-1.5 or >1.5. The Wald's test
1454 (as part of the Sleuth software) was used for assessing differential
1455 expression. The top BLAST hit (with e-value) and the common model
1456 organism top BLAST hit is also provided for each locus.

1457

1458 **Additional File 21.** (.xlsx) **Primer sequences.** All primers are given in 5'->3'
1459 orientation. 'F' and 'R' stand for 'forward' and 'reverse' primer respectively.

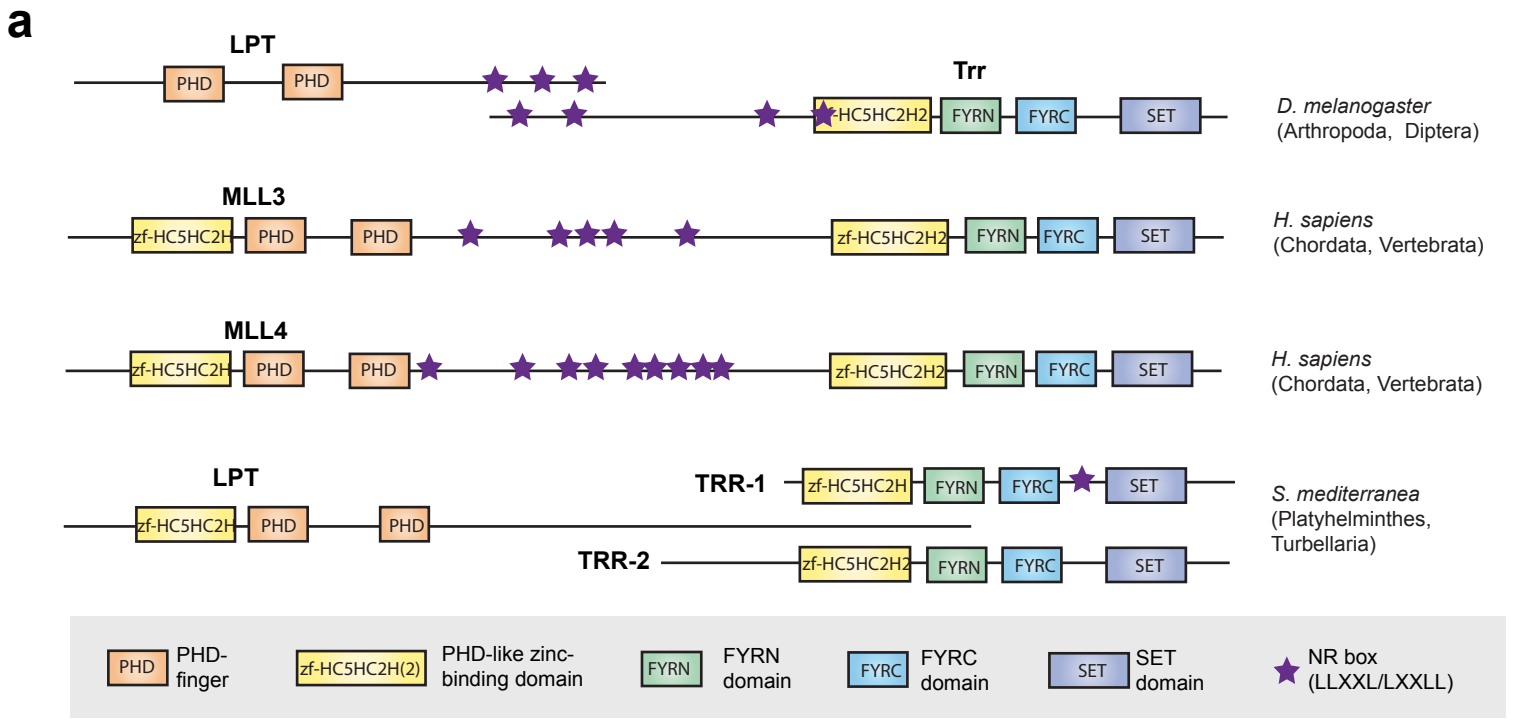
1460

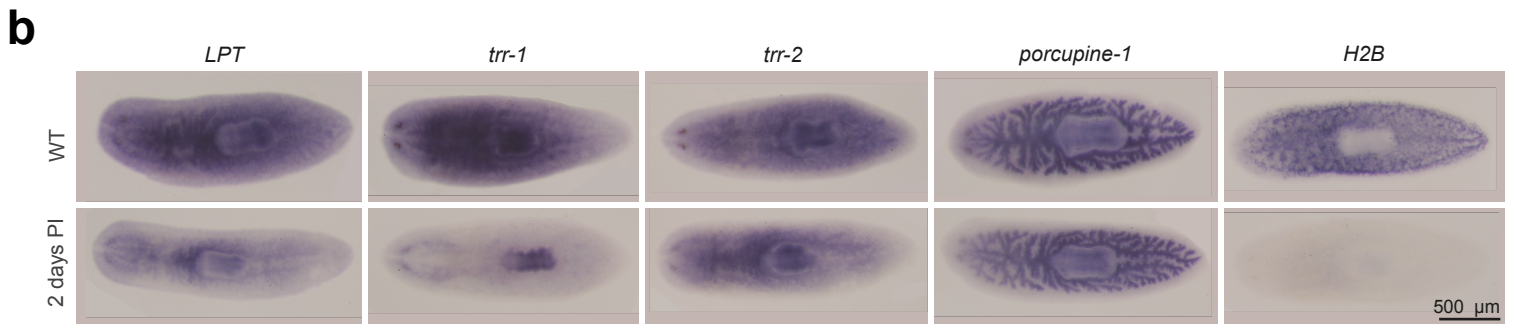
1461 **Additional File 22.** (html) **Supplementary Python Notebook.** Provides
1462 details on the ChIP-seq and RNA-seq bioinformatics analyses.

1463

1464

Figure 1


a
LPT
PHD
PHD
Trr
zf-HC5HC2H2
FYRN
FYRC
SET
D. melanogaster
(Arthropoda, Diptera)
MLL3
zf-HC5HC2H
PHD
PHD
zf-HC5HC2H2
FYRN
FYRC
SET
H. sapiens
(Chordata, Vertebrata)
MLL4
zf-HC5HC2H
PHD
PHD
zf-HC5HC2H2
FYRN
FYRC
SET
H. sapiens
(Chordata, Vertebrata)
LPT
zf-HC5HC2H
PHD
PHD
TRR-1
zf-HC5HC2H
FYRN
FYRC
SET
TRR-2
zf-HC5HC2H2
FYRN
FYRC
SET
S. mediterranea
(Platyhelminthes, Turbellaria)
PHD PHD-finger
zf-HC5HC2H(2) PHD-like zinc-binding domain
FYRN FYRN domain
FYRC FYRC domain
SET SET domain
NR box (LLXXL/LXXLL)


