Cell-type diversity and regionalized gene expression in the planarian intestine revealed by laser-capture microdissection transcriptome profiling

Authors:

David J. Forsthoefer^{1,2,*}, Nicholas I. Cejda¹, Umair W. Khan^{2,#}, Phillip A. Newmark^{2,^,*}

Affiliations:

¹ Genes and Human Disease Research Program, Oklahoma Medical Research Foundation,
Oklahoma City, Oklahoma, USA

² Howard Hughes Medical Institute, Department of Cell and Developmental Biology, University
of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[#] Current address: Graduate Program in Cell and Molecular Biology, University of Wisconsin–
Madison, Madison, Wisconsin, USA

[^] Current address: Howard Hughes Medical Institute, Morgridge Institute for Research,
Department of Integrative Biology, University of Wisconsin-Madison, Madison, Wisconsin, USA

* Co-corresponding authors:

DJF, david-forsthoefer@omrf.org

PAN, pnewmark@morgridge.org

Abstract

Organ regeneration requires precise coordination of new cell differentiation and remodeling of uninjured tissue to faithfully re-establish organ morphology and function. An atlas of gene expression and cell types in the uninjured state is therefore an essential pre-requisite for understanding how damage is repaired. Here, we use laser-capture microdissection (LCM) and RNA-Seq to define the transcriptome of the intestine of *Schmidtea mediterranea*, a planarian flatworm with exceptional regenerative capacity. Bioinformatic analysis of 1,844 intestine-enriched transcripts suggests extensive conservation of digestive physiology with other animals, including humans. Comparison of the intestinal transcriptome to purified absorptive intestinal cell (phagocyte) and published single-cell expression profiles confirms the identities of known intestinal cell types, and also identifies hundreds of additional transcripts with previously undetected intestinal enrichment. Furthermore, by assessing the expression patterns of 143 transcripts *in situ*, we discover unappreciated mediolateral regionalization of gene expression and cell-type diversity, especially among goblet cells. Demonstrating the utility of the intestinal transcriptome, we identify 22 intestine-enriched transcription factors, and find that several have distinct functional roles in the regeneration and maintenance of goblet cells. Furthermore, depletion of goblet cells inhibits planarian feeding and reduces viability. Altogether, our results show that LCM is a viable approach for assessing tissue-specific gene expression in planarians, and provide a new resource for further investigation of digestive tract regeneration, the physiological roles of intestinal cell types, and axial polarity.

Introduction

Physical trauma, disease, and aging can damage internal organs. Some animals are endowed with exceptional regenerative capacity, and can repair or even completely replace damaged tissue (1-3). Regeneration requires precise spatial and temporal control over the differentiation of distinct cell types, as well as remodeling of uninjured tissue. Furthermore, individual cell types can respond uniquely to injury and play specialized signaling roles that promote regeneration (4-8). Therefore, characterization of an organ's composition and gene expression at a cellular level in the uninjured state is an essential step in unraveling the mechanisms required for faithful re-establishment of organ morphology and physiology.

Driven by the recent application of genomic and molecular methods, the planarian flatworm *Schmidtea mediterranea* has become a powerful model in which to address the molecular and cellular underpinnings of organ regeneration (9-13). In response to nearly any type of surgical amputation injury, pluripotent stem cells called neoblasts proliferate and differentiate, regenerating brain, intestine, and other tissues lost to injury (14, 15). In addition, pre-existing tissue undergoes extensive remodeling and re-scaling through both collective migration of post-mitotic cells in undamaged tissues, as well as proportional loss of cells through apoptosis (16). These processes are coordinated to achieve re-establishment of proportion, symmetry, and function of planarian organ systems within a few weeks after injury (17).

The planarian intestine is a prominent organ whose highly branched morphology, simple cellular composition, and likely cell non-autonomous role in neoblast regulation make it a compelling model for addressing fundamental mechanisms of regeneration. In uninjured planarians, a single anterior and two posterior primary intestinal branches project into the head and tail, respectively, with secondary, tertiary, and quaternary branches extending towards lateral body margins (18, 19). Growth and regeneration of intestinal branches require

considerable remodeling of pre-existing tissue (19). Remodeling is governed by axial polarity cues (16, 20), ERK and EGFR signaling pathways (21-23), cytoskeletal regulators (24), and interactions with muscle (25-28). However, the mechanisms by which post-mitotic intestinal cells sense and respond to extrinsic signals are only superficially understood.

New intestinal cells (the progeny of neoblasts) differentiate at the severed ends of injured gut branches, as well as in regions of significant remodeling, providing an intriguing example of how differentiation and remodeling must be coordinated to achieve integration of old and new tissue (19). Only three cell types comprise the intestinal epithelium: secretory goblet cells, absorptive phagocytes (29-31), and a recently identified population of basally located “outer” intestinal cells (32). Transcription factors expressed by intestinal progenitors (“gamma” neoblasts) and their progeny have been identified (24, 32-36). However, only the EGF receptor *egfr-1* (23) has been shown definitively to be required for integration of intestinal cells into gut branches, and therefore the functional requirements for differentiation of new intestinal cells are largely undefined.

The intestine may also play a niche-like role in modulating neoblast dynamics. Knockdown of several intestine-enriched transcription factors (*nkx2.2*, *gata4/5/6-1*) causes reduced blastema formation and/or decreased neoblast proliferation (24, 37). Similarly, knockdown of the intestine-enriched HECT E3 ubiquitin ligase *wwp1* causes disruption of intestinal integrity, reduced blastema formation, and neoblast loss (38). Conversely, knockdown of *egfr-1* causes hyperproliferation and expansion of several neoblast subclasses (23). Because these genes are also expressed by neoblasts and their progeny, careful analysis will be required to distinguish their functions in the stem cell compartment from cell non-autonomous roles in the intestine. Nonetheless, because so few extrinsic signals (39-42) controlling neoblast proliferation have been identified, further investigation of the intestine as a potential source of such cues is warranted.

Addressing these aspects of intestine regeneration and function necessitates development of approaches for purification of intestinal tissue. Previously, we developed a method for purifying intestinal phagocytes from single-cell suspensions derived from planarians fed magnetic beads, enabling characterization of gene expression by this cell type (24). More recently, single-cell profiling of whole planarians has distinguished individual intestinal cell types, as well as transitional markers for neoblasts differentiating along endodermal lineages (32, 35, 36, 43). Both approaches have advanced our understanding of intestinal biology. However, methods that (a) avoid the potentially confounding effects of feeding and dissociation on gene expression and (b) overcome the need to sequence tens of thousands of planarian cells to identify intestinal cells (only 1-3% of all planarian cells, (44)), would further enhance the experimental accessibility of the intestine.

Laser-capture microdissection (LCM) was developed as a precise method for obtaining enriched or pure cell populations from tissue samples, including archived biopsy and surgical specimens (45). Since its introduction, LCM has been used to address a vast array of basic and clinical problems requiring genome, transcriptome, or proteome analysis in specific tissues or cell types (46, 47). LCM requires sample preparation including fixation, histological sectioning, and tissue labeling or staining (48). For subsequent expression profiling, tissue processing must be optimized to maintain morphology and labeling of cells of interest, as well as RNA integrity (48-51). Excision and capture of tissue with infrared and/or ultraviolet lasers is then performed using one of several commercial LCM systems (47).

Here, we report the application of LCM for expression profiling of the planarian intestine. We first identified appropriate tissue-processing conditions for extraction of intact RNA from planarian tissue sections. Then, using RNA-Seq, bioinformatics approaches, and whole-mount in situ hybridization, we characterized the intestinal transcriptome, and discovered previously unappreciated regionalization and diversity of intestinal cell types and subtypes, especially amongst goblet cells. In addition, we identified 22 intestine-enriched transcription factors,

including several required for production and/or maintenance of goblet cells, setting the stage for further studies of this cell type. The planarian intestinal transcriptome is a foundational resource for investigating numerous aspects of intestine regeneration and physiology. Furthermore, the LCM methods we introduce offer an additional, complementary strategy for assessing tissue-specific gene expression in planarians.

