





ARTICLE

DOI: 10.1038/s41467-018-06092-6

OPEN

Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells

Yuliana Mihaylova ¹, Prasad Abnave ¹, Damian Kao¹, Samantha Hughes², Alvina Lai¹, Farah Jaber-Hijazi ³, Nobuyoshi Kosaka¹ & A. Aziz Aboobaker ¹

Currently, little is known about the evolution of epigenetic regulation in animal stem cells. Here we demonstrate, using the planarian stem cell system to investigate the role of the COMPASS family of MLL3/4 histone methyltransferases that their function as tumor suppressors in mammalian stem cells is conserved over a long evolutionary distance. To investigate the potential conservation of a genome-wide epigenetic regulatory program in animal stem cells, we assess the effects of MLL3/4 loss of function by performing RNA-seq and ChIP-seq on the G2/M planarian stem cell population, part of which contributes to the formation of outgrowths. We find many oncogenes and tumor suppressors among the affected genes that are likely candidates for mediating MLL3/4 tumor suppression function. Our work demonstrates conservation of an important epigenetic regulatory program in animals and highlights the utility of the planarian model system for studying epigenetic regulation.

¹Department of Zoology, Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK. ²HAN University of Applied Sciences, Institute of Applied Sciences, Laan van Scheut 2, 6525EM Nijmegen, The Netherlands. ³Beatson Institute for Cancer Research, Switchback Road, Bearsden, Glasgow G61 1BD, UK. These authors contributed equally: Yuliana Mihaylova, Prasad Abnave. Correspondence and requests for materials should be addressed to A.A.A. (email: Aziz.Aboobaker@zoo.ox.ac.uk)

The pluripotent adult stem cell population of planarian flatworms is a highly accessible study system to elucidate fundamental aspects of stem cell function^{1,2}. These stem cells, collectively known as neoblasts (NBs), bestow these animals with an endless capacity to regenerate all organs and tissues after amputation. Comparisons of stem cell expression profiles and functional data between animals show that some key aspects of stem cell biology are deeply conserved^{3–8}, while others, like the transcription factors that define pluripotency in mammalian stem cells, appear not to be. Thus, studies of NBs have the potential to inform us about the origins of fundamental stem cell properties that underpin metazoan evolution, such as maintenance of genome stability⁹, self-renewal^{7,10}, pluripotency^{11–13}, differentiation^{14–16}, and migration¹⁷. All of these are highly relevant to understanding human disease processes, particularly those leading to cancer.

Currently, very little comparative data exists for the role of epigenetic regulation in animal stem cells. Planarian NBs offer an opportunity to ask whether the cellular and physiological roles of different epigenetic regulators might be conserved between mammalian and other animal stem cells. Additionally, as mutations in many chromatin modifying enzymes are implicated in cancer^{18–20}, using NBs as a model system may provide fundamental insight into why these mutations lead to cancer, if epigenetic regulatory programs are conserved.

The genome-wide effects of chromatin modifying enzymes make understanding how they contribute to cancer phenotypes very challenging. Complexity in the form of tissue and cell heterogeneity, life history stage and stage of pathology make resolution of epigenetic regulatory cause and effect relationships in vivo very challenging. From this perspective, planarians and their easily accessible NB population may be a very useful model system. The planarian system could be particularly suitable for investigating the early transformative changes in stem cells at the onset of hyperplasia, as the NB identity of all potentially hyperplastic cells is known a priori.

The human MLL proteins are the core members of the highly conserved COMPASS-like (complex of proteins associated with Set1) H3K4 methylase complexes. An extensive research effort has now established the evolutionary history and histone modifying activities of this protein family (Supplementary Figure 1^{21–32}). Perturbation of MLL-mediated H3K4 methylase activity is characteristic of numerous cancer types. While prominent examples include the translocation events widely reported in leukemias involving the *Mll1* gene^{33,34}, the mutation rate of *Mll3* across malignancies of different origin approaches 7%, making *Mll3* one of the most commonly mutated genes in cancer¹⁹. In attempts to model the role of *Mll3* in cancer, mice homozygous for a targeted deletion of the *Mll3* SET domain were found to succumb to ureter epithelial tumors at high frequency²⁴, an effect enhanced in a *p53*^{+/-} mutational background. Heterozygous deletions of *Mll3* in mice also lead to acute myeloid leukemia, as hematopoietic stem cells fail to differentiate correctly and over-proliferate, implicating *Mll3* in dose-dependent tumor suppression²⁰. Recent studies have revealed an increasingly complicated molecular function of MLL3, its closely related paralog MLL4, and their partial *Drosophila* orthologs—LPT (Lost PHD-fingers of trithorax-related; corresponding to the N-terminus of MLL3/4) and Trr (trithorax-related; corresponding to the C-terminus of MLL3/4)²⁶. LPT-Trr/MLL3/4 proteins have a role in transcriptional control via mono-methylating and/or tri-methylating H3K4 at promoters and enhancers^{22,23,25,26,30,35} (Supplementary Figure 1).

Links between mutations in *Mll3/4*, effects on downstream targets of MLL3/4 and human cancers remain to be elucidated. If the role of MLL3/4 in regulating stem cells is conserved in NBs,

planarians could provide an informative in vivo system for identifying potential candidate target genes that drive tumor formation. Here we exploit the accessibility of NBs to identify and investigate the role of *Mll3/4* orthologs in the planarian *Schmidtea mediterranea*, and show knockdown leads to NB over proliferation, tissue outgrowths containing proliferative NBs and differentiation defects. Investigating the regulatory effects accompanying this phenotype, we demonstrate mis-regulation of both oncogenes and tumor suppressors, providing a potential explanation for how tumor suppressor function is mediated by MLL3/4.

Results

Planarian orthologs of *Mll3/4* are expressed in stem cells. We found 3 partial orthologs of mammalian *Mll3* and *Mll4* genes. We named the planarian gene homologous to *Drosophila* LPT and the N-terminus of mammalian *Mll3/4*—*Smed-LPT* (KX681482) (Supplementary Figure 2a). *Smed-LPT* (LPT) protein contains two PHD-fingers and a PHD-like zinc-binding domain, suggesting that it has chromatin-binding properties³⁶ (Fig. 1a). There are two planarian genes homologous to *Drosophila* Trr and the C-terminus of mammalian *Mll3/4*—*Smed-trr-1* (KC262345) and *Smed-trr-2* (DN309269, HO004937), both previously described²⁷. Both SMED-TRR-1 and SMED-TRR-2 contain a PHD-like zinc-binding domain, a FYRN (FY-rich N-terminal domain), FYRC (FY-rich C-terminal domain) and a catalytic SET domain. SMED-TRR-1 (TRR-1) contains only a single NR (Nuclear Receptor) box at a non-conserved position and SMED-TRR-2 (TRR-2) has no NR boxes (Fig. 1a). This could indicate some functional divergence exists between TRR-1 and TRR-2, where only TRR-1 is capable of interacting with nuclear receptors^{26,37}.

We performed whole-mount *in situ* hybridization (WISH) and found that *LPT*, *trr-1* and *trr-2* are broadly expressed across many tissues and organs. Gamma irradiation to remove cycling cells in *S. mediterranea* revealed that the three transcripts are also likely to be expressed in NBs (Fig. 1b). A double fluorescent *in situ* hybridization (FISH) with the pan-stem cell marker *Histone 2B* (*H2B*) confirmed that over 90% of all NBs co-express *LPT*, *trr-1* and *trr-2* (Fig. 1c, d). The genes also showed clear expression in the brain, pharynx and other differentiated tissues (Fig. 1b).

***Mll3/4* RNAi causes regeneration defects and outgrowths.** In order to study the function of planarian *Mll3/4*, we investigated phenotypes after RNAi-mediated knockdown. Following *LPT* (RNAi), there was a clear failure to regenerate missing structures, including the eyes and pharynx, with regenerative blastemas smaller than controls (Fig. 2a, b). After 7 days of regeneration we observed that, as well as failure to regenerate missing structures, animals began to form tissue outgrowths (Fig. 2c). Intact (homeostatic) *LPT*(RNAi) animals also developed outgrowths, but at a lower frequency than regenerates (Fig. 2d).

Following individual knockdown of *trr-1* and *trr-2*, milder differentiation defects were observed compared to *LPT*(RNAi), with no obvious outgrowths (Supplementary Figure 2b–d), confirming results from an earlier study²⁷. However, *trr-1/trr-2* double knockdown recapitulated the phenotype of *LPT*(RNAi), but with higher penetrance and increased severity (Supplementary Figure 2e, f). Functional redundancy between the two *trr* paralogs likely accounts for the reduced severity after individual knockdown. Double knockdown animals all developed outgrowths and started dying at day 5 post-amputation. Based on these observations, we decided to focus our attention on the *LPT* (RNAi) phenotype as regeneration defects and the formation of tissue outgrowths were temporally distinct and could be studied consecutively.