b
LPT
trr-1
trr-2
porcupine-1
H2B
WT
2 days PI
500 µm

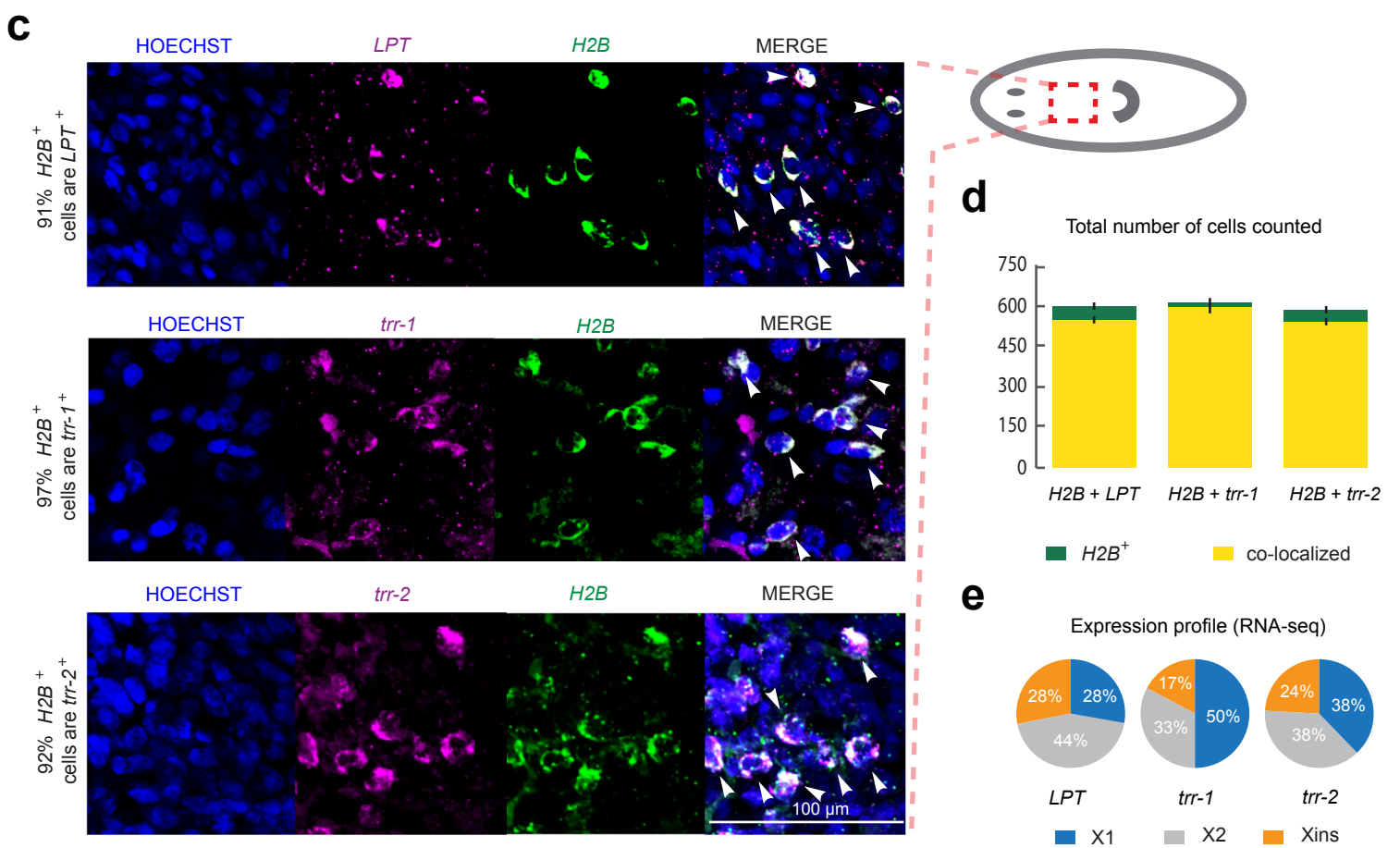

c
HOECHST
LPT
H2B
MERGE
91% H2B+ cells are LPT+
HOECHST
trr-1
H2B
MERGE
97% H2B+ cells are trr-1+
HOECHST
trr-2
H2B
MERGE
92% H2B+ cells are trr-2+
100 µm
d
Total number of cells counted
750
600
450
300
150
0
H2B + LPT
H2B + trr-1
H2B + trr-2
H2B+
co-localized
e
Expression profile (RNA-seq)
28%
28%
44%
17%
50%
33%
24%
38%
38%
LPT
trr-1
trr-2
X1
X2
Xins

Figure 2

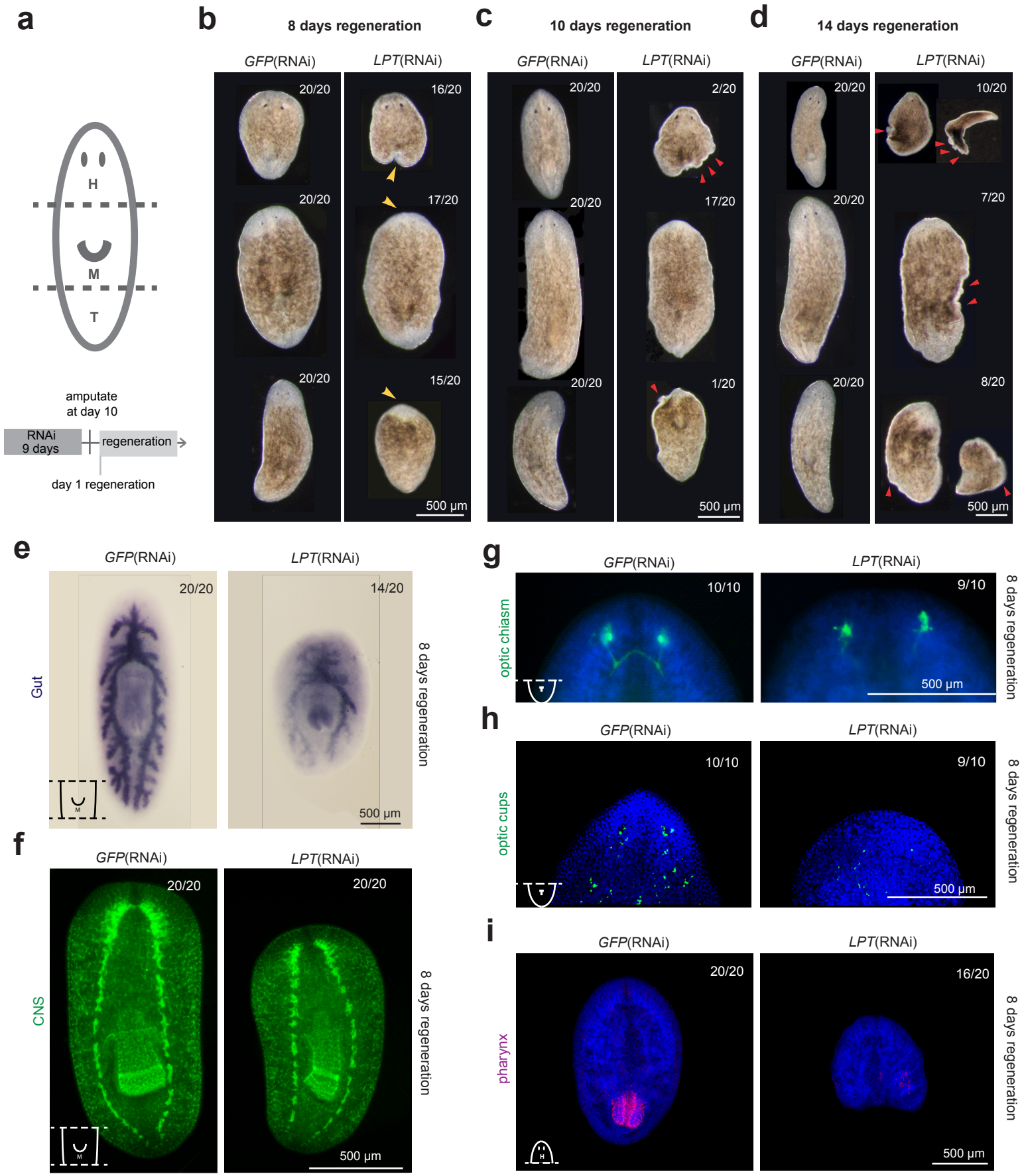
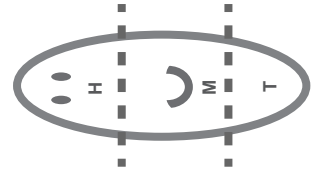
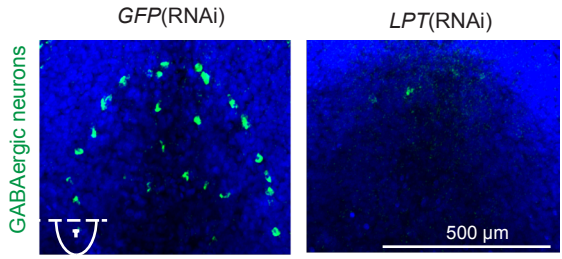

a
H
M
T
amputate
at day 10
RNAi
9 days
regeneration
day 1 regeneration
b
8 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
16/20
20/20
17/20
20/20
15/20
500 µm
c
10 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
2/20
20/20
17/20
20/20
1/20
500 µm
d
14 days regeneration
GFP(RNAi)
LPT(RNAi)
20/20
10/20
20/20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