Results

Application of laser-capture microdissection to recover RNA from the planarian intestine

Successful application of laser microdissection requires identification of sample preparation conditions that balance the need to extract high quality total RNA with preservation of specimen morphology and the ability to identify tissues or cells of interest. Fixation of whole planarians requires an initial step to relax/kill animals and remove mucus, followed by fixation (52, 53). We tested three commonly used relaxation/mucus-removal treatments and two fixatives (formaldehyde and methacarn), separately and together, using short treatment times in order to minimize potential deleterious effects on RNA (Fig S1A). None of the relaxation treatments detrimentally affected RNA quality, but methacarn (a precipitating fixative) enabled much better RNA recovery than formaldehyde (a cross-linking fixative) (Fig S1A and S1B). Next, we assessed how mucus removal affected morphology and staining of cryosections taken from methacarn-fixed planarians, again using a rapid protocol to minimize RNA degradation (Fig S1C). For all three mucus-removal methods, Eosin Y alone or with Hematoxylin enabled superior demarcation of the intestine, as compared to Hematoxylin alone (Fig S1C). For preservation of morphology, magnesium relaxation was superior; NAc and HCl treatment caused tearing and detachment of intestine from the slide (Fig S1C). Finally, we assessed RNA integrity from laser-microdissected intestine and non-intestine from magnesium-treated,

methacarn-fixed tissue (Fig S1D, Fig 1A and 1B), stained only with Eosin Y to further minimize processing time (Fig S1C). Although additional freezing, cryosectioning, staining, and drying steps required for LCM caused a modest decrease in RNA integrity relative to whole animals (compare Fig S1A and Fig S1D), prominent 18S/28S rRNA peaks (which co-migrate in planarians, as in some other invertebrates (54-57)) (Fig S1D) indicated that the combination of magnesium-induced relaxation, brief methacarn fixation, and rapid Eosin Y staining were suitable for LCM and extraction of RNA of sufficient quality for RNA-Seq.

Identification of intestine-enriched transcripts and mediolateral expression domains

Using our optimized conditions, we laser microdissected intestinal and non-intestinal tissue from four individual planarians (biological replicates) (Fig 1A, 1B and Fig S1D). For this study, we isolated tissue from the anterior of the animal (rostral to the pharynx, planarians' centrally located feeding organ), where intestinal tissue is more abundant. We microdissected tissue from medial and lateral intestine separately, since the intestine ramifies into secondary, tertiary, and quaternary branches along the mediolateral axis, but whether gene expression varies along this axis has not been addressed systematically. We then extracted total RNA, conducted RNA-Seq, and identified transcripts that were preferentially expressed in intestinal vs. non-intestinal tissue (Fig 1C-G) (Table S1A and S1B).

Altogether, we detected 13,136 of 28,069 transcripts in the reference transcriptome (46.8% coverage) in non-intestine, medial intestine, and/or lateral intestine (Fig 1C). Of these, 1,844 were upregulated in either medial or lateral intestine, or both (Fig 1C). Specifically, in medial intestine, 1,748 transcripts were significantly upregulated (fold-change>2 and FDR-adjusted p -value<0.01) compared to non-intestine (Fig 1D). In lateral intestine, 1,627 transcripts were upregulated (Fig 1D). Although most (1,531/1,844) transcripts were upregulated in both medial and lateral intestine, a small subset of transcripts was significantly upregulated *only* in medial (217/1,844) or lateral (96/1,844) intestine (Fig 1D), relative to non-intestine. To further

define medial and lateral transcript enrichment, we calculated a “mediolateral ratio” of the medial and lateral intestine/non-intestine fold-changes (Fig 1E and Table S1A). Although most transcripts were only modestly enriched (<1.5X fold-change enrichment) in medial (1,124) or lateral (567) intestine, a small number of transcripts was >1.5X enriched in medial (97) or lateral (56) intestine (Fig 1E and Table S1A).

We validated RNA-Seq results using whole-mount in situ hybridization (WISH) to test expression in fixed, uninjured planarians (58-60) (Fig 1F and Fig S2). 143/162 transcripts (~88%) had detectable expression in the intestine (Fig S2, Table S1A). Most transcripts were expressed uniformly throughout the intestine (e.g., *ATPase H⁺-transporting accessory protein 2* (*atp6ap2*), *cytochrome P450 2B19* (*cyp2b19*), *cytochrome P450 2A6* (*cyp2a6*), and *prosaposin* (*psap*); Fig 1F, blue borders, and Fig S2). By contrast, transcripts predicted by RNA-Seq to be most enriched (>1.5X) in medial intestine branches (e.g., *carboxypeptidase A2* (*cpa2*), *rapunzel 4* (*rpz4*), and *gastric triacylglycerol lipase* (*lipf*), green borders, Fig 1F) or lateral intestine branches (e.g., a *carboxypeptidase* homolog (*ct14378*), *serine peptidase inhibitor Kunitz type 3* (*spint3*), and a novel gene (“*novel*”), orange borders, Fig 1F) were indeed expressed at higher levels in these regions. Additionally, although we did not explicitly compare anterior and posterior gene expression, we also discovered anteriorly (a *C-type lectin* (*Zgc:171670/dd_79*), Fig S2) and posteriorly (*lysosomal acid lipase* (*lipa/dd_122*), Fig S2) enriched transcripts, consistent with the influence of anteroposterior polarity cues on intestinal morphology (61-66). In summary, using LCM together with RNA-Seq identified >1800 intestine-enriched transcripts, and revealed previously unappreciated regional gene expression domains in the intestine.

Diverse digestive physiology genes are expressed in the planarian intestine

In order to globally characterize functional classes of genes expressed in the planarian intestine, we assigned Gene Ontology (GO) Biological Process (BP) terms to planarian transcripts based on homology to human, mouse, zebrafish, *Drosophila*, and *C. elegans* genes,

and then identified over-represented terms among intestine-enriched transcripts (Fig 2A, 2B, Table S1A, and Table S2A). Highly represented terms fell broadly into seven groups (Fig 2A), all of which are related to the intestine's roles in digestion, nutrient storage and distribution, as well as innate immunity.

Metabolism. Metabolic processes were among the most highly represented: hundreds of transcripts were predicted to regulate catabolism, biosynthesis, and transport of a variety of macromolecules (e.g., lipids and carbohydrates) and small molecules (e.g., amino acids and ions). Consistent with an expected role in the initial macromolecular breakdown of ingested food, transcripts encoding secreted proteins known to regulate luminal digestion of proteins and lipids were intestine enriched, including *pancreatic carboxypeptidase A2* (a metalloexopeptidase) and three *gastric triacylglycerol lipase/lipase F* paralogs (Fig 1F and Table S1A). The intestine also is a major site of neutral lipid storage (29, 67). Transcripts related to lipid metabolism included *apolipoprotein B* paralogs (*apob-1*, Fig 2B) (regulators of lipoprotein particle biogenesis and secretion); *hormone sensitive lipase* (a regulator of diglyceride and cholesteryl ester hydrolysis); several *Niemann-Pick disease type C1/C2 protein* paralogs (intracellular cholesterol transporters); *7-dihydrocholesterol reductase* (cholesterol and steroid biosynthesis); *ceramide glucosyltransferase* (sphingolipid metabolism); and *fatty acid binding protein* (fatty acid transport), whose homologs are upregulated in the intestines of numerous animals (68-72) (Table S1A and Table S2A). Regulators of both complex and simple carbohydrate metabolism were also intestine enriched, such as the glucosidases *lysosomal alpha-glucosidase* (*gaa*, Fig 2B) and *beta-mannosidase*; *sorbitol dehydrogenase* (fructose biosynthesis); and *glycogenin-2*, which synthesizes glycogen, a complex polysaccharide found in cytoplasmic granules of phagocytes that may be an additional energy reserve (67, 73, 74) (Table S2A).

Transport, import, and export. Hundreds of upregulated transcripts were predicted to regulate molecular transport, vesicular trafficking, and organelle-based import and export (Fig

2A, Table S2A). Small-molecule transporters included regulators of iron transport and storage (three *ferritin* paralogs) and the intracellular copper transporter *atox1/cuc-1* (Fig 2B). Over 70 members of the solute carrier family of transmembrane transporters (e.g., *sodium-dependent serotonin transporter/serf*, Fig 2B, and Table S1) were also upregulated, reinforcing the intestine's central role in metabolic processes, and also suggesting a potential role supporting the excretory system in maintaining extracellular solute concentration (65). Enriched regulators of vesicular trafficking included several components of clathrin-associated adaptor protein (AP) complexes, and nearly 30 Ras-related Rab GTPase proteins (Table S1, Table S2A). Other putative regulators of organelle-based import and export included homologs of *phosphatidylinositol-binding clathrin assembly protein (PICALM)*, *epsin-1*, and several *rac* GTPase paralogs (endocytosis and phagocytosis); as well as *exocyst complex components 1, 6b, and 8*, *vesicle-associated membrane protein 3 (vamp3)*, and *ezrin* (exocytosis and secretion). Interestingly, several transcripts encoding SNARE proteins were intestine enriched (e.g., *syntaxin-binding protein 2*, *extended synaptotagmin-2*, *synaptogyrin-1*, Table S1), as in mammalian enteroendocrine cells (75), potentially indicating an as-yet-uncharacterized role in neuroendocrine communication between intestinal cells and/or other planarian cell types.