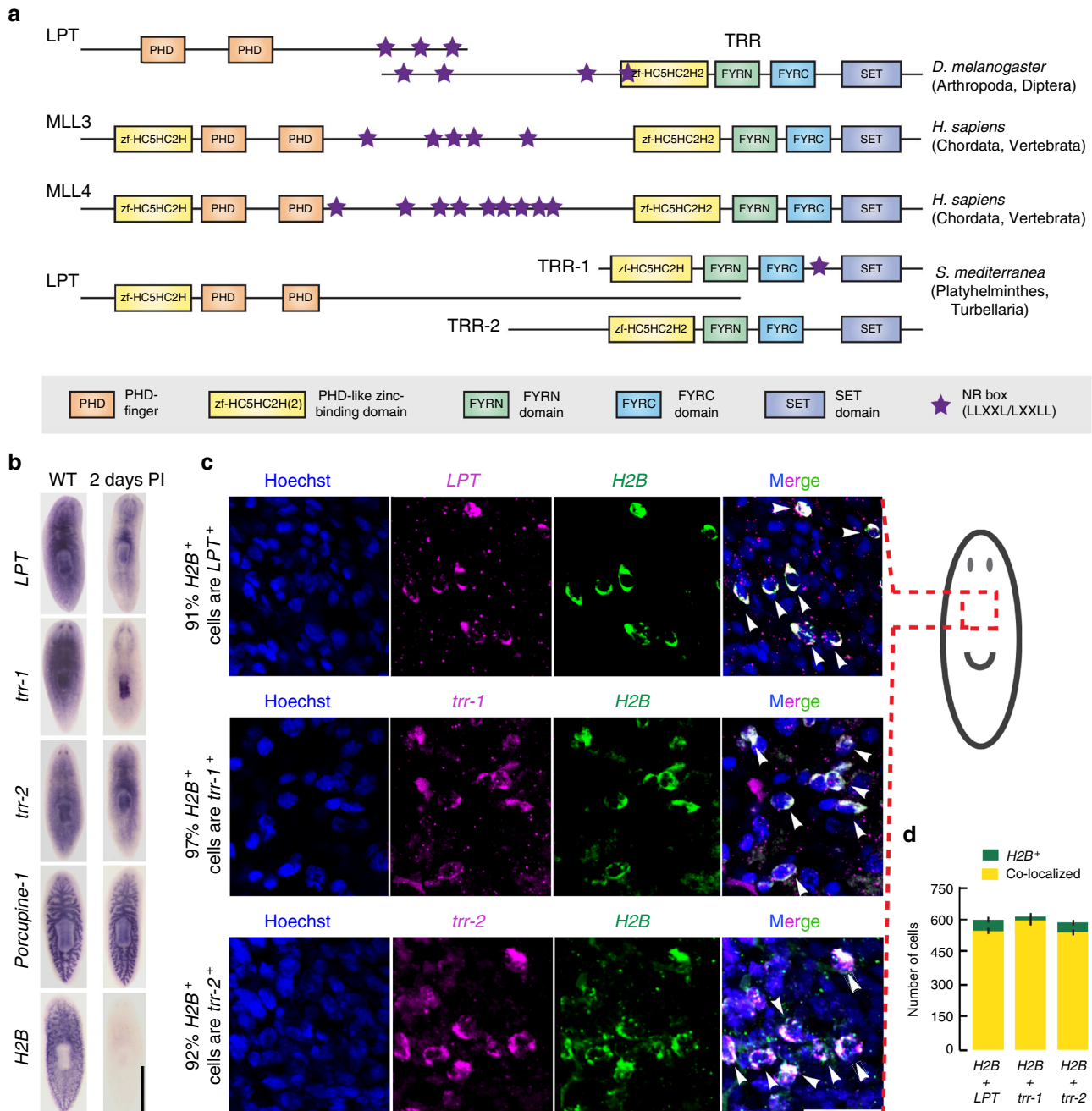


Fig. 1 *S. mediterranea* has three partial *Mll3/4* orthologs expressed in stem cells. **a** A schematic depicting the structure and domain composition of MLL3/MLL4 proteins in *D. melanogaster*, *H. sapiens* and *S. mediterranea*. **b** genes' expression pattern in wildtype (WT) and two days following a lethal dose (60 Gy) of gamma irradiation (PI=post-irradiation). *Porcupine-1* (expressed in the irradiation-insensitive cells of the differentiated gut) and *H2B* (expressed in the irradiation-sensitive neoblasts) are used as a negative and positive control respectively. Ten worms per condition were used. Scale bar: 500 μ m. **c** White arrows point to examples of cells double-positive for *Mll3/4* transcripts and *H2B* transcripts. The schematic shows the body area imaged. Scale bar: 50 μ m. **d** Graph showing the raw cell counts used for percentage estimates in (c). Green color represents all counted *H2B*-positive cells, yellow represents *H2B*-positive cells also expressing an *Mll3/4* ortholog. Error bars represent Standard Error of the Mean (SEM). Ten animals per condition were used

A more thorough study of the differentiation properties of *LPT* (RNAi) animals following amputation showed that the triclad gut structure failed to regenerate secondary and tertiary branches and to extend major anterior and posterior branches (Fig. 2e). Cephalic ganglia (CG) regenerated as smaller structures, the two CG lobes did not join in their anterior ends in *LPT*(RNAi) animals (Fig. 2f) and the optic chiasm and optic cups were mis-patterned and markedly reduced (Fig. 2g, h). We found that 80% of *LPT*(RNAi) animals did not regenerate any new pharyngeal tissue (Fig. 2i). We interpreted these regenerative defects as being

indicative of either a broad failure in stem cell maintenance and/or differentiation.

MLL3/4 controls differentiation of the epidermis and neurons. One of the structures most severely affected following loss of *LPT* function was the brain so we looked at the regeneration of different neuronal subtypes. *LPT*(RNAi) animals had reduced numbers of GABAergic (Fig. 3a), dopaminergic (Fig. 3b), acetylcholinergic (Fig. 3c) and serotonergic (Fig. 3d) neurons.

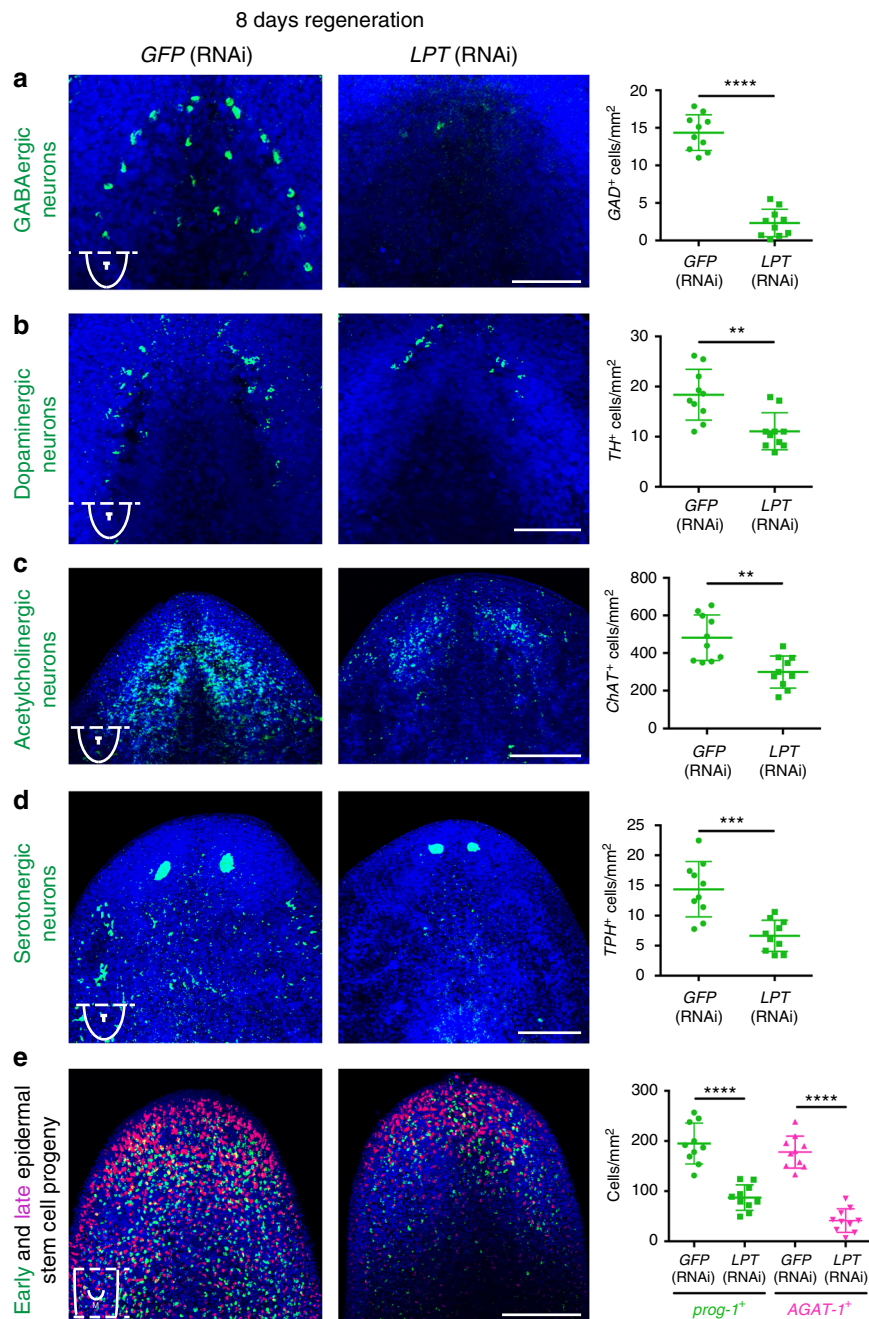


Fig. 3 LPT controls differentiation across neuronal and epidermal lineages. **a** Quantification of the number of GABAergic neurons (labeled by GAD), **b** dopaminergic neurons (labeled by TH), **c** acetylcholinergic neurons (labeled by ChAT), **d** serotonergic neurons (labeled by TPH) and **e** early (labeled by *prog-1*) and late (labeled by *AGAT-1*) epidermal stem cell progeny at 8 days of regeneration following *LPT(RNAi)*. Two-tailed t-test was used for all comparisons; ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.0001$. Lines and error bars indicate mean and SD. Five animals per condition per experiment were assessed over the course of two separate experiments. Scale bars: 200 μm

Brain defects were milder following knockdown of either *trr-1* or *trr-2* in agreement with the hypothesis of some functional redundancy between these paralogs (Supplementary Figure 3a–d).

Epidermal tissue was also affected. Both early (*Prog-1*⁺ cells) and late (*AGAT-1*⁺ cells) epidermal progeny were significantly decreased, but not entirely absent, in *LPT(RNAi)* regenerating animals (Fig. 3e). No such defect was seen in individual *trr-1* and *trr-2* knockdown animals (Supplementary Figure 3e).

These effects along with defects in pharynx and gut tissues implicate broad NB differentiation defects in *LPT(RNAi)* animals.

LPT limits normal stem cell proliferation and tissue growth.

Aside from impairment of regeneration following *LPT(RNAi)*, the other major phenotype we observed were outgrowths of tissue that appeared at unpredictable positions in regenerating pieces.

Planarian regeneration is characterized by an early burst of increased NB proliferation, 6–12 h after wounding, and a second peak of proliferation, 48 h after amputation³⁸. Following *LPT(RNAi)*, we observed significant increases in proliferation at both of these peaks and at 8 days post-amputation, as proliferation failed to return to normal homeostatic levels (Fig. 4a). It was previously reported that *Trr-1(RNAi)* leads to decreases in