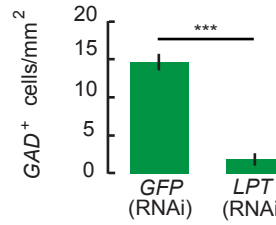
Figure 3

8 days regeneration

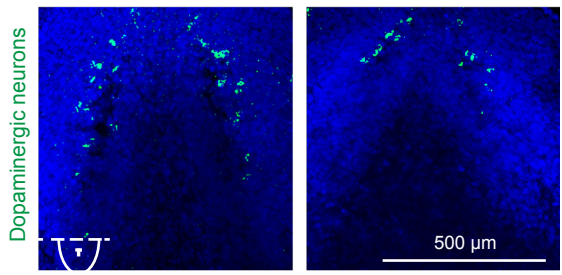


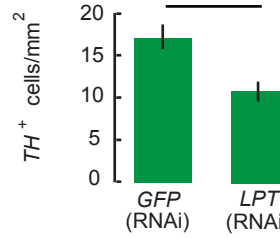
a



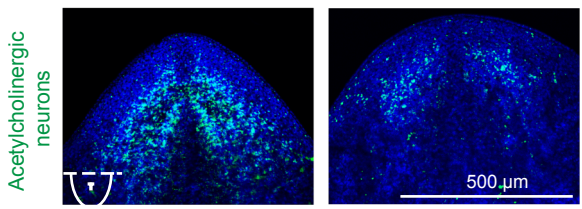


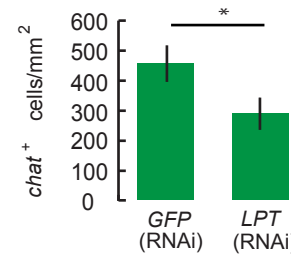
b



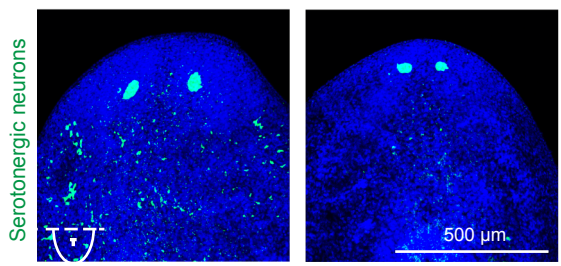


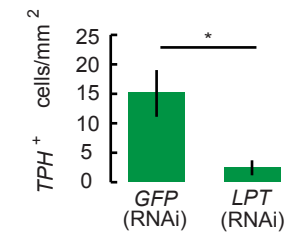
c



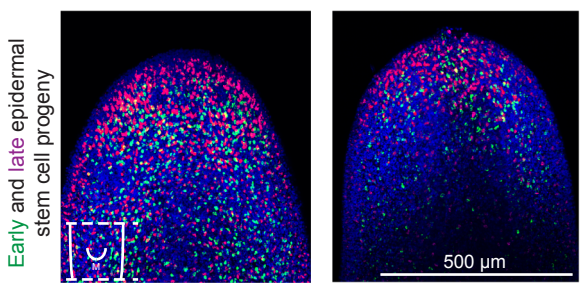


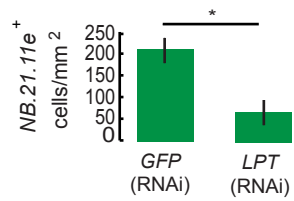
d





e





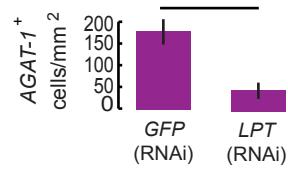
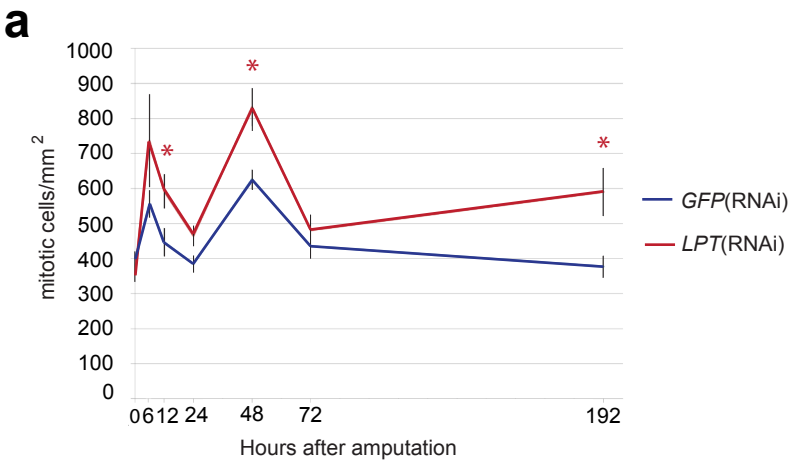
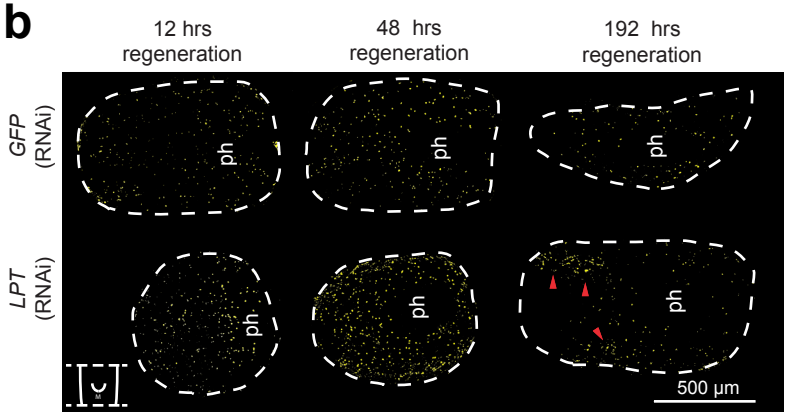
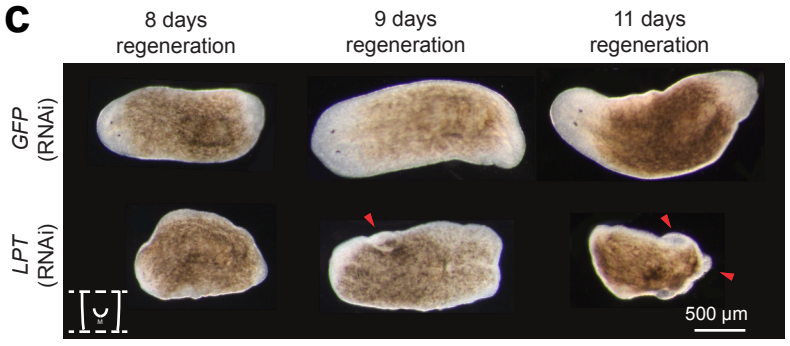
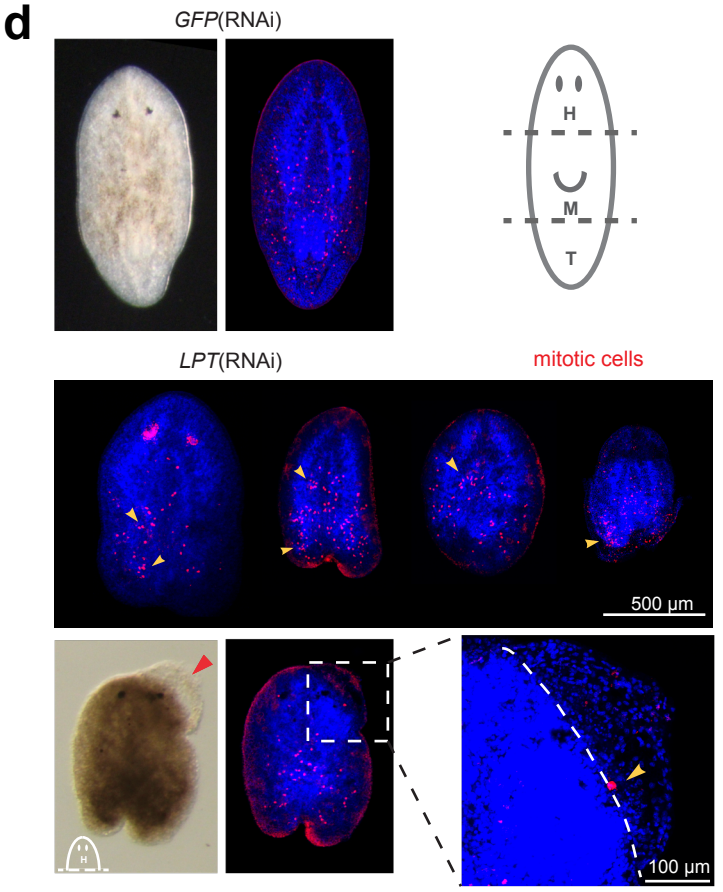