Organelle and cellular physiology. Among regulators of digestion-related organelles, transcripts predicted to coordinate phagosome, endosome, and lysosome physiology were among the most highly represented. These included several *vacuolar protein sorting-associated protein (vps)* homologs required for lysosome tethering to late endosomes and autophagosomes (76), and *V-type proton ATPase 16 kDa proteolipid subunit (atp6v0c)*, responsible for lysosome acidification (as part of the V-ATPase proton pump) and also a component of the v-ATPase-Ragulator complex that senses nutrient sufficiency/deficiency (77) (Table S1A, Table S2A). Genes regulating additional aspects of cellular physiology were also expressed in the intestine, including homologs of the autophagy-related proteins *atg3* and *atg7*, ubiquitin ligases that are required in mouse for autophagosome formation during nutrient

starvation (78, 79) (Table S1A). Autophagosomes accumulate in the intestines of starved planarians (80), suggesting functional conservation of autophagy pathways. Along with *Smed-tor* (planarian mTOR), another intestine-enriched transcript (81) (Table S1A), Atg3 and Atg7 may contribute to planarians' ability to survive extended periods of fasting (82).

Cell shape and motility. As in our previous study of phagocyte expression (24), here we identified many intestine-enriched genes predicted to regulate cell shape, motility, and adhesion. Expression of this class of transcripts is consistent with the intestine's ability to remodel its branched morphology during regeneration and growth, and its role in intracellular digestion that likely requires extensive intracellular trafficking and robust maintenance of apicobasal polarity. Examples of these transcripts include numerous cytoskeletal regulators (e.g., multiple *rho-family* GTPases, *ena/VASP-like protein EVL*, *cofilin-1*, *supervillin*, and *advillin*), and regulators of cell adhesion and interactions with extracellular matrix (e.g., two *protocadherin-1*-like transcripts, *neuroglian/neural cell adhesion molecule L1-like*, two *laminin* subunits, and *nidogen-2* (Fig 2B)). Additionally, several transcripts may coordinate cell polarity and/or cell:cell junction formation, such as *lin7b*, whose orthologs stabilize mammalian cell polarity (83) and are required for cell junction targeting of the LET-23 EGF receptor during *C. elegans* vulva morphogenesis (84), and *partitioning defective 6 (pard6b)*, which we previously demonstrated was required for planarian intestinal remodeling (24) (Table S1A, Table S2A).

Stimuli response and defense. Transcripts predicted to regulate responses to stress, microorganisms, and other stimuli were also intestine enriched (Fig 2A). Prominent among these were regulators of innate immunity, including paralogs of *peptidoglycan recognition proteins* that function in immune signaling and/or have bactericidal activity (85, 86) (*prgp-le*, Fig 2B); an ortholog of *tollip*, a modulator of responses to inflammation by innate immune cells (87) (Table S1 and Table S2); and over 30 tumor necrosis factor receptor-associated factor homologs (TRAFs), a family of adaptor proteins that is expanded in *S. mediterranea* (88) and function as effectors of receptor signaling in innate and adaptive immunity (89) (Table S1 and

Table S2). *pgrp* and *traf* homologs have functional roles in planarian responses to infection with pathogenic *Pseudomonas* bacteria (90). Notably, sixty-two intestine-enriched transcripts (from this study) were previously found to be upregulated in response to shifting planarians from recirculating to static culture conditions, which causes microbiome dysbiosis (90) (Fig S3A, Table S3). Further supporting a role for the intestine in innate immunity and/or inflammatory responses, homologs of 99 *S. mediterranea* intestine-enriched transcripts were also upregulated by ingestion of pathogenic bacteria in *Dugesia japonica*, a related planarian species (91) (Fig S3B, Table S3).

Evolutionary conservation of human digestive organ gene expression

Further illustrating the conservation of physiological roles, we found that the intestine expresses numerous homologs of transcripts that are enriched in human GI tissues (Fig 3A-F and Table S4). To make this comparison, we conducted reciprocal best homology (RBH) searches (92, 93) to identify 5,583 planarian transcripts encoding predicted open reading frames (ORFs) with high homology to ORFs in UniProt human transcripts (94), and vice versa (Fig 3A and 3B). Of these, we then identified 5,561 transcripts (including 699 gut-enriched transcripts) with human transcripts represented in the Human Protein Atlas (Fig 3B and 3C), in which transcripts whose unique or highly enriched tissue-specific expression defines 32 human tissues or organs (95). Next, we calculated the percentage of all (5,561) and gut-enriched (699) RBH transcripts that were enriched in human tissues (Fig 3D), then expressed these percentages as a fold-enrichment ratio (planarian intestine vs. non-intestine) to estimate similarity between the planarian intestine transcriptome and the transcriptomes of human tissues (Fig 3E).

Strikingly, four of the five tissues to which the planarian intestine was most similar are involved in digestion (duodenum, small intestine, and gallbladder) or energy storage/metabolism (liver) (Fig. 3E and 3F). The intestine was also similar to other human digestive tissues

(esophagus, colon, and rectum), as well as kidney, possibly suggesting a role supporting the planarian protonephridial system in filtration of extracellular solutes (Fig 3E and 3F) (65, 96, 97). Duodenum- and small intestine-enriched RBH transcripts included *hepatocyte nuclear factor 4 (hnf-4)*, a transcription factor that regulates specification of endodermal tissues like intestine and liver (33), and that is expressed by a subset of planarian neoblasts likely to be intestinal precursors (32, 33, 36); the ileal (small intestine) sodium/bile transporter *slc10a2*; an ortholog of *2-acylglycerol O-acyltransferase 2 (mogat2)*, an intracellular acyltransferase involved in dietary fat absorption in the small intestine (98); *sodium/glucose transporter 2 (slc5a9)*, predicted to regulate glucose absorption; and a membrane-associated *phospholipase B1* homolog (*plb1*) predicted to regulate lipid absorption (Fig. 3F). RBH homologs enriched in liver included *urocanate hydratase (uroc1)* and *tryptophan 2,3-dioxygenase (tdo2)*, hepatic regulators of histidine and tryptophan catabolism, respectively (99, 100) (Fig 3F); and *fructose-1,6-bisphosphatase 1 (fbp1)*, a regulator of gluconeogenesis and a coordinator of appetite and adiposity (101, 102) (Table S4H). In addition, orthologs of proteins involved in drug and toxic/xenobiotic compound metabolism (e.g., *ATP-binding cassette subfamily C member 2 (abcc2)* and two *cytochrome P450* family transcripts) (Tables S1A, S2A, and S4H), and lipoprotein particle biogenesis (103) (*microsomal triglyceride transfer protein*, (Table S4H) were enriched in duodenum, small intestine, and/or liver. Finally, examples of transcripts with homologs enriched in human gallbladder/kidney included *sulfotransferase 1C4 (sult1c4)*, *gamma-glutamyltranspeptidase 1 (ggt1)*, and *uridine phosphorylase 2 (upp2)* (Fig 3F). These observations reinforce the evolutionary conservation of intestinal gene expression, and indicate that physiological roles performed by multiple human digestive organs are consolidated in the planarian intestine.

Identification of genes expressed by three distinct intestinal cell types and comparison to single-cell RNA-Seq data

Previously, we identified genes preferentially expressed by intestinal phagocytes (24). To distinguish transcripts expressed by phagocytes and other intestinal cell types, such as goblet cells (30, 74, 104), we directly compared $\log_2$ fold-change values for 1,317 transcripts represented in our sorted phagocyte data (24) as well as laser microdissected intestinal tissue (this study) (Fig 4A, C, and E, and Table S5). 900/1,317 transcripts were significantly upregulated in both phagocytes and laser-microdissected intestine (Fig 4A). We analyzed expression of these transcripts using WISH (Fig 4B, Table S1A, and Fig S2). As expected, these transcripts, which included previously identified intestinal markers *hnf-4* and *nkx2.2* (24, 33, 105), displayed uniform, ubiquitous expression throughout the intestine, consistent with enrichment in phagocytes (Fig 4B and Fig S2).