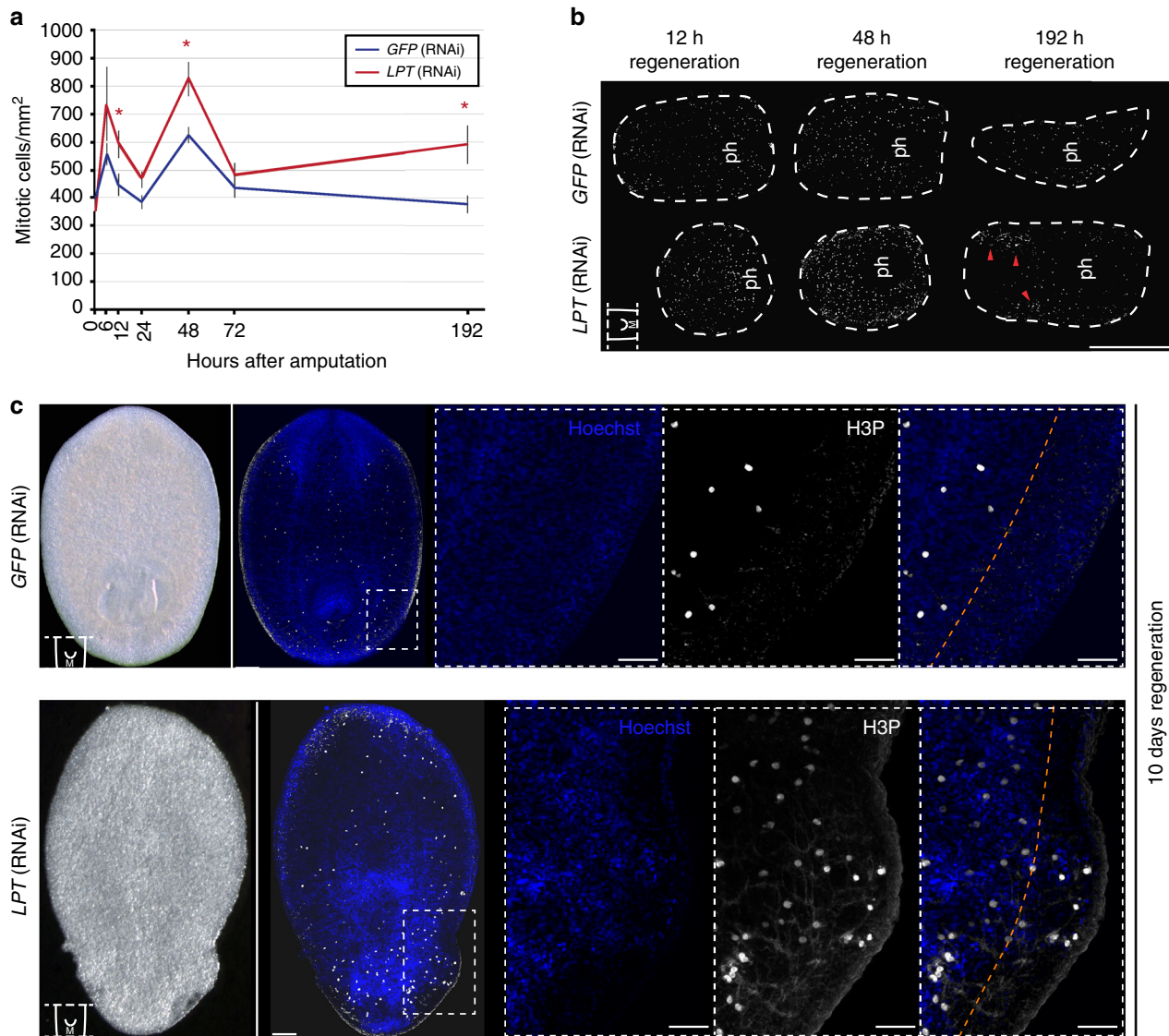


Fig. 4 Over-proliferation and mitotic cell clustering precedes and accompanies the emergence of outgrowths in *LPT(RNAi)* animals. **a** Quantification of mitotic cells (labeled by anti-H3P antibody) at different post-amputation time-points following *LPT(RNAi)*. $N=10$ animals per time-point. Two-tailed *t*-test was used for analysis; $*p < 0.05$. Error bars represent Standard Error of the Mean (SEM). **b** Examples of middle pieces at the time-points post-amputation showing significant difference in mitotic cell (white) counts according to **(a)**. 'ph' indicates the pharynx. The red arrows point towards clusters of mitotic cells in late stage regenerates (192 h/8 days). Scale bar: 500 μm . **c** Mitotic cells (white) are always restricted to mesenchyme tissue in *GFP(RNAi)* animals but penetrate into epidermal outgrowths in *LPT(RNAi)* animals. Orange line indicates the border of mesenchymal tissue. Scale bar: 50 μm

mitotic cells, whereas *Trr-2(RNAi)* doesn't affect NB proliferation²⁷.

In 8 day-regenerating *LPT(RNAi)* worms the observed over-proliferation is a result of localized clusters of mitotic cells rather than broad increase in proliferation across regenerating animals (Fig. 4b). This is different from previously reported planarian outgrowth phenotypes from our group and others, where hyperplastic stem cells are evenly distributed^{39,40}. It seems likely that these mitotic clusters might eventually be responsible for the formation of outgrowths. When we looked at outgrowths, we found mitotic cells usually restricted to mesenchymal tissue, had penetrated into epidermal outgrowths in *LPT(RNAi)* animals (Fig. 4c).

In order to understand if ectopically cycling NBs represented the breadth of known stem cell heterogeneity in planarians or only a subset of lineages, we performed FISH for markers of the *sigma* (collectively pluripotent NBs), *zeta* (NBs committed to the epidermal lineage) and *gamma* (NBs committed to the gut

lineage) cell populations⁴¹. We found that all three NB populations are represented in the outgrowths of *LPT(RNAi)* animals (Fig. 5a–c). *Sigma*, *zeta*, and *gamma* NBs are not significantly increased in pre-outgrowth *LPT(RNAi)* animals (Supplementary Figure 4), suggesting that the presence of these cells in outgrowths is not a secondary effect of increased cell number and passive spread of these cell populations, but rather local proliferation at outgrowth sites.

The epidermal progeny, marked by *Prog-1* and *AGAT-1*, were concentrated in the outgrowths of *LPT(RNAi)* animals, while being in reduced numbers in non-outgrowth tissue (Supplementary Figure 5a). The observed disarray of *Prog-1*⁺ and *AGAT-1*⁺ cells in outgrowths could be the result of perturbed patterning and polarity of the epidermal layer in *LPT(RNAi)* animals (Supplementary Figure 5b), as epidermal cells appear to have lost polarity and to be no longer capable of forming a smooth epidermal layer. Furthermore, the average epidermal nuclear size is significantly increased in *LPT(RNAi)* animals compared to

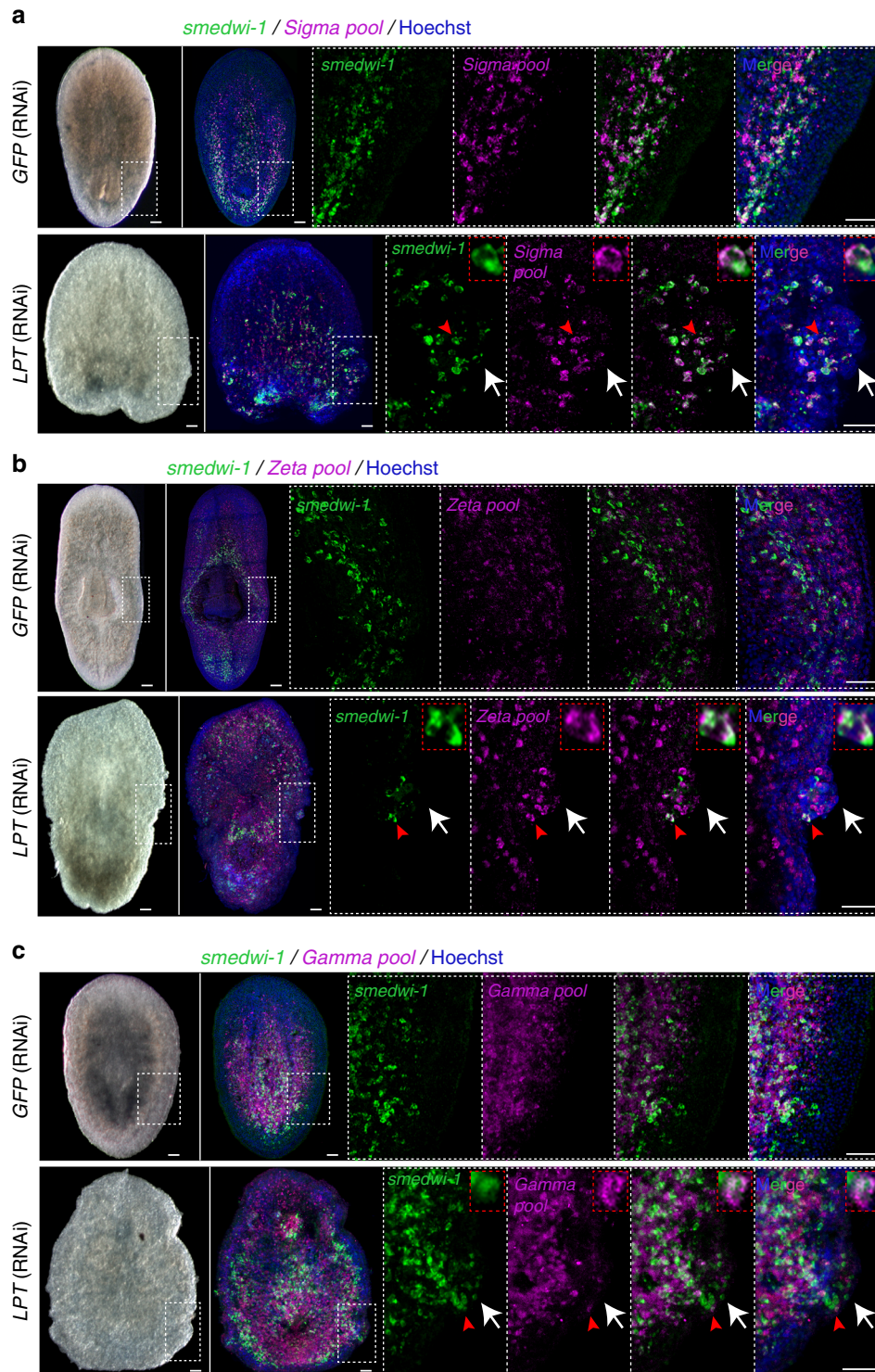


Fig. 5 Pluripotent as well as lineage restricted stem cells are present in outgrowths of *LPT*(RNAi) animals. **a** Head piece showing distribution of *Sigma* stem cells in *GFP*(RNAi) and *LPT*(RNAi) animals. *Sigma* stem cells are double positive for *smedwi-1* and the '*Sigma pool*' of RNA probes (*Soxp1*, *Soxp2*). White arrows in *LPT*(RNAi) animals point towards the outgrowth. Red arrows indicate a double-positive cell magnified in red inset box. **b** Images showing distribution of *Zeta* stem cells in *GFP*(RNAi) and *LPT*(RNAi) animals. *Zeta* stem cells are double positive for *smedwi-1* and the '*Zeta pool*' of RNA probes (*zfp-1*, *Soxp3*, *egr-1*). White arrows in *LPT*(RNAi) animals point towards the outgrowth. Red arrows indicate a double-positive cell magnified in red inset box. **c** Images showing distribution of *Gamma* stem cells in *GFP*(RNAi) and *LPT*(RNAi) animals. *Gamma* stem cells are double positive for *smedwi-1* and the '*Gamma pool*' of RNA probes (*gata4/5/6*, *hnf4*). White arrows in *LPT*(RNAi) animals point towards the outgrowth. Red arrows indicate a double-positive cell magnified in red inset box. Scale bars: 50 μ m

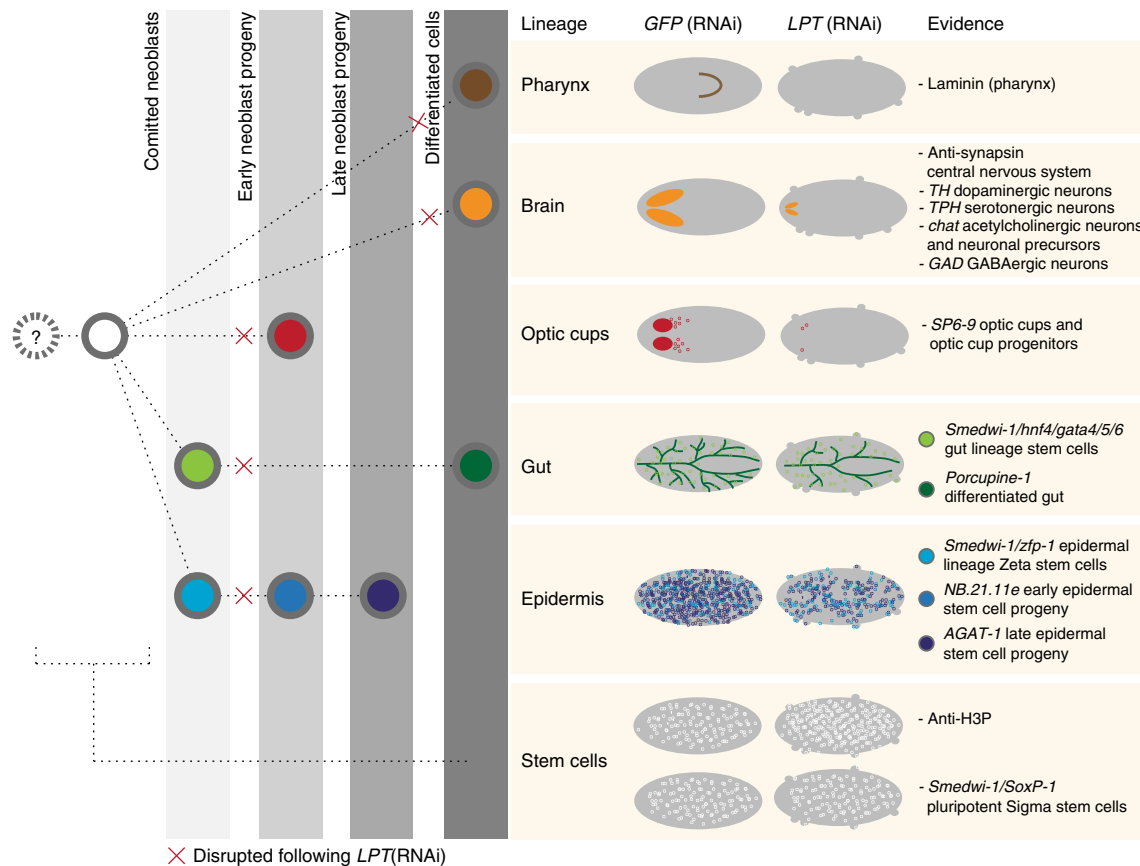


Fig. 6 LPT(RNAi) results in a cancer-like phenotype. A summary of the differentiation and neoblast proliferation data presented, together with a simplified flowchart illustrating the tested lineages' development under knockdown conditions. A red cross sign indicates where the defect in a lineage is detected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

controls (Supplementary Figure 5c), an effect similar to the pathology seen following knockdown of the tumor suppressor *SMG-1*⁴⁰. The epithelial layer in LPT(RNAi) animals also appears less well-defined than that in control animals, with a blurred distinction between epithelium and mesenchyme. Another feature of the LPT(RNAi) phenotype, encountered in a variety of malignancies⁴², are changes in nuclear shape (Supplementary Figure 5d).