Figure 4


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
100 µm
10 days regeneration

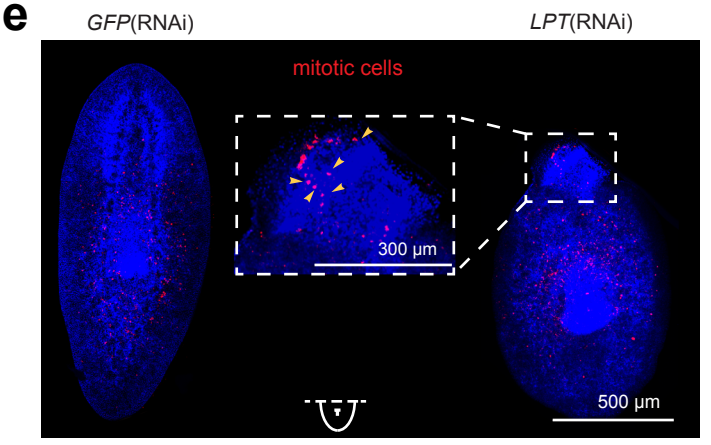

e
GFP(RNAi)
LPT(RNAi)
mitotic cells
300 µm
500 µm
12 days regeneration

Figure 5

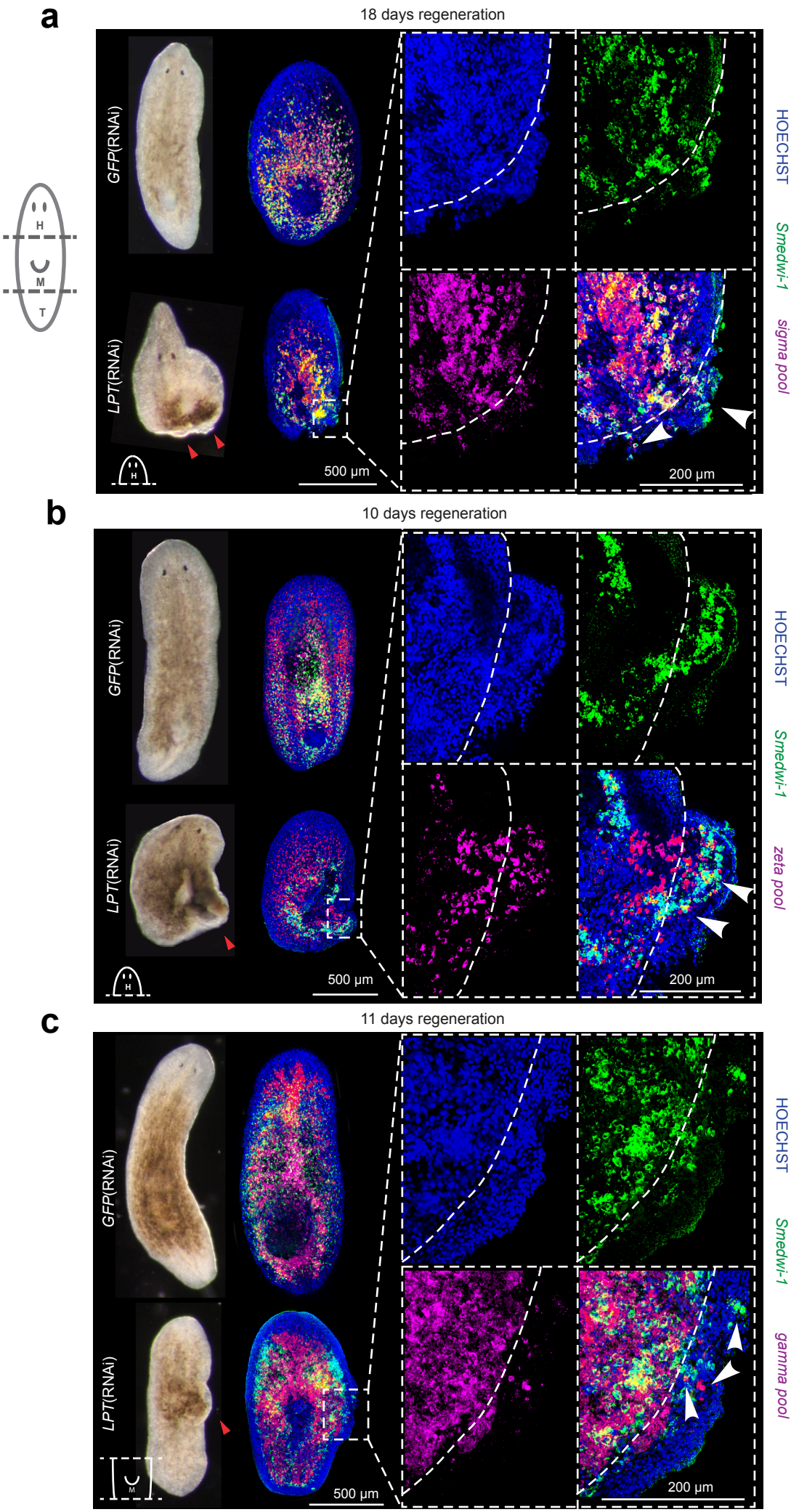

a
18 days regeneration
GFP(RNAi)
LPT(RNAi)
H
M
T
500 µm
200 µm
HOECHST
Smedwi-1
sigma pool
b
10 days regeneration
GFP(RNAi)
LPT(RNAi)
H
500 µm
200 µm
HOECHST
Smedwi-1
zeta pool
c
11 days regeneration
GFP(RNAi)
LPT(RNAi)
M
500 µm
200 µm
HOECHST
Smedwi-1
gamma pool