358 intestine-enriched transcripts were not significantly up- or down-regulated in phagocytes (Fig 4C, Fig S2, and Table S5). WISH analysis suggested these transcripts are enriched in multiple intestinal cell types (Fig 4D). Some transcripts in this group were expressed ubiquitously throughout the intestine (*ral guanine nucleotide dissociation stimulator-like 1 (rgl1)* and *family with sequence similarity 21 member C (fam21c)*, a homolog of *WASH complex subunit 2*, Fig 4D), suggesting expression in phagocytes, possibly in addition to other cell types. However, others were expressed in a distinct subset of less abundant intestinal cells (*peptidase inhibitor 16 (pi16)* and *serine peptidase inhibitor, Kunitz type 3 (spint3)*, Fig 4D). These transcripts are enriched in goblet cells, since their WISH expression pattern is highly similar to labeling of this subpopulation by lectins (106), antibodies (53, 64, 107), and other recently identified transcripts (32, 43, 64). A third set of transcripts was enriched in basal regions of the intestine (*zgc:172053*, a homolog of human C-type lectin *collectin-10*, and *calmodulin-3 (calm3)*, Fig 4D). This pattern resembles that of a planarian *gli*-family transcription factor (108) and several solute carrier-family transporters (65), and indicates expression by “outer intestinal cells”

(which we refer to as “basal cells” because of their proximity to the basal region of phagocytes) that were also recently identified in a large-scale, single-cell sequencing effort (32).

Finally, 59 intestine-enriched transcripts were significantly downregulated in phagocytes (Fig 4E, Fig S2, and Table S5). As expected, we never observed uniform, phagocyte-like expression patterns for these transcripts (Fig 4F and Fig S2). Rather, transcripts in this group were enriched only in goblet cells (e.g., *epididymal secretory protein E1/Niemann-Pick disease type C2 protein (npc2)*) or basal cells (e.g., *solute carrier family 22 member 6 (slc22a6)*). Furthermore, some goblet-cell-specific transcripts also appeared to be either medially (e.g., *metalloendopeptidase (cg7631)*) or laterally (e.g., *eppin*) enriched (Fig 4F and Fig S2), suggesting possible specialization of this cell type along the mediolateral axis.

Overall, validation by WISH identified 78 transcripts with a ubiquitous intestinal expression pattern; nearly all of these were upregulated in our sorted phagocyte data (Fig S2 and Fig S4A). By contrast, most of the 25 validated goblet-cell-enriched transcripts (Fig S4B) and 14 basal-cell-enriched transcripts (Fig S4C) were not upregulated in sorted phagocytes.

We further validated our results by comparing transcript enrichment in enterocytes/phagocytes, goblet cells, and outer intestinal cells/basal cells reported in a recent single-cell RNA-Seq (scRNA-Seq) analysis of planarian cells (32). Phagocyte- and goblet-specific transcripts mapped to similar quadrants in our phagocyte vs. laser-captured intestine plots (S4D and Fig S4E). Most outer intestinal cell-specific transcripts were downregulated in our phagocyte data set, but a few were upregulated (Fig S4F), suggesting that basal cells and phagocytes may share greater transcriptome similarity than phagocytes and goblet cells. A second scRNA-Seq study (43) also identified phagocyte- and goblet-cell-specific transcripts. Again, these mapped to similar quadrants in our phagocyte vs. laser-captured intestine plots (Fig S4G-H). Furthermore, we identified 809 intestine-enriched transcripts that single-cell studies did not find to be enriched in the intestine (or for some, any planarian cell type) (Fig S4I, Table S1A). Using WISH, we validated intestine enrichment for 23/28 of these mRNAs (Table

S1A), including some with a uniform/phagocyte-like expression pattern (e.g., *sodium-dependent serotonin transporter (SerT)* (Fig 2B); *sodium/calcium exchanger 1 (slc8a1)* (Fig S4I); and *PITH-domain containing protein CG6153 (cg6153)* (Fig S4I)), and also goblet-enriched transcripts (e.g., *carboxypeptidase A2 (cpa2)* (Fig 1F) and *kallikrein 13 (klk13)* (Fig S4I). We also note that over 1000 intestine-enriched transcripts in scRNA-Seq studies were not included in our LCM-generated transcriptome (Fig S4I). However, the vast majority (>80%, not shown) of these were enriched in multiple cell types (32, 43) (Table S1A), suggesting considerable expression in non-intestinal tissue, consistent with our data.

Taken together, our results confirm the existence of three major intestinal cell types that have unique, but overlapping, gene expression profiles (32, 43). These include phagocytes and goblet cells, as well as a third population of cells that resides in basal regions of intestinal branches. In addition, the detection of transcripts exclusively enriched in laser-captured intestine suggests that expression profiling of laser-captured bulk tissue is significantly more sensitive than current single-cell profiling approaches, and may be a preferable method for assessing tissue-specific gene expression when single-cell resolution is not required.

Multiple cell types and subtypes reside in the planarian intestine

To further characterize intestinal cell types and subtypes, we used fluorescence in situ hybridization (FISH) to investigate co-expression of intestine-enriched transcripts. First, we verified the existence of three distinct cell types, using highly expressed phagocyte, goblet, and basal-specific markers (Fig 5A-C). Expression of the most phagocyte-enriched transcript, *cathepsin La (ctsla)*, was ubiquitous throughout intestinal branches, but did not overlap with *npc2*, a goblet-cell-enriched mRNA (Fig 5A), or with *slc22a6*, a basally enriched transcript (Fig 5B). A second phagocyte-enriched transcript, *cathepsin L1 (ctsl.1)*, was co-expressed with *ctsla*, but not with *npc2* or *slc22a6* (Fig S5A). A third phagocyte-enriched mRNA, *apolipoprotein b-2 (apob-2)*, was expressed in more basal regions of *ctsla*+ phagocytes, but again, *apob-2*

expression was clearly distinct from *npc2*+ goblet cells and *slc22a6*+ basal cells (Fig S5A). Goblet cell-enriched *npc2* was expressed by cells with minimal *slc22a6* expression (Fig 5C), further reinforcing that *npc2*+ goblet cells are distinct from *ctsla*+ phagocytes as well as *slc22a6*+ basal cells. We also found that basal cells were distinct from numerous visceral muscle fibers that surround intestinal branches, occupying basal regions around digestive cells (109, 110), consistent with another study (28). Basal-specific *slc22a6* mRNA did not colocalize with either a muscle-specific mRNA (*troponin I 4 (tni-4)*, Fig 5D) (5), or with labeling by an antibody that recognizes a muscle-specific antigen (mAb 6G10, Fig 5E) (53). Thus, *slc22a6*+ cells represent a novel cell type in the intestine that is distinct from visceral muscles, phagocytes, and goblet cells, and that has, to our knowledge, not been described by numerous previous histological and ultrastructural studies. Our data independently confirm the identification of this cell type in a recent single-cell sequencing effort (32).

Using additional markers, we also discovered previously unappreciated heterogeneity in gene expression amongst both goblet and basal cells. For example, *metalloendopeptidase (cg7631)* was expressed by a subpopulation of goblet cells restricted to medial regions of the intestine, mainly localized to primary branches (Fig 6A and 6C). By contrast, *spint3*-expressing cells were found laterally, within secondary, tertiary, and quaternary branches (Fig 6B and 6C). Only rarely did cells in these medial and lateral domains co-mingle, or co-express both markers; such occurrences were restricted to the mediolateral boundaries between primary and secondary branches (Fig 6C). We also identified subpopulations of basal cells. For example, *zgc:172053 (collectin-10-like)*, see also Fig 4D) was expressed by some, but not all basal cells, especially in lateral regions of the intestine (Fig 6D). Similarly, *si:dkey-241* (another C-type lectin) was enriched in a subset of laterally enriched basal cells, but not phagocytes or goblet cells (Fig S5B). These data reveal subpopulations of goblet and basal cells in distinct intestinal domains, suggesting potentially specialized roles.

We also identified transcripts expressed by multiple cell types. For example, *fatty acid binding protein 7 (fabp7)* mRNA was expressed by both basal cells (Fig 6E) and phagocytes (Fig 6F), but negligibly in goblet cells (Fig 6G). By contrast, *lysosomal alpha-glucosidase (gaa)* was expressed at high levels in goblet cells, at low levels in phagocytes, and at negligible levels in basal cells (Fig S5C). *oysgedart (oys)*, a lysophospholipid acyltransferase, was enriched in phagocytes, and at low levels in basal cells, but not in goblet cells (Fig S5D). These results, together with bioinformatic comparisons (Fig 4 and Fig S4), demonstrate that unique digestive cells co-express many transcripts in different combinations and levels, illustrating the complexity of gene expression even in a tissue with relatively few cell types.