In summary, LPT controls NB proliferation and restricts stem cells to pre-defined tissue compartments as well as being responsible for the successful differentiation of several lineages. Taken together, our data demonstrate that disturbance of the function of planarian LPT leads to development of both differentiation and proliferation defects (Fig. 6), allowing us to conclude that the function of LPT/trr/Mll3/4 proteins as an epigenetic tumor suppressor function is conserved over a large evolutionary distance.

Transcriptional changes driving proliferation in stem cells. A key insight missing from the literature for *Mll3* and *Mll4* mutations, is the downstream targets that are mis-regulated in disease states, for example in hematopoietic stem cells that cause leukemia²⁶. Given the conserved tumor suppressor function in planarians, we decided to focus on early regeneration when LPT (RNAi) animals that do not yet exhibit any outgrowth phenotype, providing the possibility to describe early regulatory changes that are potentially causal of outgrowths, rather than consequential. We performed RNA-seq on X1 (G2/M) fluorescence activated cell sorted (FACS) NBs from LPT(RNAi) and GFP(RNAi) planarians

at 3 days of regeneration. Our analysis revealed that 540 transcripts are down-regulated (fold change ≤ -1.5 , $p < 0.05$) and 542 up-regulated (fold change ≥ 1.5 , $p < 0.05$) in X1 stem cells from LPT(RNAi) animals when compared to controls (Supplementary Data 1).

A recent meta-analysis of all available *S. mediterranea* RNA-seq data allowed classification of all expressed loci in the planarian genome by their relative expression in FACS sorted cell populations representing stem cells, stem cell progeny and differentiated cells⁸. Superimposing the differentially expressed genes following LPT(RNAi) onto a gene expression spectrum reflecting FACS compartments, shows that LPT(RNAi) has a broad effect on gene expression in X1 cells (Fig. 7a), affecting genes normally expressed in different planarian cell types (Fig. 7b).

Analysis of Gene Ontology (GO) terms revealed a clear enrichment for cell cycle and cell division-associated terms in the list of up-regulated genes (Fig. 7c), in agreement with the observed hyper-proliferation in LPT(RNAi) phenotype. The list of down-regulated genes is also enriched for cell cycle-related terms, as well as cell differentiation and metabolism-related processes (Fig. 7c). These findings suggest a broad link between gene expression changes caused by LPT loss of function and NB over-proliferation.

Promoter H3K4 methylation and transcription are correlated. Previous studies tie MLL3/4/LPT-Trr function directly to mono- and tri-methylation of H3K4^{22–25} and indirectly to tri-methylation of H3K27, because the H3K27me3 demethylase

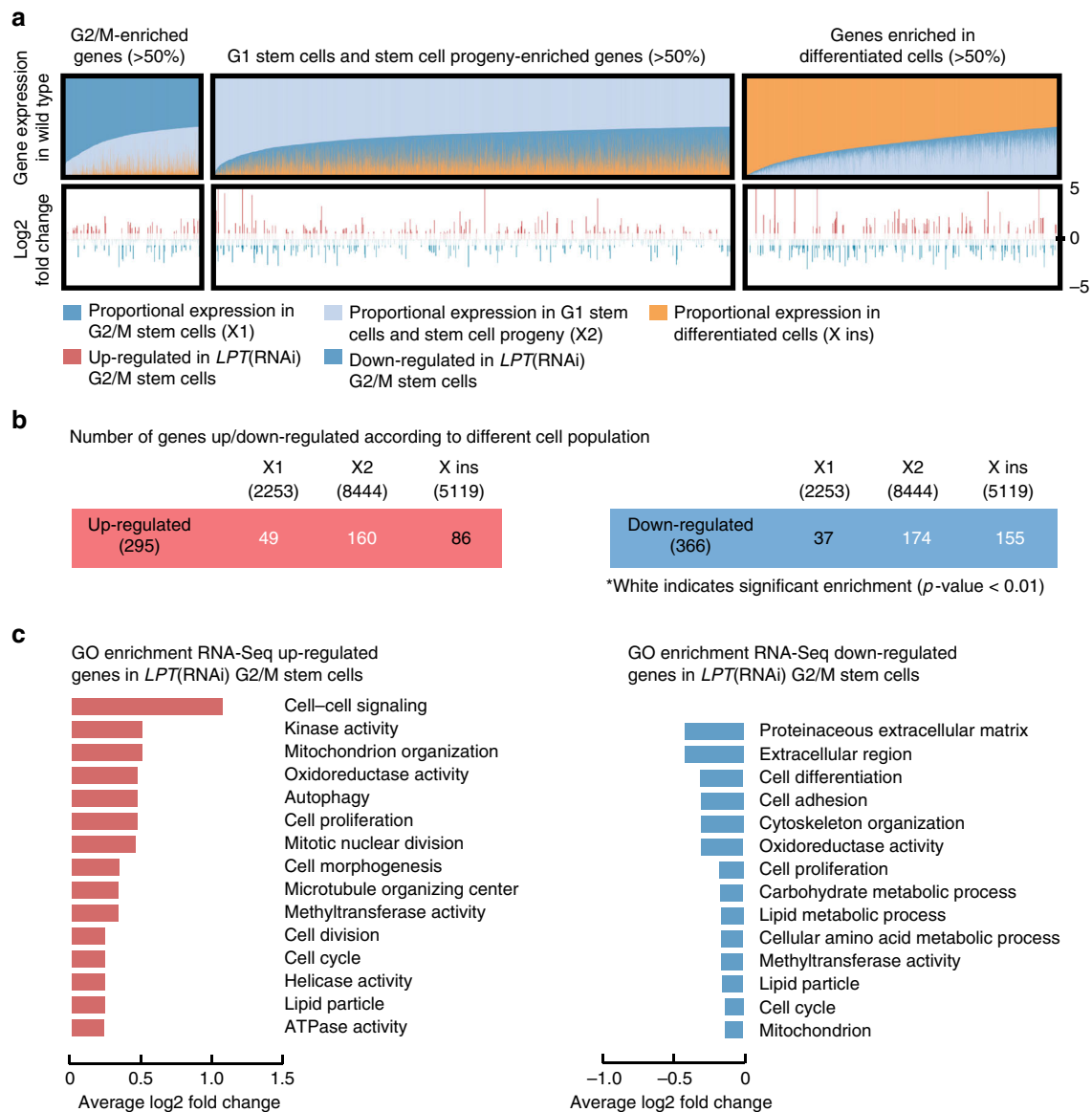


Fig. 7 RNA-seq of G2/M stem cells following *LPT*(RNAi) reveals effects on genes enriched in different cell populations. **a** Genes were classified according to their proportional expression in the X1 (G2/M stem cells; dark blue), X2 (G1 stem cells and stem cell progeny; light blue) and X ins (differentiated cells; orange) FACS populations of cells. Genes were defined as enriched in certain population(s) if >50% of their expression is observed in that population in wild type animals. Each vertical line represents a gene. Under the population expression enrichment track is a track with all the significantly up- and down-regulated genes in G2/M stem cells following *LPT*(RNAi). The genes with fold change >1.5 ($p < 0.05$) are shown in red following a log₂ fold change transformation. The genes with fold change <-1.5 ($p < 0.05$) are shown in blue following a log₂ fold change transformation. The Wald's test (as part of the Sleuth software) was used for assessing differential expression. **b** Enrichment for genes in each of the three classes was calculated for the up- and down-regulated genes' list (red and blue respectively). The number of genes in each group is indicated in brackets under the group's name. Numbers in white represent significant enrichment ($p < 0.01$) according to a hypergeometric enrichment test. **c** Gene Ontology (GO) enrichment analysis on the genes significantly up-regulated (red) and down-regulated (blue) in G2/M stem cells following *LPT*(RNAi). Categories are sorted by average log₂ fold change of the up- or down-regulated genes falling in each category. In bold are shown terms that relate to the described *Mil3/4* loss of function phenotype

UTX is present in the same protein complex⁴³. We set out to understand potential epigenetic causes of the transcriptional changes following *LPT*(RNAi) in planarians. To this end, we performed Chromatin Immunoprecipitation sequencing (ChIP-seq) on isolated planarian G2/M stem cells and used *Drosophila* S2 cells as a spike-in control to normalize for any technical differences between samples⁴⁴. The profile of H3K4me₃, H3K4me₁ and H3K27me₃ at the transcriptional start sites (TSSs) of genes in control X1 cells was in agreement with their conserved roles in transcriptional control^{8,25} (Fig. 8a, Supplementary Figure 6), suggesting our methodology was robust. *LPT*(RNAi) led to only

subtle changes in the overall level of H3K4me₃ and H3K4me₁ at TSSs throughout the genome. A decrease in H3K4me₁ and H3K4me₃ was apparent proximal to the TSS in genes where expression is normally enriched in differentiated cells (Fig. 8a). Concomitant with this, we also observed an increase in H3K4me₁ signal downstream of the predicted TSS for genes in enriched in stem cells (Fig. 8a). For the H3K27me₃ mark, no consistent pattern was observed as a result of *LPT*(RNAi) in any group of genes subdivided by expression profiles (Fig. 8a).