Figure 6

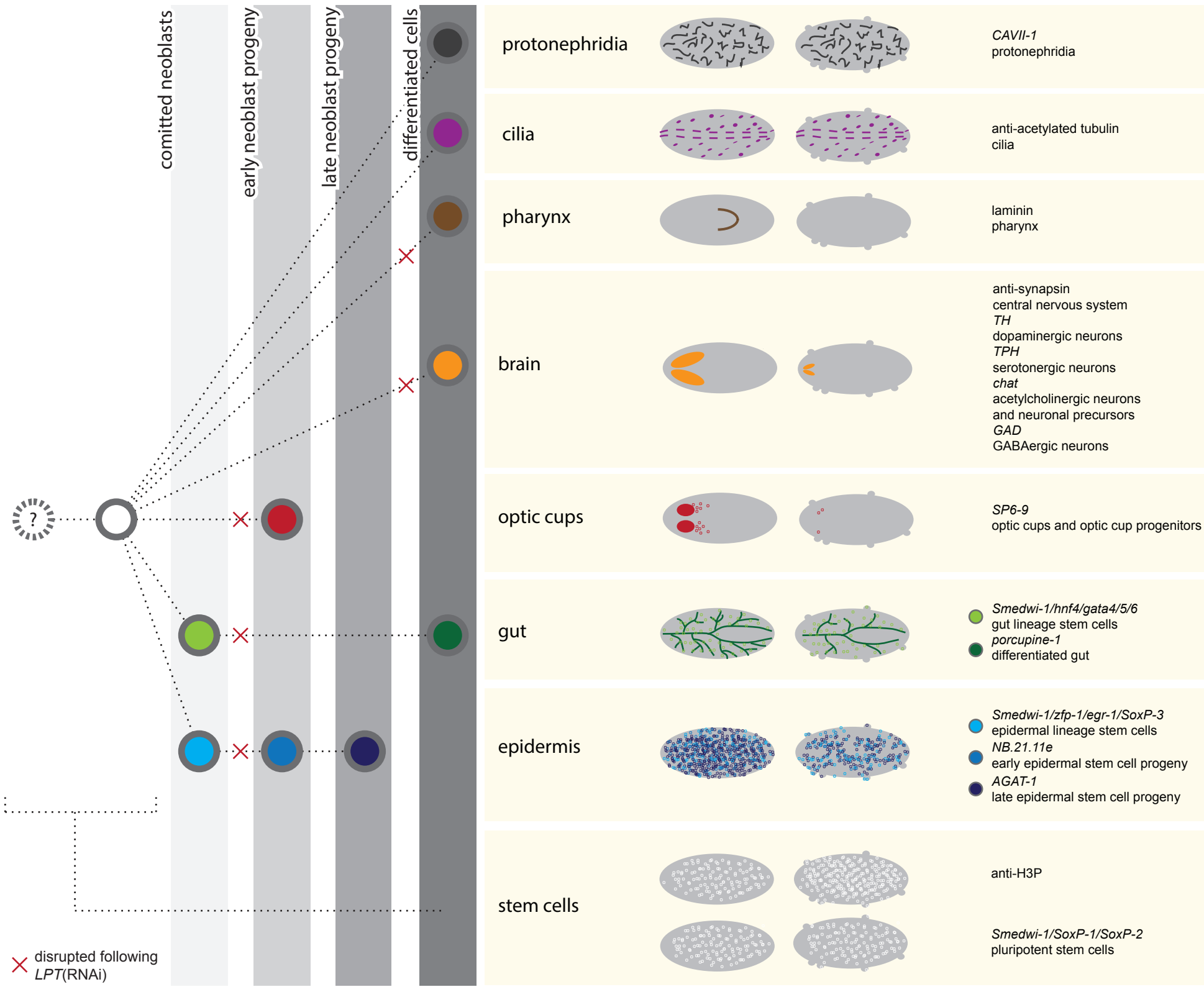
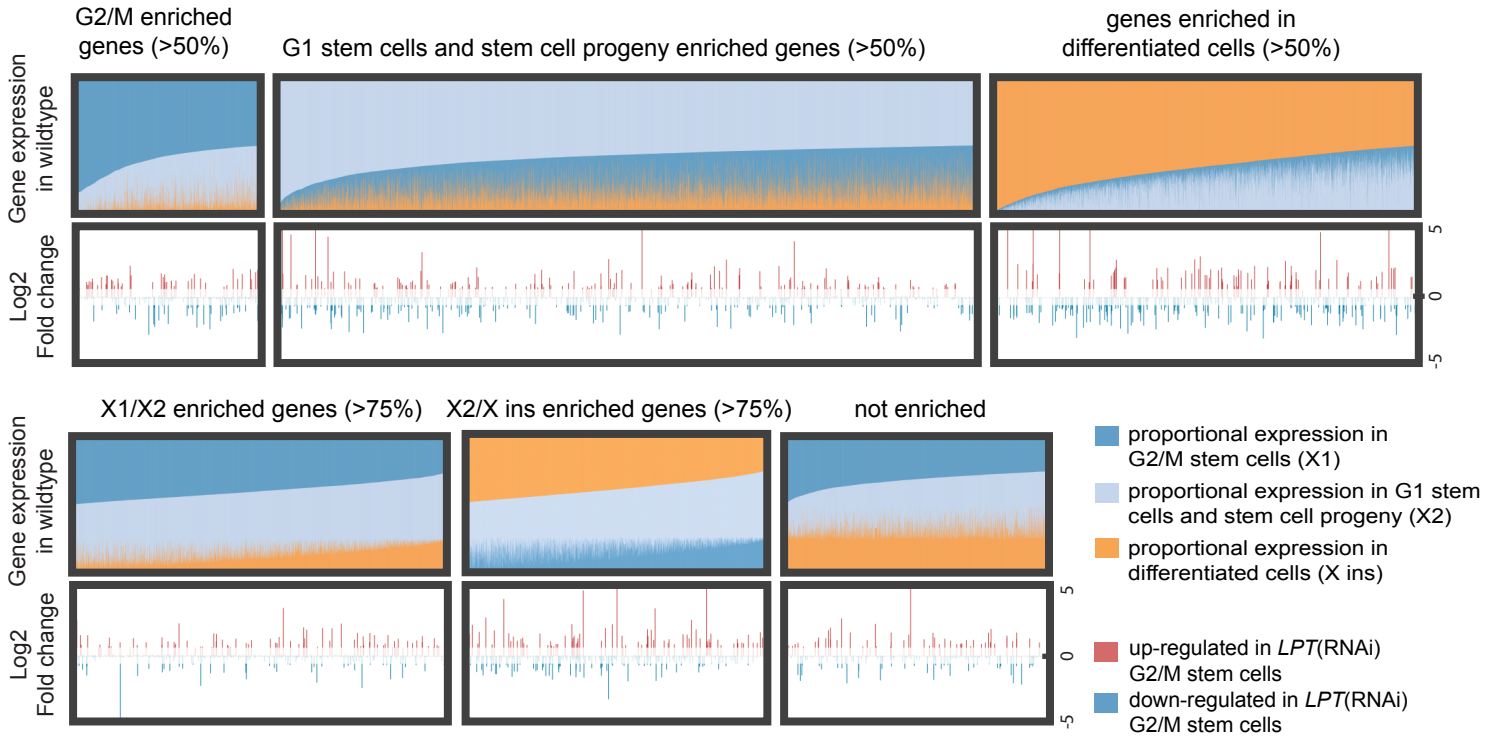

lineage
GFP(RNAi)
LPT(RNAi)
evidence
comitted neoblasts
early neoblast progeny
late neoblast progeny
differentiated cells
?
protonephridia
CAVII-1
protonephridia
cilia
anti-acetylated tubulin
cilia
pharynx
laminin
pharynx
brain
anti-synapsin
central nervous system
TH
dopaminergic neurons
TPH
serotonergic 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)