Intestine-enriched transcription factors regulate goblet cell differentiation and maintenance

Definition of the intestinal transcriptome enables identification of genes required for regeneration and functions of distinct intestinal cell types. Here, to initiate this effort, we focused on goblet cells, which expressed the majority of medially and laterally enriched transcripts we identified (Table S1A, Fig S2). Planarian goblet cells (44, 111) (also called “Minotian gland cells” (112), “granular club” cells (30), “sphere cells” (31), or “pyriform cells” (113)) have been proposed to secrete digestive enzymes into the intestinal lumen (104), or store protein reserves (29). Ultrastructurally, planarian goblet cells possess numerous large proteinaceous granules and abundant rough endoplasmic reticulum (30, 104, 113), resembling mammalian goblet cells that produce a protective mucous barrier and mount innate immune responses (114-118). However, although numerous markers and reagents have been identified that label planarian goblet cells (32, 43, 53, 64, 106, 107), to our knowledge, genes required for goblet cell differentiation, maintenance, or physiological roles have not been reported.

Reasoning that transcription factors (TFs) would regulate goblet cell generation and/or maintenance, we focused on transcripts encoding 22 intestine-enriched TFs, only 10 of which

were previously known to be enriched in intestinal cells (Table S6A). Using FISH, we validated expression of all but one of these TFs in the intestine (Fig S6). In a dsRNA-mediated RNA interference screen to specifically assess goblet cell regeneration, we found that knockdown of three TFs dramatically reduced expression of a goblet cell marker in regenerating head, trunk, and tail fragments (Table S6A and Fig 7). Knockdown of *mediator of RNA polymerase II transcription subunit 21 (med 21)* reduced goblet cells and blastema formation (Table S6A), but also caused severe disruption of intestinal integrity in our previous study (24). This suggested a non-goblet-cell-specific role, and we did not investigate *med21* further. Knockdown of a second transcription factor, *gli-1* (a transducer of hedgehog signaling (108, 119)), caused failure of goblet cells to regenerate at the midline of the new anterior branch in amputated tail fragments regenerating a new head (Table S6A; Fig 7A and 7B; Fig S7A). Regeneration of goblet cells was also reduced in new tail branches of *gli-1(RNAi)* head fragments (Fig S7B) and in anterior and posterior branches in *gli-1(RNAi)* trunk fragments (Fig S7C). These effects on goblet cells were specific, since phagocytes (Fig 7A, Fig S7A-C) and basal cells (Fig 7B) regenerated normally. This was true even in some fragments with smaller posterior blastemas characteristic of reduced hedgehog signaling (Fig S7C) (108, 119, 120).

By contrast to the *gli-1* phenotype, goblet cells arose normally in regenerating intestine upon knockdown of a third TF, *ras-responsive element binding protein 2 (RREB2)*, including anterior intestinal branches in tail fragments (Fig 7A and 7B; Fig S7A), posterior branches in head fragments (Fig S7B), and both anterior and posterior branches in trunk fragments (Fig S7C). However, in pre-existing intestinal regions, goblet cell numbers were dramatically reduced. These included the posterior of tail fragments (Fig 7A and 7B; Fig S7A), the anterior of head fragments (Fig S7B), and central regions of trunk fragments (Fig S7C). As with *gli-1*, phagocytes and basal cells were unaffected (Fig 7A and 7B; Fig S7A-C). Thus, the *RREB2* phenotype represents a striking "mirror image" of the phenotypes in *gli-1(RNAi)* animals. Together, these results suggest that *gli-1* regulates neoblast fate specification and/or

differentiation of neoblast progeny into goblet cells, while *RREB2* may control maintenance or survival of goblet cells after they initially differentiate.

In uninjured animals, both *gli-1* and *RREB2* also reduced goblet cell numbers. Knockdown of *gli-1* reduced goblet cell numbers in uninjured animals, especially in anterior, posterior, and lateral intestine branches (Fig 7C), but many goblet cells remained in medial, primary intestinal branches. In *RREB2(RNAi)* animals, goblet cells were more dramatically reduced in all regions of the intestine (Fig 7C). Although these phenotypes were broadly consistent with observations in regenerates, they also suggested that *gli-1* and *RREB2* might be required for differentiation and/or maintenance of distinct mediolateral goblet cell subpopulations in uninjured animals. Indeed, we found (using additional goblet cell markers) that almost no lateral goblet cells remained in *gli-1(RNAi)* uninjured animals, while again, medial goblet cells were much less affected (Fig S8A). By contrast, in *RREB2(RNAi)* uninjured animals, reduction of both medial and lateral goblet cell numbers was pronounced, but nonetheless some lateral goblet cells remained (Fig S8A). These differential effects on mediolateral subpopulations were not observed in regenerates. For example, both subpopulations failed to differentiate in the primary anterior intestinal branch in *gli-1(RNAi)* tail regenerates (Fig S8B), and both subpopulations were severely reduced in pre-existing, posterior branches of *RREB2(RNAi)* tail regenerates (Fig S8B).

Taken together, these results suggest that, in uninjured animals undergoing normal homeostatic growth and renewal, *gli-1* is primarily required for differentiation of new lateral intestinal cells. Alternatively, as goblet cells fail to renew, remaining goblet cells might somehow migrate to more medial intestinal branches, or survival of goblet cells in primary branches could be prolonged. Conversely, *RREB2* seems to be required for maintenance of both medial and lateral goblet cells, although lateral cell numbers are slightly less affected by *RREB2* knockdown, compared to *gli-1*. Finally, in regenerates, *gli-1* and *RREB2* knockdown affects medial and lateral goblet cells similarly (but in regenerating vs. pre-existing intestine). These

complex observations suggest that mechanisms regulating goblet cell differentiation and maintenance may differ during homeostasis and regeneration, or that compensatory mechanisms capable of sustaining goblet cell production/survival in uninjured *gli-1(RNAi)* and *RREB2(RNAi)* animals are insufficient to meet the increased demand for new tissue during regeneration.

Despite their specific effects on goblet cells, neither *gli-1* nor *RREB2* mRNAs are specifically enriched in this cell type. In fact, *gli-1*, which was previously shown to be expressed by intestine-associated cells (108, 121), was most highly enriched in phagocytes, basal cells, and intestine-associated muscle cells (visceral muscle), but was also expressed at lower levels in some goblet cells (Fig S9A). *RREB2* was enriched in goblet cells and basal cells, but expression was also observed in some phagocytes and intestine-associate muscle (Fig S9B). Intriguingly, we found that knockdown of two other TFs, *48 related 1 (fer1)* (also called *pancreas transcription factor 1 subunit alpha* or *PTF1A* (32)), and *LIM homeobox 2 (lhx2b)*, also modestly reduced goblet cells in pre-existing branches (Table S6A, Fig S10A-C). mRNAs encoding these TFs are enriched in basal regions of the intestine (Fig S6). In addition, single-cell transcriptome data suggest *PTF1A* expression is elevated in differentiating neoblast progeny in the basal cell lineage (32), and *PTF1A* also reduces basal intestinal cell numbers (32). Thus, although our data support a role for *gli-1* and *RREB2* in goblet cells and/or their precursors, they also raise the possibility that basal cells (and possibly phagocytes or muscle cells) may non-autonomously influence goblet cell differentiation and/or survival.