We next looked specifically at the promoter histone methylation status of those genes whose transcript levels were affected by

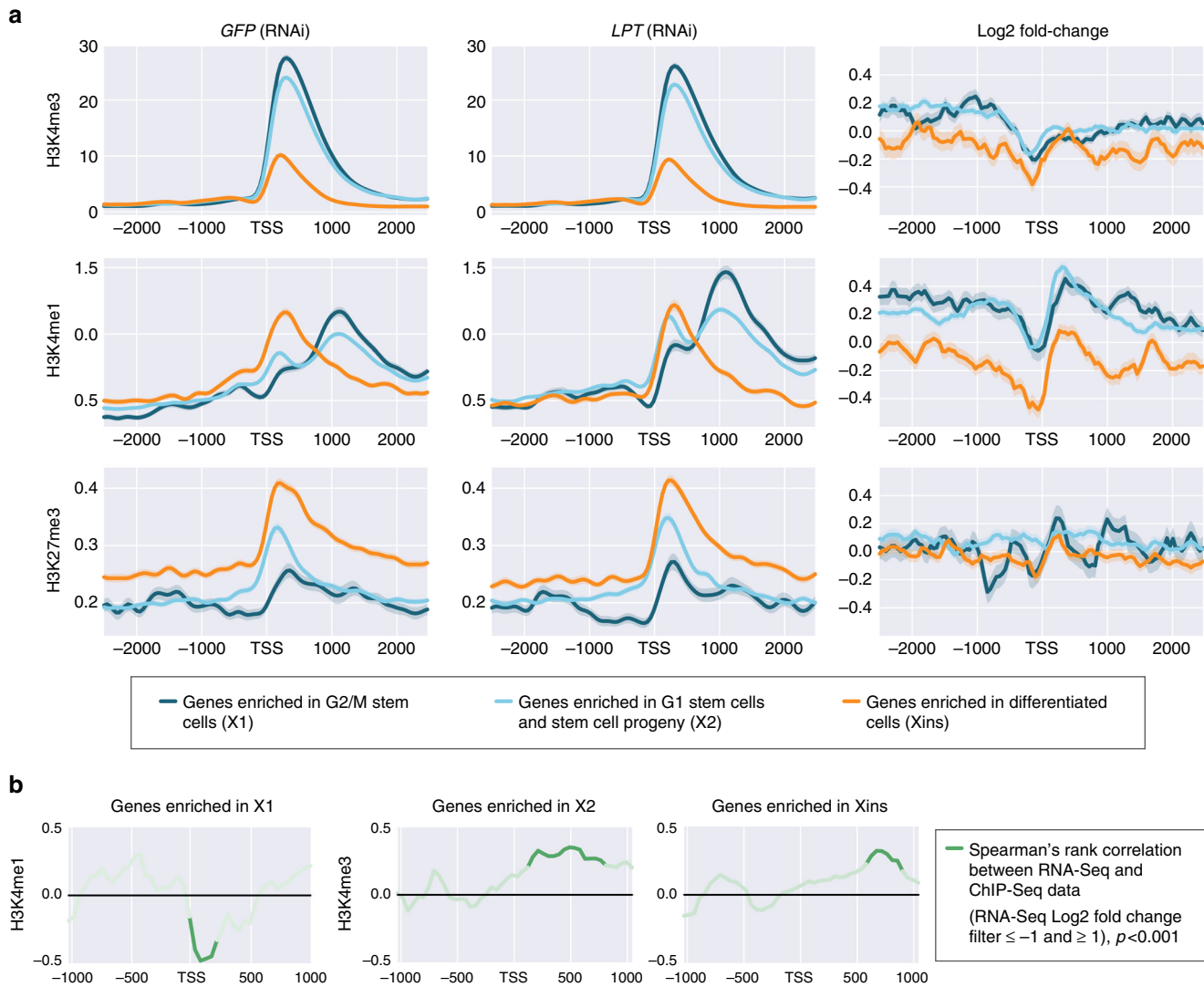


Fig. 8 *LPT*(RNAi) is mainly manifested in changes in H3K4me1 and H3K4me3 around the TSS in G2/M (X1) stem cells. **a** Graphs presenting the average read coverage across the genome for H3K4me3, H3K4me1 and H3K27me3 after performing ChIP-seq on X1 (G2/M stem cells). The graphs are centered on the TSS (showing 2 kb upstream and downstream) and the data is normalized to *Drosophila* S2 signal spike-in. The input coverage is subtracted. Log2 fold change graphs are also shown for each histone modification, where signal above zero shows increase following *LPT*(RNAi) and signal below zero represents a decrease. Three colors are used for different gene classes—dark blue (genes enriched in G2/M stem cells, i.e., X1), light blue (genes enriched in G1 stem cells and stem cell progeny, i.e., X2), orange (genes enriched in differentiated cells, i.e., Xins). Standard deviation is shown by a faded color around each line. **b** Spearman's rank correlation between changes in RNA-seq signal following *LPT*(RNAi) and H3K4me1 or H3K4me3 ChIP-seq signal for the region around the TSS of genes from different enrichment classes (only examples where a significant correlation exists are shown). The green line shows a correlation where RNA-seq fold change data was filtered for Log2 fold changes ≤ -1 and $\geq +1$. Faded areas of the lines represent results not significant at $p < 0.001$, while darker colors represent results significant at $p < 0.001$

LPT(RNAi). For genes with enriched expression in NBs, we observed a significant inverse correlation between expression following *LPT*(RNAi) and the amount of TSS-proximal H3K4me1 (Fig. 8b). This indicates that *LPT*(RNAi) leads to a reduction of this repressive mark at some loci and an up-regulation of the cognate transcript expression in stem cells, consistent with the role of H3K4me1 as a repressive mark. For mis-regulated genes not normally enriched in X1 NBs, we observed instead a positive correlation between changes in transcriptional expression (upregulation) and changes in H3K4me3 levels at the TSS and gene bodies (Fig. 8b).

Overall, our data suggest that reductions in H3K4me1 following *LPT*(RNAi) cause up-regulation of some of the stem cell genes implicated by RNA-seq data from *LPT*(RNAi) animals.

Our data are consistent with MLL3/4's known role in H3K4 methylation and identify gene expression profiles following *LPT* (RNAi) that are broadly correlated with the amount of H3K4me1 and H3K4me3 at gene promoters.

Mis-regulation of oncogenes and tumor suppressors. After observing the global changes in expression and histone modification patterns following *LPT*(RNAi), we wanted to investigate individually mis-regulated genes that could be major contributors to the differentiation and outgrowth phenotypes (Figs. 9, 10). Within our list of up- and down-regulated genes, we saw mis-regulation of tumor suppressors, oncogenes and developmental genes (Supplementary Data 1). Detailed inspection of changes in promoter patterns of H3K4 methylation revealed example loci

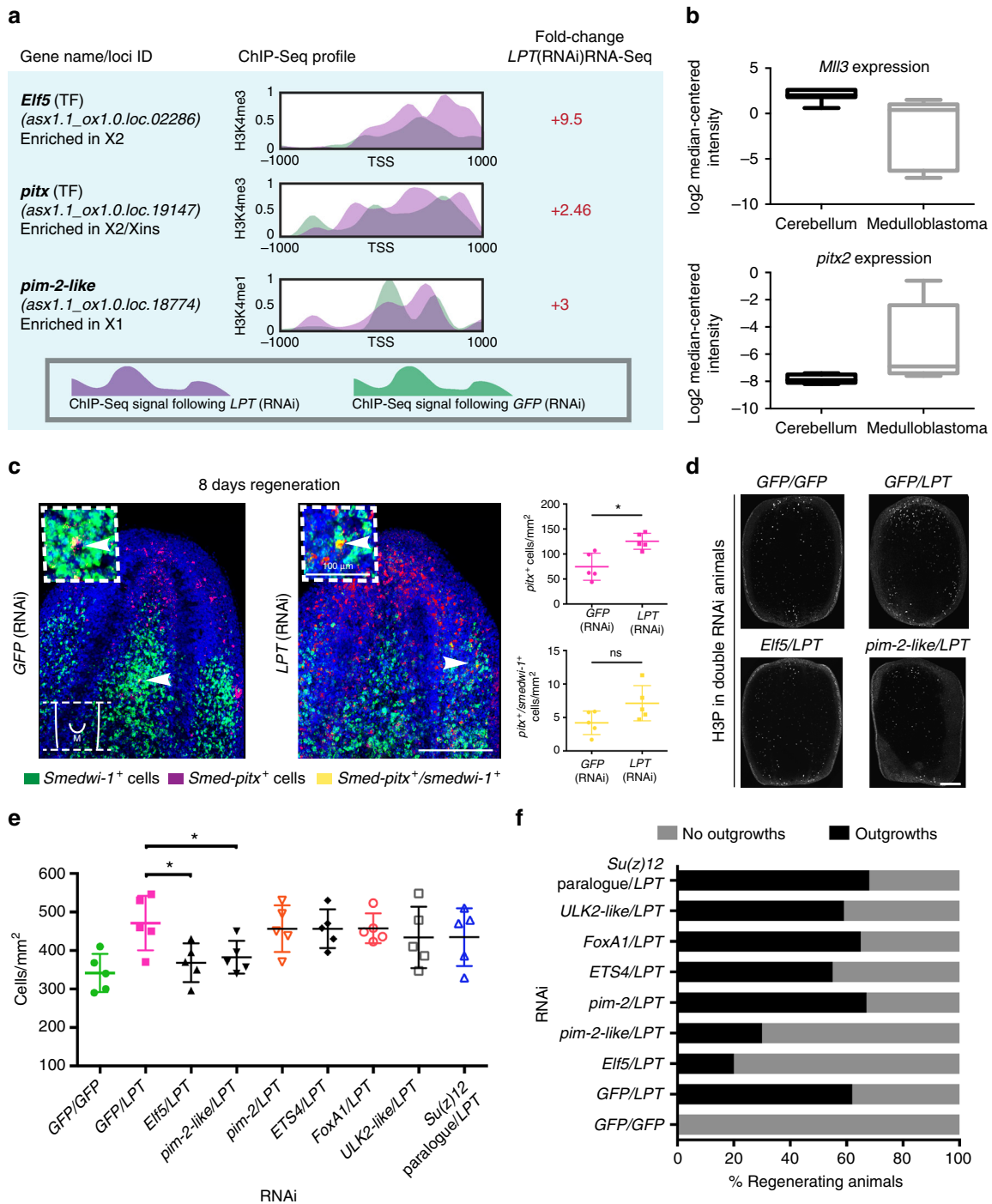


Fig. 9 Double knockdown with *Elf5* or *pim-2-like* alleviates the *LPT*(RNAi) over-proliferation and outgrowth phenotype. **a** Examples of genes significantly ($p < 0.05$) up-regulated in G2/M stem cells following *LPT*(RNAi). The ChIP-seq profile for H3K4me3 or H3K4me1 in the 2 kb region around the TSS of each gene is presented. Purple color represents normalized signal following *LPT*(RNAi) and green color is used to show the normalized signal following *GFP*(RNAi). The bold font of the gene names signifies a correlation between the genes' up-regulation and their respective H3K4me3/me1 profile. 'TF' stands for 'transcription factor'. All three genes show correlation between ChIP-seq profile and up-regulation in RNA-seq data. **b** *in silico* analysis (www.oncoPrint.org; *t*-test, $p < 0.0001$) of *Mli3* and *pitx2* expression in normal tissue (cerebellum) and cancer tissue (medulloblastoma). Lines and error bars indicate mean and SD. **c** *pitx* and *smedwi-1* *in situ* hybridization at 8 days of regeneration of middle pieces following *LPT*(RNAi). White arrows show double-positive cells. Lines and error bars indicate mean and SD. Two-tailed *t*-test used for analysis ($n = 5$), * $p < 0.05$, ns is not significant. **d** Representative examples of mitotic cells (labeled by anti-H3P antibody) in double RNAi condition at 48 h post amputation. **e** Graph showing number of mitotic cells in double RNAi animals at 48 h post amputation. Each dot represents average number of mitotic cells in single worm ($n = 5$). Lines and error bars indicate mean and SD. Two-tailed *t*-test: * $p < 0.05$. **f** Graph showing percentage quantification of double knockdown regenerates developing outgrowths. Scale bars: 200 μm