Figure 7

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

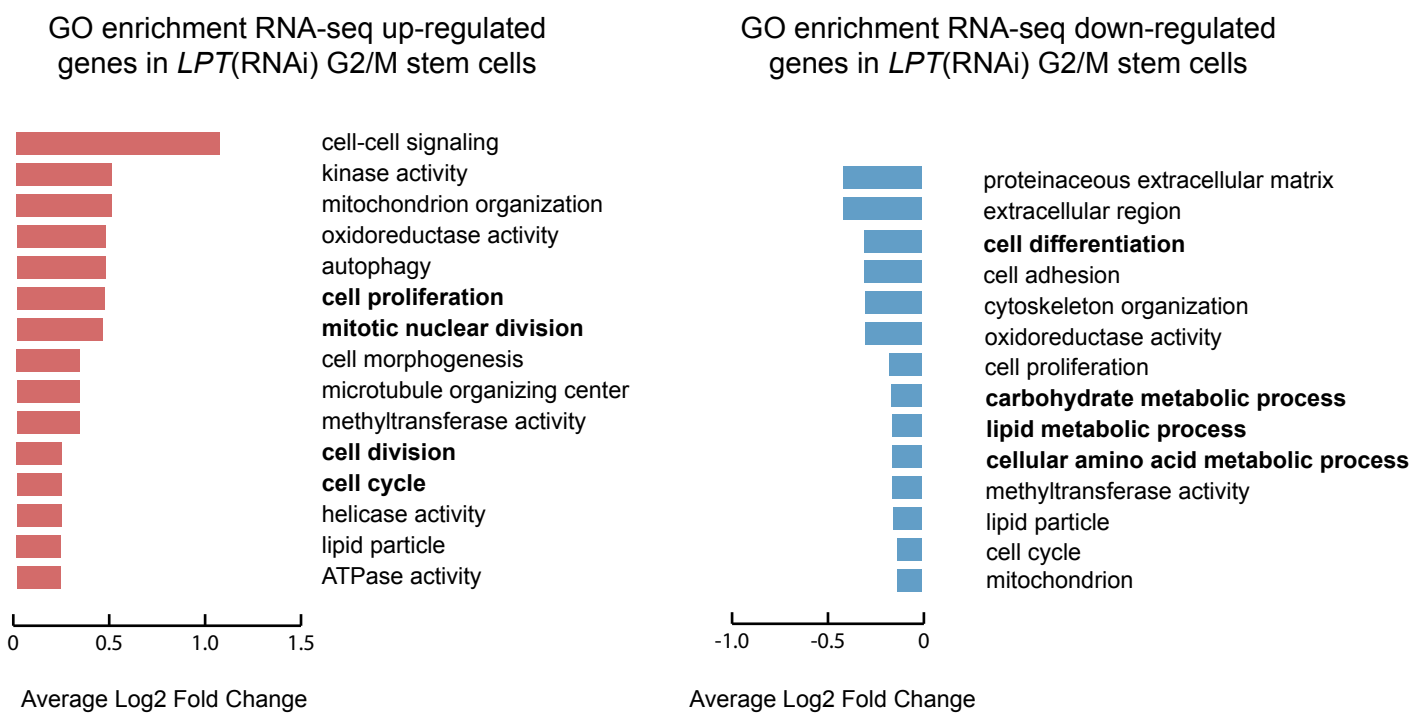
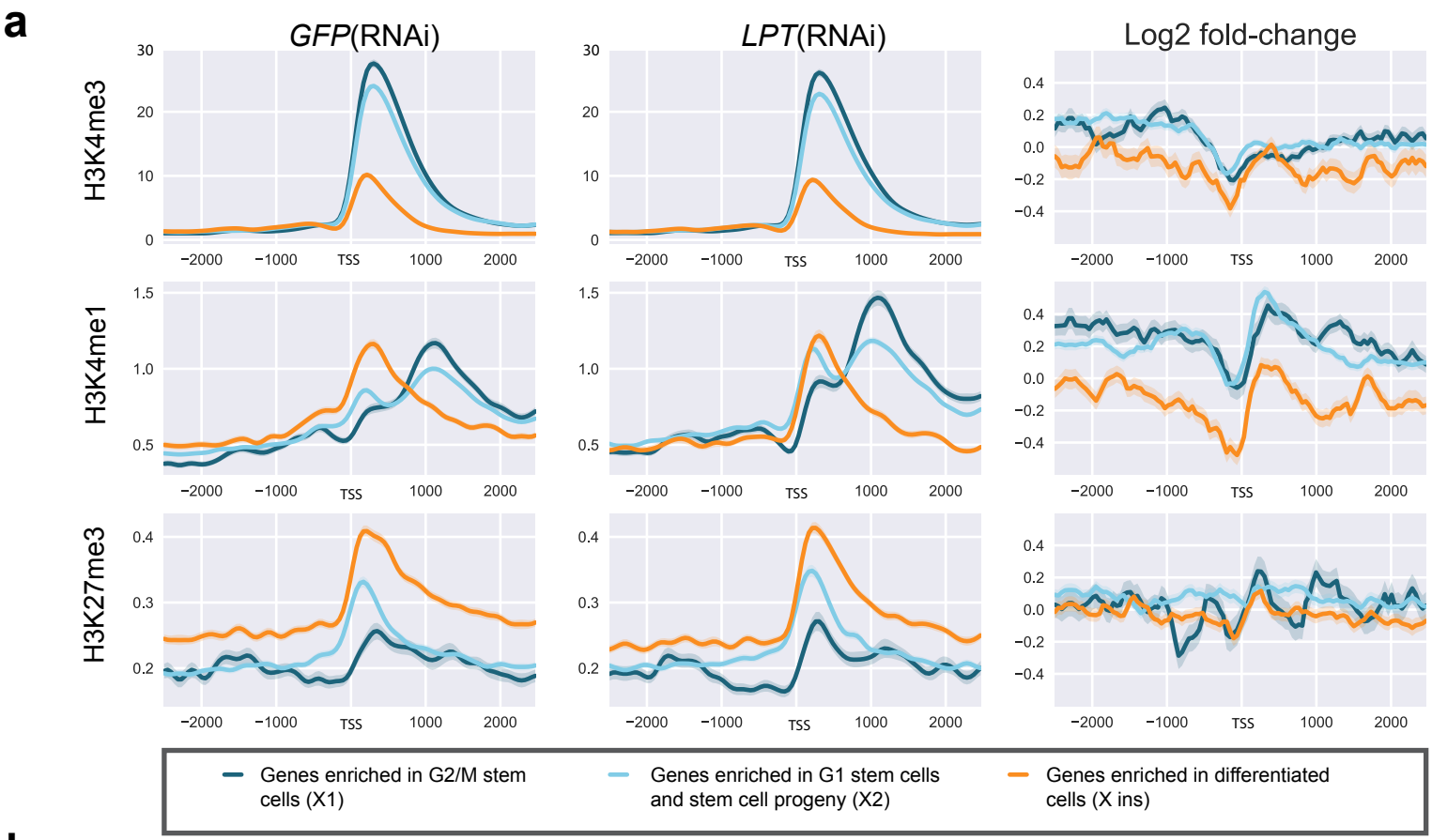


Figure 8


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)

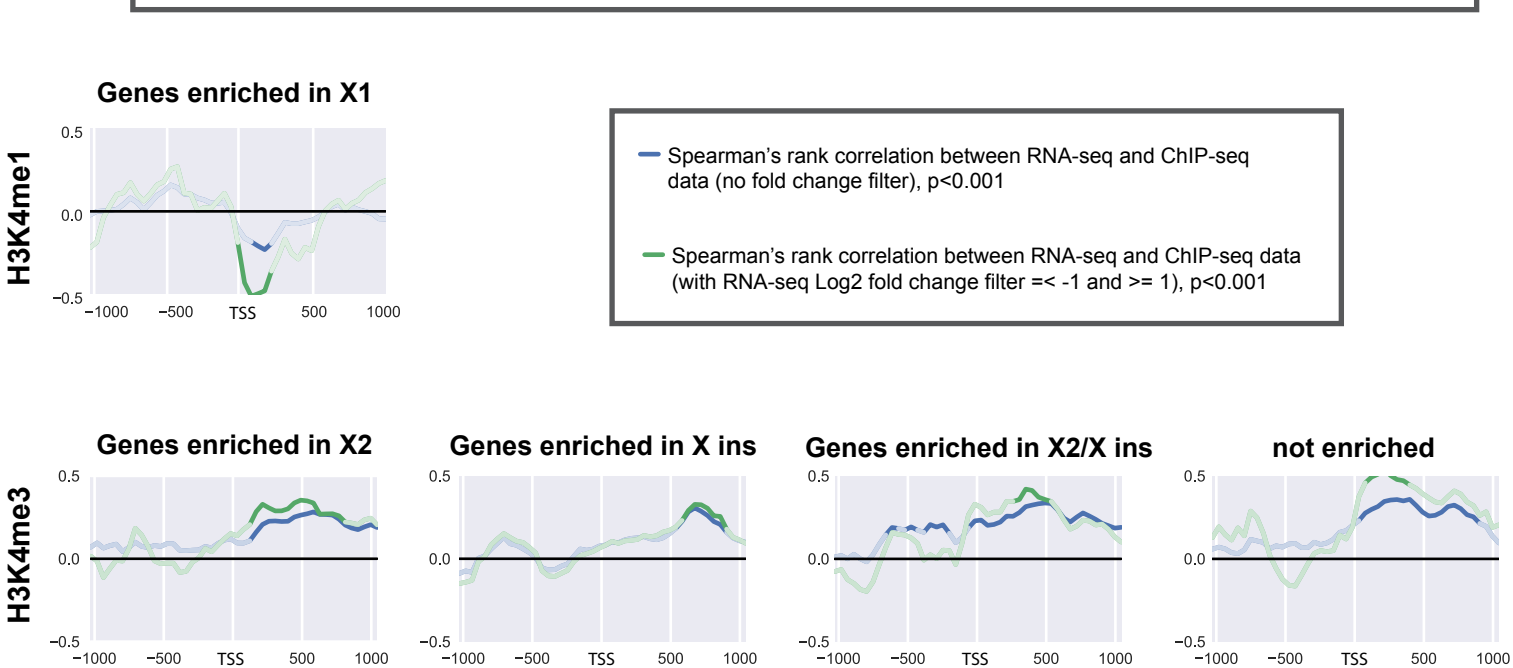

b
Genes enriched in X1
H3K4me1
Spearman's rank correlation between RNA-seq and ChIP-seq data (no fold change filter), p<0.001
Spearman's rank correlation between RNA-seq and ChIP-seq data (with RNA-seq Log2 fold change filter =< -1 and >= 1), p<0.001
Genes enriched in X2
Genes enriched in X ins
Genes enriched in X2/X ins
not enriched
H3K4me3
TSS