Goblet cell reduction compromises feeding behavior and viability

We asked whether goblet cell depletion affected planarian viability, behavior, or regeneration. Over a 6-week dsRNA feeding regimen, some uninjured *gli-1(RNAi)* and *RREB2(RNAi)* planarians refused food and failed to eat after 4-5 weeks (Fig 7D). In both *gli-1(RNAi)* and *RREB2(RNAi)* animals, some animals eventually lysed, curled, or died (Fig S11A),

suggesting that goblet cells are required for viability. Additionally, we also observed a modest decrease in animal size (Fig S11B-C) in *RREB2(RNAi)* planarians relative to controls. Interestingly, we only observed feeding failure and other phenotypes in animals fed the dsRNA/liver mix 1x/week (every 7 days) for 6 weeks; animals that were fed 2x/week (every 3-4 days), but still 6 times, did not refuse to eat, lyse, curl, or die (not shown). This suggests that goblet cells may regulate hunger (or other aspects of digestive physiology) primarily in starved animals. Next, to assess regeneration, we amputated planarians fed 2x/week for 8 feedings, in order to eliminate the influence of feeding failure on possible regeneration phenotypes. In both *gli-1(RNAi)* and *RREB2(RNAi)* regenerates, goblet cell depletion was robust (Fig 7A and 7B), but neither gene was required more generally for regeneration (with the infrequent exception of reduced posterior blastemas in *gli-1(RNAi)* regenerates, mentioned above), as the brain (Fig 7E), pharynx (Fig 7E), other intestinal cell types (Fig 7A and 7B), and new intestinal branches (Fig 7A and 7B) all regenerated without noticeable defects. Together, these results suggest that goblet cells are dispensable for regeneration, and that their primary role may be to regulate appetite or other aspects of intestinal physiology that are critical for viability.

Goblet cells likely have multiple functions

The intestinal transcriptome (Table S1A) includes numerous goblet-cell-enriched transcripts that provide insights into the potential physiological roles of goblet cells (Fig S12A and S12B). First, in support of previous suggestions that goblet cells promote luminal digestion (111, 112, 122), we identified several goblet-enriched transcripts predicted to encode secreted regulators of protein catabolism (*pancreatic carboxypeptidase A2*) and triglyceride catabolism (*lipase F/gastric triacylglycerol lipase*) (Fig S12A). Second, we also identified three goblet-enriched *kallikreins* (*klk*) (Fig S12A and Fig S2), secreted proteases whose mammalian homologs produce vasoactive plasma kinin, but also hydrolyze extracellular matrix molecules, growth factors, hormone proteins, and antimicrobial peptides in numerous tissues (123).

Kallikreins are expressed by goblet cells in rat, cat, and mouse intestines (124, 125), and have been implicated in inflammatory bowel disease and gastrointestinal cancers (126, 127), suggesting additional conservation of planarian goblet cell physiology. Third, goblet cells express *peptidoglycan recognition protein (pgrp-1b)* (Fig S12A), whose vertebrate and invertebrate homologs modulate innate immune signaling and play direct bactericidal roles (85, 86), suggesting goblet cells may coordinate immune responses and/or regulate microbiome composition. Consistent with this idea, a second planarian paralog, *pgrp-1e* (Fig 2B and Fig S2), is upregulated in the intestine (but not restricted to goblet cells) in response to *Pseudomonas* infection (90). Fourth, a homolog of *prohormone convertase (pcsk1)* is enriched in goblet cells, as well as peripharyngeal cells surrounding the pharynx (Fig S12B). Prohormone convertases (PCs) cleave neuropeptide and peptide hormone preproteins to generate bioactive peptides. In both vertebrates and invertebrates, PCs function in neurons, but also in endocrine cell types such as pancreatic islet cells and digestive tract enteroendocrine cells, where they regulate glucose levels, energy homeostasis, and appetite (125, 128, 129). Although another planarian *pcsk* paralog, *Smed-pc2*, processes neuropeptides required for germline development (130), to our knowledge, no intestine-enriched prohormones have been reported. Nonetheless, *pcsk1* expression suggests goblet cells may have an enteroendocrine-like role, consistent with our observation (Fig 7D) that goblet cell reduction perturbs feeding behavior. Finally, because mucin production is a defining feature of goblet cells in vertebrates (117) we also identified three planarian homologs of genes encoding gel-forming mucins, including MUC2, an abundant intestinal mucin (Fig S12C) (131, 132). Somewhat surprisingly, these were expressed by peripharyngeal cells that send projections into the pharynx (52) (Fig S12C), but not goblet cells. Thus, although mucin proteins may be delivered to the intestinal lumen through the pharynx, negligible goblet cell expression of *mucin* transcripts is a notable difference between planarians and vertebrates.

Intriguingly, many transcripts with the greatest medial or lateral enrichment (e.g., >2X in medial vs. lateral tissue or vice versa) in the intestine were expressed by goblet cells (Table S6B-C, Fig 4C-F, Fig S4B, Fig S2), suggesting possible functional specialization. Such transcripts included the digestive enzymes, proteases, and pattern recognition receptor just described, as well as transcripts encoding predicted extracellular matrix-associated proteins (e.g., *nidogen-2* and *eppin*) (Fig 2B, Fig 4F, Fig S2, and Table S6B-C). To attempt to determine the functional significance of mediolateral goblet cell subpopulations, we used RNAi to assess the roles of 16 of the most medial and 8 of the most lateral goblet-enriched transcripts (Table S6B-C). However, we did not observe failure to feed, decreased viability, or defects in blastema formation, even after 8 weeks of knockdown (feeding 1x/week) (Table S6B-C). This might be due to functional redundancy, since multiple lipases, carboxypeptidases, and kallikriens are expressed by goblet cells. Alternatively, other intestinal cell types might play overlapping roles with respect to some functions. For example, analysis of biological process GO term enrichment among all 1221 medial and 623 lateral transcripts – without regard to cell type specificity – revealed numerous putative regulators of innate immunity and macromolecular catabolism among medial transcripts, and of extracellular matrix organization among lateral transcripts (Fig S13A-C). Lastly, some goblet cell functions may be required only when planarians are challenged by stresses like bacterial infection or extended starvation, possibilities that will require further investigation.

Discussion

We have developed methods for applying laser microdissection to planarian tissue, which we used to define gene expression in the intestine. The intestine expresses genes involved in metabolism, nutrient storage and transport, innate immunity, and other physiological roles, demonstrating considerable functional homology with digestive systems of other animals, including humans. Comparison of gene expression in microdissected tissue to that of intestinal

phagocytes (previously isolated by sorting) enabled identification of transcripts enriched in two other cell types, goblet cells and basal cells. We also discovered previously unappreciated intestinal cell-type diversity, especially amongst goblet cells, which reside in distinct medial and lateral intestinal domains. Identification of medially and laterally enriched transcripts suggests an additional paradigm for addressing axial influences on organ regeneration, as well as evolution of patterning mechanisms influencing the regionalization of bilaterian digestive systems (133, 134). Finally, we identified intestine-enriched transcription factors that play distinct roles in goblet cell differentiation and maintenance, and found that depletion of goblet cells reduces planarians' willingness to feed and viability, with only negligible effects on regeneration of other intestinal cell types or non-intestinal tissues.

Characterization of the planarian intestinal transcriptome provides a framework for several next steps. First, a detailed description of intestinal cell types will facilitate further studies of digestive cell types. Here, we identified two transcription factors, *gli-1* and *RREB2*, whose RNAi phenotypes suggest that distinct mechanisms govern differentiation and maintenance of goblet cells. We are unaware of studies in other organisms that have uncovered direct roles for either *gli-1* or *RREB2* in goblet cells. However, in mice, modulation of Sonic Hedgehog (an upstream activator of Gli-family TFs) levels affects goblet cell numbers (135, 136), and RREB1 cooperatively regulates expression of the peptide hormone secretin in enteroendocrine cells (137). By contrast, in the *Drosophila* midgut, Hedgehog (Hh) signaling promotes intestinal stem cell proliferation (138), while the RREB-1 ortholog *hindsight/pebbled* suppresses midgut intestinal stem cell proliferation and is required for differentiation of absorptive enterocytes (139). In planarians, hedgehog (Hh) regulates anteroposterior polarity, neoblast proliferation, and neurogenesis (108, 120, 121, 140); our data suggest a possible additional role for Hh in intestinal morphogenesis. Similarly, although a second planarian RREB paralog, *RREBP1*, is partially required for eye regeneration and may promote differentiation (141), implication of *RREB2* in goblet cell maintenance/survival suggests that Ras-mediated

signaling could also influence cellular composition in the intestine (although direct links to Ras remain to be established). Thus, although characterization of the precise roles of *gli-1* and *RREB2* will be required, our results suggest broad similarity between planarians and other animals with respect to evolutionarily conserved signaling pathways that govern cell dynamics in the digestive tract. Furthermore, although we focused on goblet cells, previous studies suggest that several genes (*gata4/5/6-1*, *hnf4*, *egfr-1*, *PTF1A*) expressed by cycling neoblast subpopulations (23, 33, 34, 36) may be required for differentiation of multiple intestinal cell types (23, 32, 37, 142). Thus, enumeration of intestinal cell types and subtypes here, along with endoderm-specific progenitors in several recent single-cell analyses (32, 36, 43), provides a rich list of candidate regulators for further elucidation of differentiation, maintenance, and functions of planarian digestive cells and their progenitors.