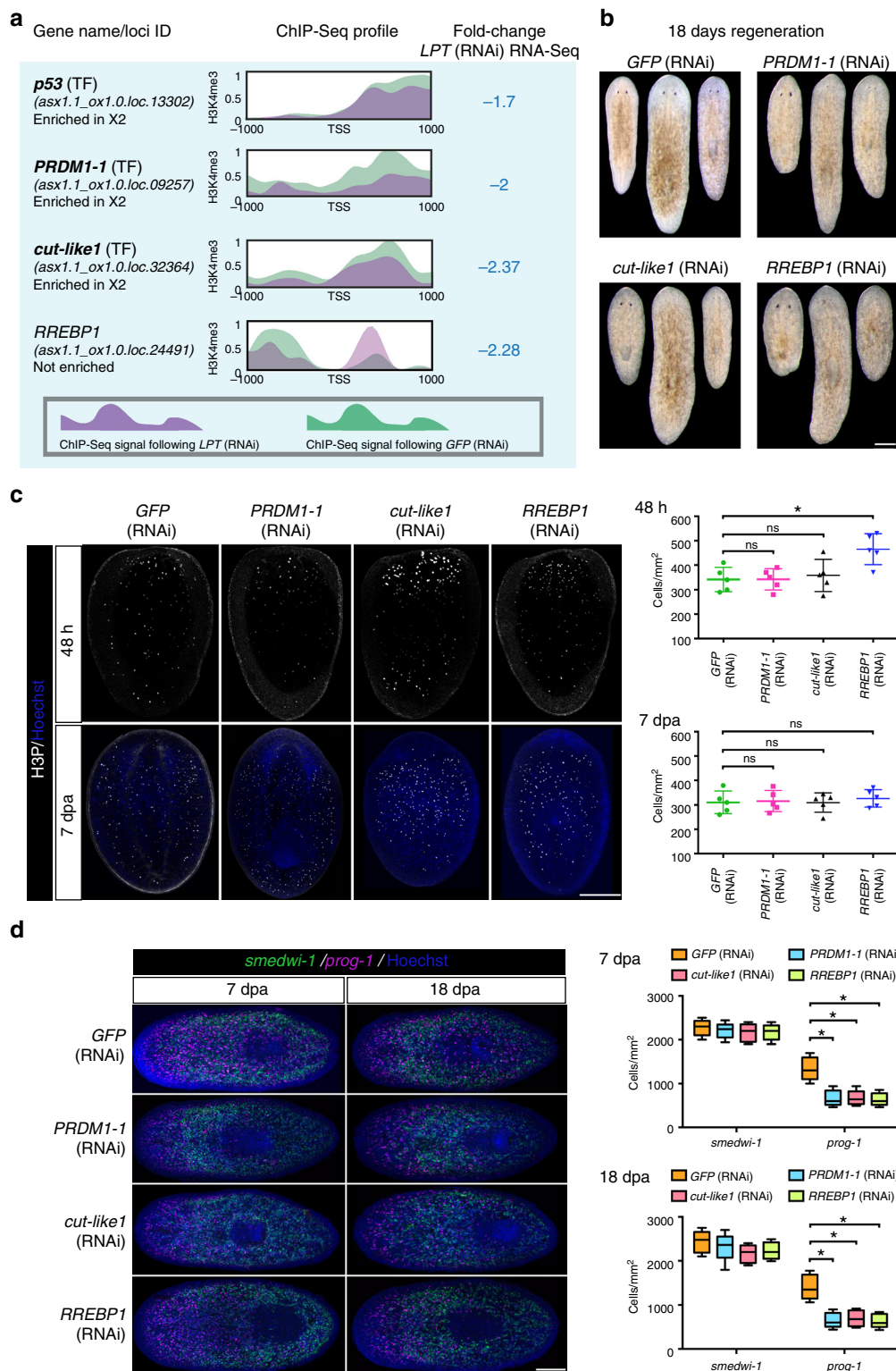


Fig. 10 LPT(RNAi) down-regulates the expression of genes involved in stem cell proliferation and differentiation. **a** Examples of genes significantly ($p < 0.05$) down-regulated in G2/M stem cells following LPT(RNAi). The ChIP-seq profile for H3K4me3 in the 2 kb region around the TSS of each gene is presented. Bold font of a gene name illustrates an example where there is a correlation between H3K4me3 profile and down-regulation in RNA-seq data. 'TF' stands for 'transcription factor'. **b** Representative bright field images of 18 day regenerating animals following different RNAi. PRDM1-1(RNAi), cut-like1 (RNAi) and RREBP1(RNAi) animals show defective anterior regeneration compared to GFP(RNAi) animals. **c** Representative images of mitotic cells (labeled by anti-H3P antibody) in different RNAi animals at 48h and 7 days post-amputation. Graphs show mitotic cell quantification. Each dot represents average number of mitotic cells in single worm ($n = 5$). Lines and error bars indicate mean and SD. Student's t test: $*p < 0.05$, ns is not significant. **d** Representative images showing stem cells (*smedwi-1*⁺) and early epidermal progeny (*prog-1*⁺) at 7 and 18 day regenerating animals from different RNAi conditions. Graphs show quantification of stem cells and early epidermal progeny ($n = 5$). Lines and error bars indicate mean and SD. Student's t test: $*p < 0.05$. Scale bars: 200 μ m

where changes in methylation status were both consistent and inconsistent with changes in transcript levels (Figs. 9a, 10a, Supplementary Figure 7a, Supplementary Figure 9a). For example, we find that the up-regulated expression of the planarian orthologs of the transcription factors *Elf5* and *pituitary homeobox (pitx)* are associated with increased levels of TSS-proximal H3K4me3 signal following *LPT*(RNAi) (Fig. 9a). Furthermore, up-regulation of some X1-enriched genes, such as *pim-2-like*, is associated with a decrease in H3K4me1 signal on the TSS, consistent with alleviated repression (Fig. 9a). On the other hand, some transcriptional changes following *LPT*(RNAi), such as the down-regulation of *Ras-responsive element-binding protein 1 (RREBP1)*, are not correlated with the expected alterations in histone modification patterns on promoters (Fig. 10a). Such examples could potentially represent secondary (not related to histone modifications) or enhancer-dependent changes in the *LPT*(RNAi) phenotype. In the absence of similar RNA-seq/ChIP-seq data in mammals, our data provide an important insight beyond the deep evolutionary conservation of MLL3/4 function.

While mis-regulation of well-known oncogenes and tumor suppressors, like *p53* (Fig. 10a), would be expected to broadly correlate with mis-regulation of *Mll3/4* in cancer gene expression datasets, shared correlative expression changes for some selected genes not previously associated with *Mll3/4* loss of function provide some independent evidence of a conserved regulatory program. In the Pomeroy Brain Oncomine dataset⁴⁵, *Mll3* and *p53* expression levels were both significantly down-regulated (Supplementary Figure 10), supporting the functional link between the two genes established by previous work²⁴ and emerging in our study. We identified more *LPT*(RNAi) mis-regulated genes, which have some known association with growth and/or cancer, but have not been previously implicated in *Mll3/4* loss of function phenotypes. An investigation of the Brune⁴⁶ and Compagno⁴⁷ lymphoma Oncomine datasets demonstrates that these genes are mis-regulated in different cancers in a similar manner to the mis-regulation we observe in planarians stem cells as a consequence of a decreased *LPT* expression (Supplementary Figure 10).

One gene of interest in this data set was the planarian *pitx* gene ortholog. In planarians, *pitx* is expressed in the serotonergic neuronal precursor cells^{48,49} and is required for their differentiation. Thus, *pitx* is not directly implicated in planarian stem cell proliferation, but rather in differentiation. Nonetheless, *pitx*'s up-regulation was of great interest to us since in human medulloblastomas down-regulation of *Mll3* and over-expression of *pitx2* are co-occurrences (Pomeroy Brain Oncomine dataset⁴⁵ (www.oncomine.org)) (Fig. 9b). To investigate the cellular basis for *pitx* overexpression, we performed FISH for this gene in *LPT*(RNAi) animals. We observed an accumulation of *pitx*-positive cells in *LPT*(RNAi) regenerates (Fig. 9c). Given that production of terminally differentiated serotonergic neurons is decreased (Fig. 3d), the increase of *pitx*-positive cells following *LPT*(RNAi) marks the accumulation of serotonergic neuronal precursors that fail to differentiate. We conclude that planarian *LPT* normally regulates *pitx*-mediated differentiation of serotonergic neurons and that regulation of *pitx* is a possible example of a conserved feature of MLL3/4 function that is mis-regulated in some cancer types⁴⁵.

Double RNAi validates overexpressed target genes. In the planarian model, up-regulated genes in our data set provided the opportunity to identify genes whose overexpression contributes to the *LPT/Mll3/4* loss of function phenotype using double RNAi experiments. In addition to the transcription factor *E74-like factor 5 (Elf5)*, we observed up-regulation of planarian orthologs of

other developmental or cancer associated genes. Among them were orthologs of the serine/threonine kinase oncogene *pim-2* (two genes called *Smed-pim-2* and *Smed-pim-2-like*), a paralog of the epigenetic regulator *Suppressor of zeste (Su(z)12)*, the transcription factors *ETS4* and *FoxA1* and an ULK2-like serine/threonine protein kinase (Fig. 9a, Supplementary Figure 7a). Our ChIP-seq data suggested *pim-2*, *ETS4* and *ULK2-like* are either indirectly regulated or regulated by *LPT* at enhancers, while *pim-2-like*, *FoxA1*, *Su(z)12 paralog* and *Elf5* appear to be direct targets of *LPT* as their increased expression was associated with either reduced H3K4me1 or increased H3K4me3 signal at promoters (Fig. 9a, Supplementary Figure 7a).

We decided to focus on these genes because they all have a known role in a wide range of cancers. Overexpression of the cell fate decision determinant *Elf5* is a known driving force behind breast cancer progression and metastasis⁵⁰. The PIM family of proteins is involved in the integration of growth and survival signals⁵¹ and their overexpression has been associated with hematological malignancies and solid tumors⁵². Inhibition of PIM2 is a promising avenue for the treatment of multiple myeloma⁵¹. *Su(z)12* is part of the ubiquitous polycomb repressive complex 2 (PRC2) and is overexpressed in numerous cancers, including gastric cancer where *Su(z)12* levels were associated with increased metastasis and unfavorable survival prognosis⁵³. The ETS family of transcription factors has demonstrated significant involvement in all stages of tumorigenesis⁵⁴, while *FoxA1* is known as a “pioneer transcription factor in organogenesis and cancer progression”⁵⁵. Finally, ULK2 is an autophagy regulator overexpressed in prostate cancer cells⁵⁶. This selected panel of genes represents some of the best candidates for major effects amongst those genes with significant up-regulation in expression following *LPT*(RNAi).

In order to test whether the up-regulated expression of any of these genes is a potentially significant contributor to the *LPT*(RNAi) outgrowth phenotype, we performed *LPT*(RNAi) rescue experiments in the form of double RNAi knockdowns (Fig. 9d–f, Supplementary Figure 7). At 48 h post-amputation, *LPT*(RNAi) regenerates have a significantly increased NB proliferation (Fig. 4a, b) and so do *GFP/LPT*(RNAi) double knockdown animals (Fig. 9d, e). Whereas *pim-2/LPT*(RNAi), *ETS4/LPT*(RNAi), *FoxA1/LPT*(RNAi), *ULK2-like/LPT*(RNAi) and *Su(z)12 paralogue/LPT*(RNAi) regenerates still have elevated NB proliferation, both *pim-2-like/LPT*(RNAi) and *Elf5/LPT*(RNAi) regenerates have a significantly decreased NB proliferation compared to *GFP/LPT*(RNAi) (Fig. 9d, e and Supplementary Figure 7b). Furthermore, not only did *pim-2-like/LPT*(RNAi) and *Elf5/LPT*(RNAi) regenerating animals show improved blastema formation (Supplementary Figure 7c), but also less than half as many animals in these two conditions went on to form outgrowths compared to *GFP/LPT*(RNAi) (Fig. 9f), demonstrating a rescue of the outgrowth phenotype. Importantly, individual knockdown of *Elf5* and *pim-2-like* did not lead to regenerative, proliferation or outgrowth-related defects (Supplementary Figure 8). These findings suggest that the up-regulation of both *pim-2-like* and *Elf5* is specifically involved in driving the *LPT*(RNAi) outgrowth phenotype and demonstrate the utility of our data for validating the role of MLL3/4 targets.