Figure 9

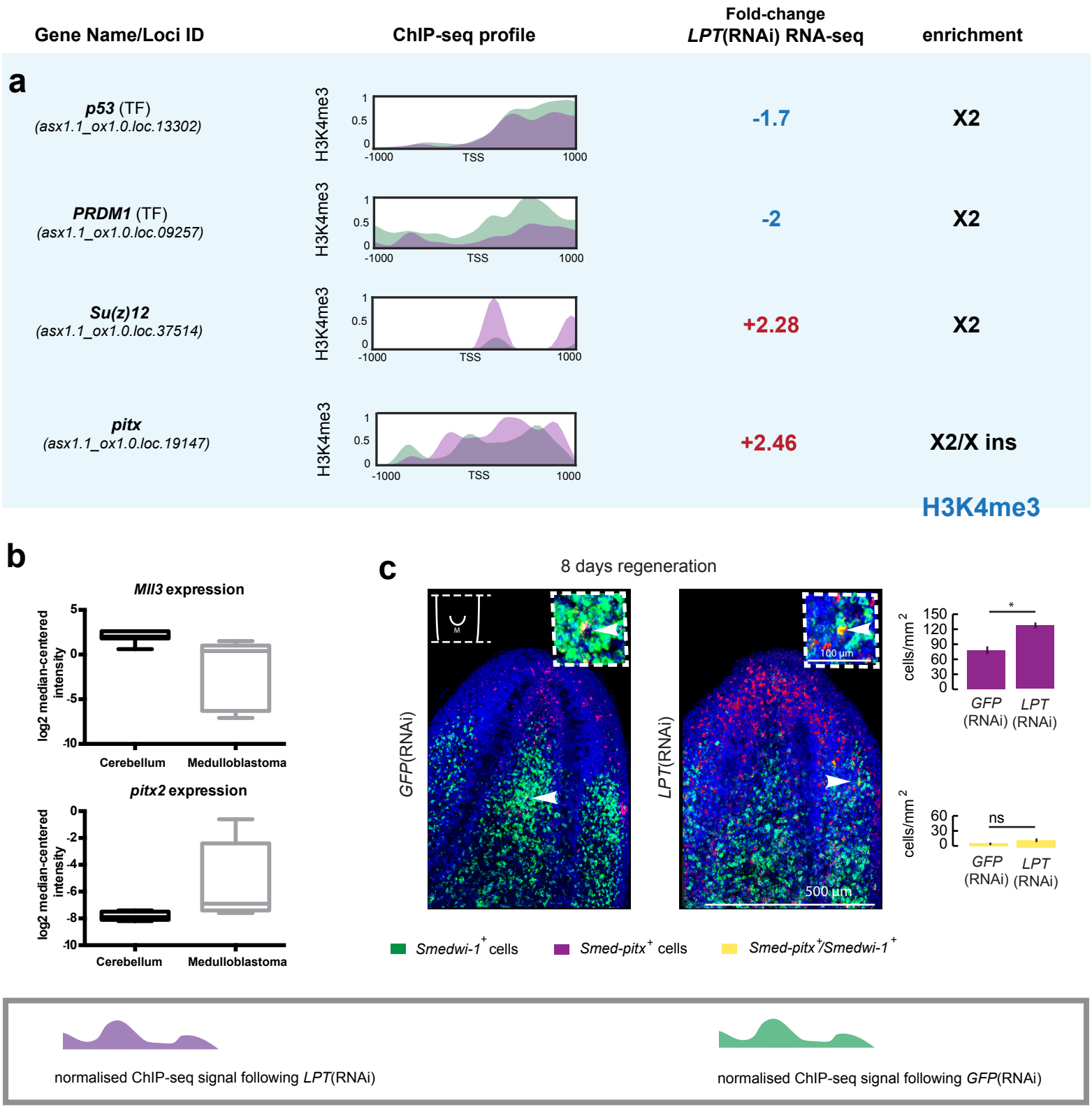
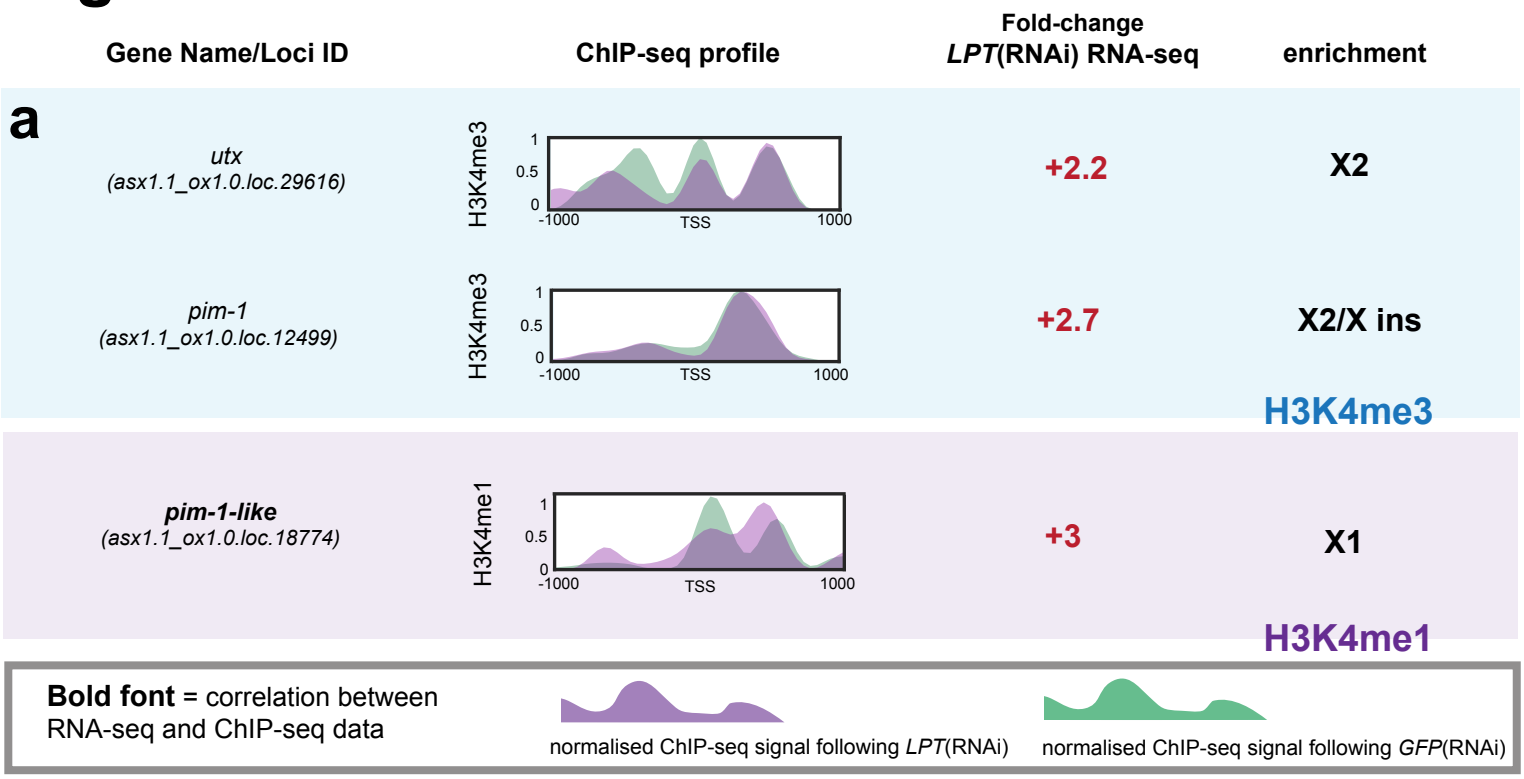
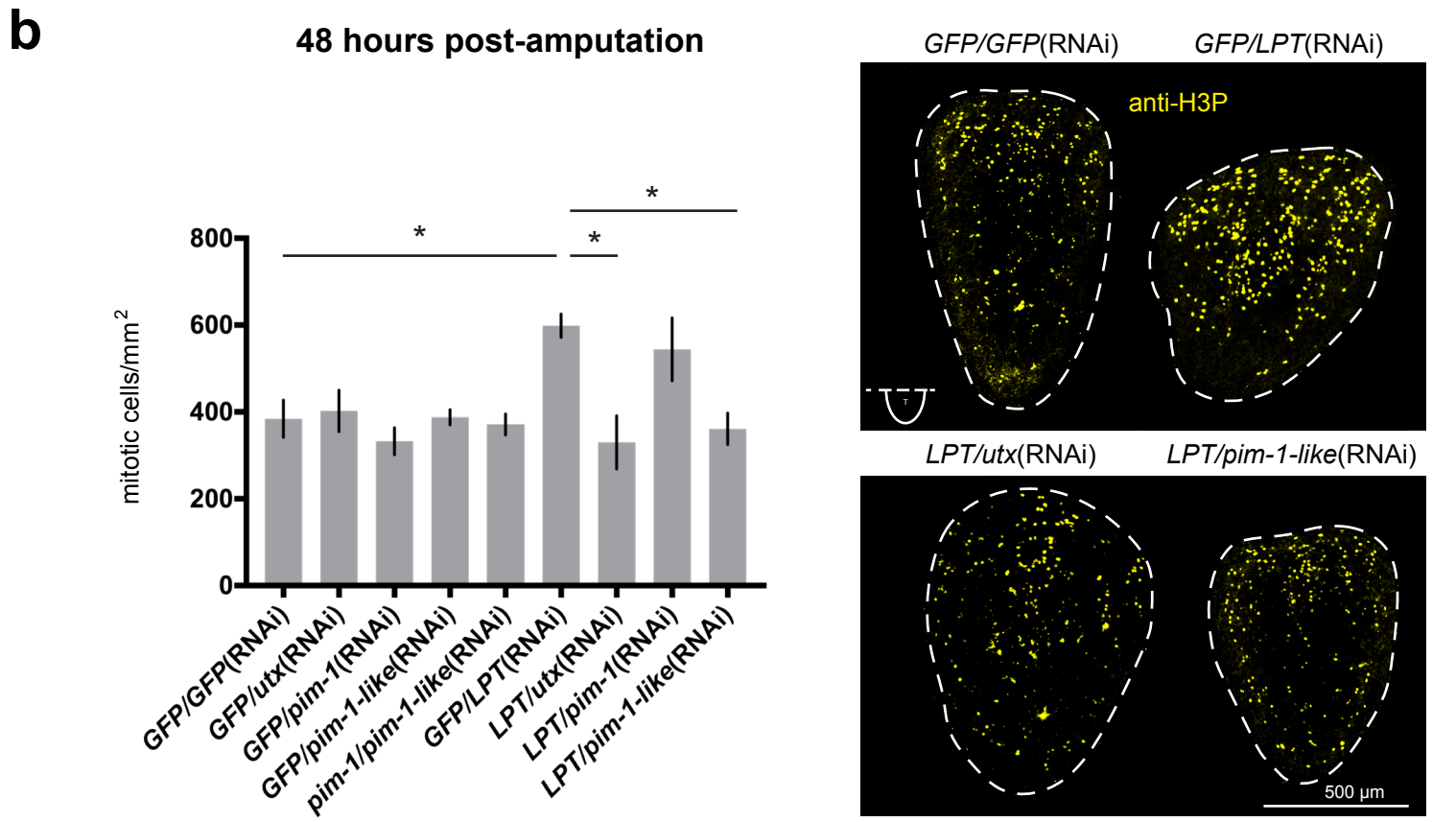
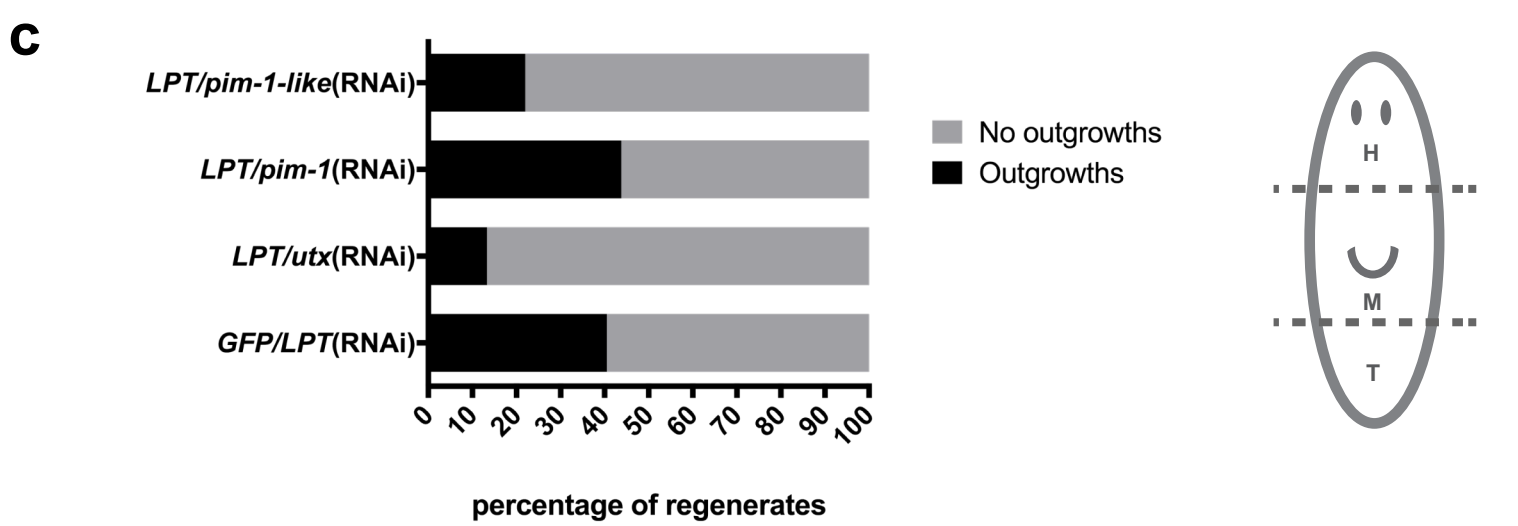