Second, functional analysis of intestine-enriched transcripts will help to resolve the intestine's role in regulation of the stem cell microenvironment. Knockdowns of the intestine-enriched TFs *nkx2.2* or *gata4/5/6-1* result in reduced proliferation and blastema production (19, 37). In addition, a subset of *tetraspanin group-specific gene-1* (*tgs-1*)-positive neoblasts likely to be pluripotent neoblasts resides near intestinal branches (36), further hinting at a niche-like role for the intestine. Numerous intestine-enriched transcripts encode regulators of metabolite processing and transport, as well as putatively secreted proteins, suggesting multiple possible cell non-autonomous influences on neoblast dynamics. Because gut-enriched TFs like *nkx2.2* and *gata4/5/6-1* are also expressed by neoblasts and other cell types (32), LCM also provides an efficient approach for clarifying which candidate stem cell regulators are dependent on these or other TFs for their intestine-specific expression.

Third, LCM provides a complementary method for assessing injury-induced gene expression changes in the intestine and other planarian tissues that may overcome the shortcomings of other available methods. Previously, we developed a method for purification of intestinal phagocytes from planarians that ingested magnetic beads (24, 52). Laser

microdissection may be preferable, because it enables detection of transcripts expressed by other intestinal cell types (not just phagocytes), and also eliminates potentially confounding gene expression changes caused by feeding. Similarly, although single-cell sequencing (SCS) approaches in planarians have driven significant advances in our understanding of planarian cell types and their responses to injury (32, 34, 36, 43, 143, 144), LCM may provide a more direct and efficient way to assess gene regulation in tissues with rarer cell types. For example, because intestinal cells comprise only 1-3% of total planarian cells (44), without prior enrichment SCS would potentially require analysis of tens of thousands of cells to reliably detect gene expression changes in the intestine. In addition, although chemistry and computational approaches for SCS are improving rapidly (145, 146), LCM may enable more sensitive detection of low-copy transcripts or subtle fold changes in bulk tissue, and also bypass “noise” caused by dissociation or other SCS-related technical artefacts.

Regeneration of digestive organs is not well understood. Development of a robust method for isolating intestinal tissue, and characterization of the intestinal transcriptome, will facilitate mechanistic studies in planarians. The methods and resources presented here will also support comparative analyses, complementing current and future efforts to understand digestive tract regeneration in platyhelminths, sea cucumbers, annelids, ascidians, amphibians, and mammals (147-156). In addition, studies in planarians and other regeneration models are likely to generate new insights into cellular processes (e.g., proliferation, differentiation, metabolism, and stress responses) whose dysregulation underlies human gastrointestinal pathologies associated with aging and disease.

Materials and Methods

Ethics Statement

No vertebrate organisms were used in this study.

Planarian maintenance and care

Asexual *Schmidtea mediterranea* (clonal line CIW4) (157) were maintained in 0.5 g/L Instant Ocean salts with 0.0167 g/L sodium bicarbonate dissolved in Type I water (158), and fed with beef liver paste. For all experiments, planarians were starved seven days prior to fixation. For LCM, planarians were 6-9 mm in length. For WISH and FISH, planarians were 2-4 mm in length. All animals were randomly selected from large (300-500 animals) pools, with the exception that animals with blastemas (e.g., those that had recently fissioned) were excluded.

Optimization of planarian fixation for RNA extraction and histology

Planarians were relaxed in 0.66 M MgCl₂ or treated with 7.5% N-acetyl-L-cysteine or 2% HCl (ice-cold) for 1 minute to remove mucus as described (52). Planarians were fixed in 4% formaldehyde/1X PBS, or methacarn (6 ml methanol, 3 ml chloroform, 1 ml glacial acetic acid) for 10 min at room temperature as described (52). Formaldehyde-fixed samples were washed 3 times (5 minutes each) in 1X PBS. Methacarn-fixed samples were first rinsed 3 times in methanol, then rehydrated in 1:1 methanol:PBS for 5 minutes, followed by 3 washes (5 minutes each) in PBS. For analysis of RNA integrity, 5-10 planarians were immediately homogenized in Trizol, and RNA was extracted using two chloroform extractions and high-salt precipitation buffer according to the manufacturer's instructions. RNA samples were analyzed using Agilent RNA ScreenTape on an Agilent TapeStation 2200 according to the manufacturer's protocol.

For histology on methacarn-fixed samples, animals were relaxed and fixed individually in glass vials to minimize adherence to other samples. After fixation and rehydration, animals were incubated in 5%, 15%, and 30% sucrose (in RNase-free 1X PBS), for 5-10 min each. Samples were then mounted and frozen in OCT medium, and cryosectioned at 20 µm thickness onto either Superfrost Plus glass slides (Fisher 12-550-15) (for staining optimization) or PEN membrane slides (Fisher/Leica No. 11505158) (for LCM). Prior to cryosectioning, PEN membrane slides were treated for 1 min with RNaseZAP (Invitrogen AM9780), then rinsed by dipping 10 times (1-2 seconds per dip) in three successive conical tubes filled with 30 ml DEPC-treated water, followed by 10 dips in 95% ethanol and air-drying for 5-10 min. After cryosectioning, slides were stored on dry ice for 2-4 hours prior to staining. Cryostat stage and blades were wiped with 100% ethanol prior to sectioning.

For histological staining, slides were warmed to room temp. for 5-10 min. All slides were stained individually in RNase-free conical tubes by manually dipping (1-2 seconds per dip) in 30 ml solutions. Forceps used for dipping were treated with RNaseZAP and ethanol. All ethanol solutions were made with 200 proof ethanol and DEPC-treated water.

For Hematoxylin staining: 70% ethanol (20 dips); DEPC-treated water (20 dips); Mayer's Hematoxylin (Sigma Aldrich MHS16-500ML) (15 dips); DEPC-treated water (10 dips); Scott's Tap Water (2 g sodium bicarbonate plus 10 g anhydrous MgSO₄ per liter of nuclease-free water) (10 dips); 70% ethanol (10 dips); 95% ethanol (10 dips); 95% ethanol (10 dips); 100% ethanol.

For Eosin Y staining: 70% ethanol (20 dips); DEPC-treated water (20 dips); 70% ethanol (20 dips); Alcoholic Eosin Y (Sigma Aldrich HT110116-500ML) (100%, 10%, or 2% diluted into 200 proof ethanol) (15 dips); 95% ethanol (10 dips); 95% ethanol (10 dips); 100% ethanol. In some cases fewer dips (8-10) in Eosin Y were required for better differentiation of gut tissue.

For combined Hematoxylin and Eosin Y staining: 70% ethanol (20 dips); DEPC-treated water (20 dips); Mayer's Hematoxylin (15 dips); DEPC-treated water (10 dips); Scott's Tap

Water (10 dips); 70% ethanol (10 dips); 10% Alcoholic Eosin Y (15 dips); 95% ethanol (10 dips); 95% ethanol (10 dips); 100% ethanol.

The entire staining protocol was completed in less than 5 minutes. Slides were air dried for 5 minutes, then stored in plastic slide boxes on dry ice for 2-4 hr before LCM. Although we tested overnight storage at -80°C, we found that section morphology and RNA quality were best when conducting all steps, from fixation to LCM, on the same day.

Laser-capture microdissection and RNA extraction

Stained PEN slides were removed from dry ice and immediately immersed for 30 seconds in ice-cold 100% ethanol, then room temperature 100% ethanol to minimize condensation/rehydration of sections and maintain RNase inactivation during warming. Slides were then air dried for 2-3 minutes, and mounted in a Leica LMD7 laser microdissection microscope. Samples were dissected at 10X magnification using the following parameters: Power-30; Aperture-20; Speed-5; Specimen Balance-1; Head Current-100%; Pulse Frequency-392 Hz. Regions were dissected into empty RNase-free 0.5 ml microcentrifuge caps (Axygen PCR-05-C). We separately captured medial intestine, lateral intestine, and non-intestine regions from all sections (8-10) per slide within 45-50 min. After capture, 20 µl Buffer XB (Arcturus PicoPure RNA Isolation Kit 12204-1) was added to captured tissue, then tubes were immediately frozen on dry ice and stored at -80°C prior to RNA extraction.