Defects caused by transcription factor down-regulation. We chose a selection of genes down regulated by *LPT*(RNAi) for further investigation to see if they might contribute to the observed *LPT*(RNAi) phenotype (Supplementary Figure 9a–d, Fig. 10a–d). *LPT*(RNAi) leads to the down-regulation of the tumor suppressor *p53*, *PRDM1-1* and *cut-like1* in X1 stem cells (Fig. 10a). These genes' down-regulation is associated with a

decrease in H3K4me3 levels on their promoters. On the other hand, RREBP1's down-regulation, does not correlate with the amount of H3K4me3 on its promoter. While the function of *p53* in planarians has already been described⁵⁷, the role of *PRDM1-1*, *cut-like1* and *RREBP1* remain unexplored in planarian biology. Knockdown of each of these three genes resulted in impaired regeneration, characterized mainly by the inability to form eyes (Fig. 10b). *RREBP1*(RNAi) resulted in significantly increased proliferation compared to control, while *cut-like1*(RNAi) and *PRDM1-1*(RNAi) animals did not show a significant change in mitotic cell numbers (Fig. 10c). All three knockdown conditions showed a decreased number of *prog-1*-positive epidermal progenitors, but intact *smedwi-1*-positive cell numbers (Fig. 10d). These findings suggest that the observed differentiation phenotypes are not associated with an inability to maintain the stem cell pool following knockdown. Instead, stem cells seem to be restricted in their ability to differentiate correctly.

Our data suggest that decreased expression of *PRDM1-1*, *cut-like1* and *RREBP1* in LPT(RNAi) animals could be contributing to the differentiation defects seen after perturbation of LPT/Trr/MLL3/4 function, and that downregulation RREBP1 may be contributing to the early over proliferation of stem cells observed in LPT(RNAi).

Discussion

In mammals, *Mll3* and *Mll4* have been implicated in different malignancy landscapes¹⁹, with clear evidence for tumor suppressor roles in mammalian systems^{20,24}. However, relatively little is known about how these effects are mediated. Our study demonstrates that loss of function of the planarian LPT (an *Mll3/4* ortholog) also results in the emergence of an outgrowth phenotype characterized by differentiation and proliferation defects. Our work also shows that LPT, TRR-1, and TRR-2 control differentiation to form the gut, eyes, brain, and pharyngeal cellular lineages. Future work in planarians combining ChIP-seq and RNA-seq will allow closer investigation of these and other epigenetic effects on stem cell differentiation.

We found that clusters of mitotic cells preceded the appearance of outgrowths in LPT(RNAi) regenerating animals, possibly pre-empting where the outgrowths would subsequently form. The observation of clusters of cells and the formation of outgrowths in some, but not all, RNAi animals is evidence of heterogeneity in stem cell responses to LPT(RNAi). This may reflect the stochastic nature of the broad genome-wide epigenetic changes mediated by MLL3/4 proteins that will lead to variability between cells after knockdown, such that only some NBs cycle out of control and cause outgrowths after the initial proliferative peaks associated with regeneration. A contributory cause to outgrowth formation in addition to proliferation could be failure of NBs to differentiate appropriately and instead continue to cycle at inappropriate positions. We also observed that outgrowth tissue contained different classes of stem cells. Among these stem cells, the presence of *sigma* NBs, thought to include truly pluripotent stem cells⁴¹, is of particular significance. When mis-regulated, these cells could share fundamental similarities with cancer stem cells (CSCs), thought to be founder cells in human malignancies⁵⁸. CSCs have been described as one of the main factors in cancer aggressiveness and resistance to treatment⁵⁹. Studying such cells in a simple in vivo stem cell model, provided by the planarian system, should bring further insight into important control mechanisms that are mis-regulated in different cancers. Our work here provides a useful example of this approach.

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

(RNAi) were mostly associated with cell proliferation, differentiation, and metabolic processes. A subset of mis-regulated genes where RNA-seq and ChIP-seq data correlate is likely a direct consequence of LPT(RNAi) affecting promoter histone methylation status. Genes with altered expression where there is no such correlation, may represent indirect (secondary) changes or, alternatively, may have enhancers that have altered histone modifications as a result of LPT(RNAi). Future work will develop the use of planarians as a model of epigenetic gene regulation and it should also be possible to study enhancer function and evolution.

Among mis-regulated genes, we saw many tumor suppressors with reduced expression and oncogenes with increased expression. We also found a number of genes, including the transcription factor *pitx*, which were similarly mis-regulated in LPT(RNAi) planarians and human cancers with reduced MLL3 expression. Together, these data suggest that, as well as physiological function in controlling stem cell proliferation, there may be deep regulatory conservation of MLL3/4 function in animal stem cells. One advantage of our approach is that we were able to sample expression and histone states in NBs at an early time point before tumors formed. This could be an advantage for the identifying targets that act early to drive hyperplasia, rather than later secondary regulatory changes.

As proof of principle that genes mis-regulated by LPT(RNAi) directly contributed to the phenotype, we performed double RNAi experiments. Planarian homologs of the oncogene *pim-2*, called *Smed-pim-2* and *Smed-pim-2-like*, that were overexpressed in stem cells following LPT(RNAi), were chosen as likely candidates, based on previous data on the roles of these genes from mammals⁶⁰. We found that double RNAi with *pim-2-like*, was able to ameliorate LPT loss of function over-proliferation and outgrowth phenotypes induced by LPT(RNAi). In addition, double knockdown with the breast cancer oncogene *Elf5*⁵⁰ resulted in an even more dramatic rescue of the LPT loss of function phenotype. This provides strong support for the hypothesis that the over-expression of these two genes was significant in driving stem cell hyperplasia. Future work can now study how these genes function in stem cells and why over-expression leads to over proliferation.

We also identified downstream candidates that could be contributing to the lack of differentiation phenotype following LPT(RNAi). Knockdown of *PRDM1-1*, *cut-like* and *RREBP-1* (down-regulated in LPT(RNAi) animals) indicated their mis-regulation might contribute to the decreased epidermal differentiation observed following LPT knockdown.

Together these experiments, demonstrate the value of our approach for identifying potential downstream targets and implicate regulatory interactions driving the *Mll3/4* loss of function phenotype. These targets can now be tested for conservation in mammalian experimental systems.

Overall, our work shows how perturbation of a conserved physiological role of LPT leads to mis-regulation of genes well-known to control cell proliferation, causing hyperplasia and tumors in planarians. We find other genes that are mis-regulated in planarians that are also similarly mis-regulated in cancer expression studies that have reduced *Mll3* expression. Some of these, like *pitx*, may represent deeply conserved regulatory interactions. In the absence of similar RNA-seq/ChIP-seq data in mammals, our data provide an important insight into *Mll3/4* loss of function, as well as revealing a deep evolutionary conservation in animal stem cells. These findings demonstrate the strength of the planarian system for understanding fundamental animal stem cell biology highly relevant to cancer and the potential for investigation of epigenetic mechanisms in stem cells.

Methods

Animal husbandry. Asexual freshwater planarians of the species *S. mediterranea* were used. The culture was maintained in 1x Montjuic salts water⁶¹. Planarians were fed organic calf liver once a week. After every feeding, the water was changed. Planarians were starved for 7 days prior to each experiment. They were also starved throughout the duration of each experiment.

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

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

DsRNA was delivered via microinjection using Nanoject II apparatus (Drummond Scientific) with 3.5" Drummond Scientific (Harvard Apparatus) glass capillaries pulled into fine needles on a Flaming/Brown Micropipette Puller (Patterson Scientific). Each animal received around 100 nl dsRNA each day. *H2B* (RNAi) was performed for three consecutive days, as per Solana et al.⁷ protocol. For single and double *LPT*, *trr-1* and *trr-2* knockdown, a course of 7 days of microinjections was performed (3 consecutive days + 2 days rest + 4 consecutive days). *Set1* (RNAi) and *utx* (RNAi) were performed for 4 consecutive days. For all other single and double knockdowns, a course of 10 days of microinjections was performed (3 consecutive days + 4 days rest + 3 consecutive days).

Primers used for amplification of DNA for dsRNA synthesis can be found in Supplementary Table 1.

In situ hybridization. RNA probes labeled with digoxigenin and fluorescein were generated via anti-sense transcription of DNA cloned in pCRII (Invitrogen) or PGemTEasy (Promega) vector. In situ hybridization was performed after fixation in 4% paraformaldehyde⁶² for fluorescent experiments. For *LPT*, *trr-1*, *trr-2*, *sigma*, *zeta*, and *gamma* fluorescent in situ procedures, a pooled probes method was used using PCR products from across transcriptional units as templates for RNA probes⁴¹. Colorimetric in situ hybridization procedures were performed using alkaline phosphatase based detection of hybridized RNA probes⁶³. Primers used for amplification of DNA for RNA probe synthesis can be found in (Supplementary Table 1).

Immunohistochemistry. Immunohistochemistry was performed after either paraformaldehyde of Carnoy fixative based fixation⁶⁴. Antibodies used were: anti-H3P (phosphorylated serine 10 on histone H3; Millipore; 09-797; 1:1000 dilution), anti-VC1 (kindly provided by Prof. Hidefumi Orii (check title); 1:10,000 dilution), anti-SMEDWI-1 (kindly provided by Prof. Jochen Rink; 1:500 dilution), anti-SYNORF-1 (3C11; Developmental Studies Hybridoma Bank; 1:50 dilution), anti-acetylated tubulin (Developmental Studies Hybridoma Bank; 1:200 dilution).

Imaging and image analysis. Colorimetric images were taken on Zeiss Discovery V8 (Carl Zeiss) microscope with a Canon EOS 600D or Canon EOS 1200D camera. Fluorescent images were taken on either Inverted Olympus FV1000 or FV1200 Confocal microscope. Cells were counted via Adobe Photoshop CS6 or FIJI software and the count was normalized to image area in mm².

Flow cytometry. A simple FACS protocol^{8,65} was used on dissociated planarians cells using hoechst, calcein and propidium iodide to sort living cells in G2/M of the cell cycle. A FACS Aria III machine equipped with a violet laser was used for the sort. BD FACSDiva and FlowJo software was used for analysis and gate-setting.

Western blot. 2x Laemmli buffer (Sigma Aldrich), 1M DTT and complete protease inhibitors (Roche) were used for protein extraction from 10 to 15 animals per condition. Protein extract was quantified with Qubit Protein Assay kit (Thermo Fisher Scientific). NuPAGE Novex 4–12% Bis-Tris protein gels (Thermo Fisher Scientific) were used, followed by a wet transfer in a Mini Trans-Blot Electrophoretic Transfer Cell machine. Ponceau S (Sigma Aldrich) whole-protein stain was used prior to antibody incubation. The antibodies used were: anti-H3 (unmodified histone H3; rabbit polyclonal; Abcam; ab1791; 1:10,000 dilution), anti-H3K4me3 (rabbit polyclonal; Abcam; ab8580; 1:1000 dilution), anti-H3K4me1 (rabbit polyclonal; Abcam; ab8895; 1:1000 dilution), anti-H3K27me3 (mouse monoclonal; Abcam; ab6002; 1:1000 dilution), anti-mouse IgG HRP-linked antibody (Cell Signalling; 7076P2), anti-rabbit IgG HRP-linked antibody (Cell Signalling; 7074P2). Western blot experiments were done to validate the specificity of the histone modification antibodies used for ChIP-seq (Supplementary Figure 11). The rationale behind these experiments was to knock down a methylase (*set1*), part of a methylase complex (*LPT*) or a demethylase (*utx*) known to affect H3K4 and H3K27 methylation levels and to observe whether global H3K4me1, H3K4me3, and H3K27me3 levels would change in the expected way.