Gene Name/Loci ID
ChIP-seq profile
Fold-change
LPT(RNAi) RNA-seq
enrichment
a
p53 (TF)
(asx1.1_ox1.0.loc.13302)
-1.7
X2
PRDM1 (TF)
(asx1.1_ox1.0.loc.09257)
-2
X2
Su(z)12
(asx1.1_ox1.0.loc.37514)
+2.28
X2
pitx
(asx1.1_ox1.0.loc.19147)
+2.46
X2/X ins
H3K4me3
-1000
TSS
1000
b
Mll3 expression
pitx2 expression
log2 median-centered intensity
Cerebellum
Medulloblastoma
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)

Figure 10


Gene Name/Loci ID
ChIP-seq profile
Fold-change
LPT(RNAi) RNA-seq
enrichment
a
utx
(asx1.1_ox1.0.loc.29616)
H3K4me3
+2.2
X2
pim-1
(asx1.1_ox1.0.loc.12499)
H3K4me3
+2.7
X2/X ins
H3K4me3
pim-1-like
(asx1.1_ox1.0.loc.18774)
H3K4me1
+3
X1
H3K4me1
-1000
TSS
1000
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
mitotic cells/mm2
800
600
400
200
0
GFP/GFP(RNAi)
GFP/utx(RNAi)
GFP/pim-1(RNAi)
GFP/pim-1-like(RNAi)
pim-1/pim-1-like(RNAi)
GFP/LPT(RNAi)
LPT/utx(RNAi)
LPT/pim-1(RNAi)
LPT/pim-1-like(RNAi)
GFP/GFP(RNAi)
GFP/LPT(RNAi)
anti-H3P
LPT/utx(RNAi)
LPT/pim-1-like(RNAi)
500 µm


c
LPT/pim-1-like(RNAi)
LPT/pim-1(RNAi)
LPT/utx(RNAi)
GFP/LPT(RNAi)
0
10
20
30
40
50
60
70
80
90
100
percentage of regenerates
No outgrowths
Outgrowths
H
M
T