ChIP-seq. In total 600,000–700,000 planarian G2/M cells were FACS-sorted (using 3-day knockdown regenerates) in PBS and pelleted at 4 °C. During the pelleting, S2 cells were added (corresponding to roughly 15% of the number of planarian ×1 cells) for the purpose of downstream data normalization⁴⁴. Samples were then used to prepare Illumina sequencing libraries⁹. The process is summarized in Supplementary Figure 12. The libraries were sequenced on an Illumina NextSeq machine. Three biological replicates were prepared. The raw reads are available in the NCBI Short Read Archive (PRJNA338116).

RNA-seq. In total 300,000 G2/M cells were FACS-sorted in RNALater (Ambion) from knockdown and control RNAi animals at 3 days of regeneration. Cells were pelleted at 4 °C and Trizol-based total RNA extraction was performed. The amount of total RNA used for each library preparation was 0.8–1 µg. Illumina TruSeq Stranded mRNA LT kit was used for library preparation according to the manufacturers instructions. Libraries were quantified with Qubit, Agilent Bioanalyzer and KAPA Library Quantification qPCR kit. Samples were sequenced on an Illumina NextSeq machine. Two biological replicates were prepared. The raw reads are available in the Short Read Archive (PRJNA338115).

ChIP-seq data analysis. ChIP-seq reads were trimmed with Trimmomatic 0.32⁶⁶ and aligned to the *S. mediterranea* SmedGD asexual genome 1.1⁶⁰ and *D. melanogaster* genome r6.10⁶⁷ with BWA mem 0.7.12. Picard tools 1.115 was used to remove read duplicates after mapping. Python scripts were used to filter and separate out read pairs belonging to either genome. ChIP-seq coverage tracks were then generated and normalized according to Orlando et al. in order to account for any technical variation between samples⁴⁴. For more in-depth methods, including code, refer to Supplementary Software.

RNA-seq data analysis. Raw reads were trimmed with Trimmomatic 0.32⁶⁶ and pseudo-aligned to a set of asexual genome annotations⁸ with Kallisto 0.42⁶⁸. Differential expression was subsequently performed with Sleuth 0.28.1⁶⁹. For more in-depth methods, including code, refer to Supplementary Software.

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

For analysis of RNA-seq data, Wald's test (as part of the Sleuth⁶⁹ software) was used for assessing differential expression. Spearman's rank correlation was used for assessing the correlation between RNA-seq and ChIP-seq data. Hypergeometric tests were used for assessing gene enrichment in the RNA-seq data.

Code Availability. Code used in the analysis of data are presented in Supplementary Software.

Data availability

The ChIP-seq and RNA-seq datasets are deposited in the Short Read Archive database under the accession codes: PRJNA338116 and PRJNA338115, respectively. The 'Pomeroy Brain' dataset⁴⁵ from the oncomine database (<https://www.oncomine.com>) was used for assessing expression level of potential target genes and *Mil3* in human medulloblastoma versus normal cerebellum. All other data are made available within the article and its supplementary information.

Received: 10 May 2017 Accepted: 13 August 2018
Published online: 07 September 2018

References

1. Aboobaker, A. A. Planarian stem cells: a simple paradigm for regeneration. *Trends Cell Biol.* **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. *EMBO J.* **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 Res.* **21**, 1193–1200 (2011).
5. Labbé, R. M. et al. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**, 1734–1745 (2012).
6. Solana, J. et al. Conserved functional antagonism of CELF and MBNL proteins controls stem cell-specific alternative splicing in planarians. *eLife* **5**, 1193 (2016).

7. Solana, J. et al. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. *Genome Biol.* **13**, R19 (2012).
8. Dattani, A. et al. Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells. Preprint at <https://www.biorxiv.org/content/early/2018/08/09/122135> (2018).
9. Shibata, N. et al. Inheritance of a nuclear PIWI from pluripotent stem cells by somatic descendants ensures differentiation by silencing transposons in planarian. *Dev. Cell.* **37**, 226–237 (2016).
10. Salvetti, A. DjPum, a homologue of *Drosophila* Pumilio, is essential to planarian stem cell maintenance. *Development* **132**, 1863–1874 (2005).
11. Reddien, P. W. Specialized progenitors and regeneration. *Development* **140**, 951–957 (2013).
12. Jaber-Hijazi, F. et al. Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Dev. Biol.* **384**, 141–153 (2013).
13. Scimone, M. L., Meisel, J. & Reddien, P. W. The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian *Schmidtea mediterranea*. *Development* **137**, 1231–1241 (2010).
14. Zhu, S. J., Hallows, S. E., Currie, K. W., Xu, C. & Pearson, B. J. A mex3 homolog is required for differentiation during planarian stem cell lineage development. *eLife* **4**, 304 (2015).
15. Cowles, M. W., Omuro, K. C., Stanley, B. N., Quintanilla, C. G. & Zayas, R. M. COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians. *PLoS Genet.* **10**, e1004746–12 (2014).
16. Barberan, S., Fraguas, S. & Cebrià, F. The EGFR signaling pathway controls gut progenitor differentiation during planarian regeneration and homeostasis. *Development* **143**, 2089–2102 (2016).
17. Abnave, P. et al. Epithelial-mesenchymal transition transcription factors control pluripotent adult stem cell migration in vivo in planarians. *Development* **144**, 3440–3453 (2017).
18. Parsons, D. W. et al. The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435–439 (2011).
19. Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
20. Chen, C. et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **25**, 652–665 (2014).
21. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
22. Herz, H. M. et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homologue of mammalian Mll3/Mll4. *Genes Dev.* **26**, 2604–2620 (2012).
23. Hu, D. et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell Biol.* **33**, 4745–4754 (2013).
24. Lee, J. et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proc. Natl Acad. Sci. USA* **106**, 8513–8518 (2009).
25. Cheng, J. et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol. Cell* **53**, 979–992 (2014).
26. Chauhan, C., Zraly, C. B., Parilla, M., Diaz, M. O. & Dingwall, A. K. Histone recognition and nuclear receptor co-activator functions of *Drosophila* Cara Mitad, a homologue of the N-terminal portion of mammalian MLL2 and MLL3. *Development* **139**, 1997–2008 (2012).
27. Hubert, A. et al. Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone methyltransferases. *Epigenetics* **8**, 79–91 (2013).
28. Duncan, E. M., Chitsazan, A. D., Seidel, C. W. & Alvarado, A. S. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. *CellReports* **13**, 2741–2755 (2015).
29. Lee, J.-E. et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2**, 2817–2825 (2013).
30. Wang, C. et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl Acad. Sci. USA* **113**, 11871–11876 (2016).
31. Denisov, S. et al. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526–537 (2014).
32. Hsieh, J. J.-D., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S. J. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol. Cell Biol.* **23**, 186–194 (2003).
33. Thirman, M. J. et al. Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N. Engl. J. Med.* **329**, 909–914 (1993).
34. Sobulo, O. M. et al. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc. Natl Acad. Sci. USA* **94**, 8732–8737 (1997).
35. Lee, S. et al. Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc. Natl Acad. Sci. USA* **103**, 15392–15397 (2006).
36. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31**, 35–40 (2006).
37. Lee, S., Lee, J., Lee, S.-K. & Lee, J. W. Activating Signal Cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. *Mol. Endocrinol.* **22**, 1312–1319 (2008).
38. Wenemoser, D. & Reddien, P. W. Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Dev. Biol.* **344**, 979–991 (2010).
39. Oviedo, N. J., Pearson, B. J., Levin, M. & Sanchez Alvarado, A. Planarian PTEN homologs regulate stem cells and regeneration through TOR signaling. *Dis. Models Mech.* **1**, 131–143 (2008).
40. Gonzalez-Estevéz, C. et al. SMG-1 and mTORC1 act antagonistically to regulate response to injury and growth in planarians. *PLoS Genet.* **8**, e1002619–17 (2012).
41. van Wolfswinkel, J. C., Wagner, D. E. & Reddien, P. W. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Stem Cell* **15**, 326–339 (2014).
42. Zink, D., Fische, A. H. & Nickerson, J. A. Nuclear structure in cancer cells. *Nat. Rev. Cancer* **4**, 677–687 (2004).
43. Lee, M. G. et al. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).
44. Orlando, D. A. et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *CellReports* **9**, 1163–1170 (2014).
45. Pomeroy, S. L. et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**, 436–442 (2002).
46. Brune, V. et al. Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. *J. Exp. Med.* **205**, 2251–2268 (2008).
47. Compagno, M. et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717–721 (2009).
48. Currie, K. W. & Pearson, B. J. Transcription factors *lhx1/5-1* and *pitx* are required for the maintenance and regeneration of serotonergic neurons in planarians. *Development* **140**, 3577–3588 (2013).
49. März, M., Seebeck, F. & Bartscherer, K. A *Pitx* transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. *Development* **140**, 4499–4509 (2013).
50. Gallego-Ortega, D. et al. ELF5 drives lung metastasis in luminal breast cancer through recruitment of Gr1+ CD11b+ myeloid-derived suppressor cells. *PLoS Biol.* **13**, e1002330 (2015).
51. Nair, J. R. et al. Novel inhibition of PIM2 kinase has significant anti-tumor efficacy in multiple myeloma. *Leukemia* **31**, 1715–1726 (2017).
52. Jiménez-García, M. P. et al. The role of PIM1/PIM2 kinases in tumors of the male reproductive system. *Sci. Rep.* **6**, 38079 (2016).
53. Xia, R. et al. SUZ12 promotes gastric cancer cell proliferation and metastasis by regulating KLF2 and E-cadherin. *Tumour Biol.* **36**, 5341–5351 (2015).
54. Sizemore, G. M., Pitarresi, J. R., Balakrishnan, S. & Ostrowski, M. C. The ETS family of oncogenic transcription factors in solid tumours. *Nat. Rev. Cancer* **17**, 337–351 (2017).
55. Zhang, G. et al. FOXA1 defines cancer cell specificity. *Sci. Adv.* **2**, e1501473–e1501473 (2016).
56. John Clotaire, D. Z. et al. MiR-26b inhibits autophagy by targeting ULK2 in prostate cancer cells. *Biochem. Biophys. Res. Commun.* **472**, 194–200 (2016).
57. Pearson, B. J. & Alvarado, A. S. A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* **137**, 213–221 (2009).
58. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
59. Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nat. Rev. Cancer* **5**, 275–284 (2005).
60. Robb, S. M. C., Gotting, K., Ross, E. & Sánchez Alvarado, A. SmedGD 2.0: the *Schmidtea mediterranea* genome database. *Genesis* **53**, 535–546 (2015).
61. Cebria, F. Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development* **132**, 3691–3703 (2005).
62. King, R. S. & Newmark, P. A. In situ hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea*. *BMC Dev. Biol.* **13**, 8 (2013).
63. Gonzalez-Estevéz, C., Arseni, V., Thambyrajah, R. S., Felix, D. A. & Aboobaker, A. A. Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int. J. Dev. Biol.* **53**, 493–505 (2009).
64. Cebria, F. & Newmark, P. A. Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. *Development* **134**, 833–837 (2007).

65. Romero, B. T., Evans, D. J. & Aboobaker, A. A. *Progenitor Cells* Vol. **916**, 167–179 (Humana Press, 2012).
66. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
67. Attrill, H. et al. FlyBase: establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Res.* **44**, D786–D792 (2016).
68. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
69. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-Seq incorporating quantification uncertainty. *Nat. Methods* **14**, 678–690 (2017).

Acknowledgements

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

Authors contributions

A.A.A., P.A., and Y.M. conceived and designed the study. Y.M. and P.A. performed the experiments. D.K. performed the bioinformatics analyses. S.H. participated in the optimization of the ChIP-seq protocol. A.G.L. provided technical support. F.J.H. performed initial work on the project, including generating the first *LPT*(RNAi) results. N.K. provided technical support. Y.M., P.A., and A.A.A. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-06092-6>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018