For RNA extraction, samples were thawed for 5 min at room temp. Next, tissue from two tubes/slides (16-20 sections from the same planarian) was pooled for each biological replicate, incubated at 42°C for 30 minutes, and RNA was extracted using the Arcturus PicoPure RNA Isolation Kit following the manufacturer's instructions. 40-400 ng of total RNA was obtained from each sample, measured using a Denovix UV Spectrophotometer. RNA quality was analyzed

using Agilent HS RNA ScreenTape on an Agilent TapeStation 2200 according to the manufacturer's protocol.

Library preparation and RNA sequencing

Concentration of RNA was ascertained using a Thermo Fisher Qubit fluorometer. RNA quality was verified using the Agilent Tapestation. Libraries were generated using the Lexogen Quantseq 3' mRNA Library Prep Kit according to the manufacturer's protocol, with 5 ng total RNA input for each sample. Briefly, first-strand cDNA was generated using 5'-tagged poly-T oligomer primers. Following RNase digestion, second strand cDNA was generated using 5'-tagged random primers. A subsequent PCR step with additional primers added the complete adapter sequence to the initial 5' tags, added unique indices for demultiplexing of samples, and amplified the library. Final libraries for each sample were assayed on the Agilent Tapestation for appropriate size and quantity. Libraries were then pooled in equimolar amounts as ascertained by fluorometric analysis. Final pools were quantified using qPCR on a Roche LightCycler 480 instrument with Kapa Biosystems Illumina Library Quantification reagents. Sequencing was performed using custom primers on an Illumina Nextseq 500 instrument with High Output chemistry and 75 bp single-ended reads.

Short-read mapping and gene-expression analysis

Adapters and low quality reads were trimmed from fastq sequence files with BBduk (<https://sourceforge.net/projects/bbmap/>) using Lexogen data analysis recommendations (<https://www.lexogen.com/quantseq-data-analysis/>): k=13 ktrim=r forcetrimleft=11 useshortkmers=t mink=5 qtrim=t trimq=10 minlength=20. Sequence quality was assessed before and after trimming using FastQC (159). Reads were then mapped to a version of the *de novo* dd_Smed_v6 transcriptome (11) restricted to 28,069 unique transcripts (i.e., those transcripts whose identifiers ended with the suffix "_1") using Bowtie2 (v2.3.1) (160) with default

settings. Resulting SAM files were converted to BAM files, sorted, and indexed using Samtools (v1.3) (161). Raw read counts per transcript were then generated for each BAM file using the “idxstats” command in Samtools, and consolidated into a single Excel spreadsheet.

The resulting read counts matrix was imported into R, then analyzed in edgeR v3.8.6 (162). First, all transcripts with counts per million (CPM) < 1 in 4/12 samples (e.g., lowly expressed transcripts) were excluded from further analysis (13,136 / 28,069 transcripts were retained). Next, after recalculation of library size, samples were normalized using trimmed mean of M-values (TMM) method, followed by calculation of common, trended, and tagwise dispersions. Finally, differentially expressed transcripts in intestinal vs. non-intestinal samples were determined using the generalized linear model (GLM) likelihood ratio test. 1911 transcripts had a fold-change of more than 2 ($\log_{2}FC > 1$) and an FDR-adjusted p value < .01 in either medial or lateral intestine, relative to non-intestinal tissue. We further limited to 1844 transcripts with a minimum transcripts-per-million (TPM) of 2 in 4 of any 8 intestinal biological replicates (medial or lateral), since transcripts with lower expression values were at the lower limit of detection by ISH, and their removal also modestly increased the robustness of LCM vs. phagocyte analysis.

Human Protein Atlas comparison

We queried (TBLASTX) 28,069 unique dd_Smed_v6 nucleotide sequences against 20,726 nucleotide sequences in the human reference proteome downloaded from UniProt (www.uniprot.org) (Release 2017_12, 20-Dec-2017). 13,362 dd_Smed_v6 transcripts hit human sequences (7,309 unique) with E -value $\leq 1 \times 10^{-3}$. These human sequences were then used to conduct reciprocal TBLASTX queries against dd_Smed_v6 transcripts: 7,220 hit 5,808 unique dd_Smed_v6 sequences with E -value $\leq 1 \times 10^{-3}$. In total, 5,583 dd_Smed_v6 transcripts had reciprocal best hits (RBHs) in the human proteome, with >94% having E -value $\leq 1 \times 10^{-10}$ in either direction. Of 1844 intestine-enriched transcripts, 701 had RBHs in the human proteome.

Next, using RBH UniProt Accession numbers, we extracted RNA-Seq tissue enrichment data from the Human Protein Atlas (95). 699/701 intestinal transcripts' RBH homologs were present in the HPA data (Table S4); of these, 130 were enriched in one or more human tissues. 5,561/5,583 dd_Smed_v6 transcripts' RBH homologs were present in the HPA data; of these, 1,011 were enriched in one or more human tissues. The number of intestine and non-intestine RBH homologs enriched in each of 32 human tissues was calculated, and expressed as a percentage of total transcripts (699 intestine or 5,561 non-intestine) with RBH homologs. Finally, ratio of the percentage of intestine-enriched RBH homologs to the percentage of all dd_Smed_v6 RBH homologs was then calculated to determine fold-enrichment for each tissue. Two tissues ("Appendix" and "Smooth Muscle") were excluded from final analysis since <0.1% (6/5561) of all dd_Smed_v6 transcripts had RBH homologs enriched in these tissues.

TBLASTX queries were conducted using NCBI BLAST+ standalone suite. Extraction and analysis of tissue enrichment data from HPA was conducted in R and Excel.

Gene Ontology annotation, nomenclature, and analysis

BLASTX homology searches of UniProtKB protein sequences for *H. sapiens*, *M. musculus*, *D. rerio*, *D. melanogaster*, and *C. elegans* were conducted using all 28,069 unique transcripts in the "dd_Smed_v6" transcriptome in PlanMine as queries. In all figures, gene names/abbreviations are based on the best (lowest *E*-value) UniProt homolog, except: (1) genes were named after the best human homolog for HPA analysis (Figure 3); and (2) we used *Smed* nomenclature when genes (or paralogs) were previously named by us or others, including *hnf-4*/dd_1694_0_1 (33), *nkx2.2*/dd_2716_0_1 (24), *gata4/5/6*/dd_4075_0_1 (33), *apob-1*/dd_636_0_1 [DJF, unpublished], *apob-2*/dd_194_0_1 [DJF, unpublished], *slc22a6*/dd_1159_0_1 (65), *gli-1*/dd_7470_0_1 (108), and *RREB2*/dd_10103_0_1 (141). Biological Process GO terms (also obtained from UniProtKB) from the top hit for each species (with *E*-value $\leq 1 \times 10^{-5}$) were assigned to each dd_Smed_v6 transcript. 9344 of 28,069 total

transcripts and 1379 of 1844 intestine-enriched transcripts were annotated with GO terms. Enrichment for specific terms among all intestine-enriched transcripts, medially enriched transcripts, or laterally enriched transcripts was then evaluated with BiNGO (163). Regional transcript enrichment was calculated as a ratio of fold changes in medial and lateral intestinal tissue: transcripts were considered to be medially enriched if $FC_{\text{medial}}/FC_{\text{lateral}} > 1$ (1221 transcripts), or laterally enriched if $FC_{\text{medial}}/FC_{\text{lateral}} < 1$ (623 transcripts). All 13,136 (of 28,069) transcripts detected in our experiment (above) were used as a background set, and hypergeometric testing with a Benjamini & Hochberg False Discovery Rate (FDR) of .05 was considered significant. We additionally restricted our summarization to terms that were annotated to more than one percent of transcripts that received annotations in each group: 1379/1844 intestine-enriched transcripts, 924/1221 medially enriched transcripts, or 455/623 laterally enriched transcripts.

Comparison to gene sets involved in innate immunity

1456 *Dugesia japonica* transcripts upregulated in response to either *L. pneumophila* or *S. aureus* infection (91) were queried (TBLASTX) against 28,069 dd_smed_v6_unique transcripts. 981 dd_smed_v6 transcripts hit with evalules < 1e-03. After removing duplicate hits, 783 transcripts remained. 701/783 transcripts were present in the full 13,136 LCM dataset; 99/783 were intestine-enriched.

30,021 SMED_20140614 transcripts were mapped to 28,069 dd_smed_v6_unique transcripts, generating 22,889 dd_smed_v6_unique hits with evalules < 1e-03. After removing duplicates, 19,338 transcripts remained. 741/19,338 transcripts were up- or down-regulated in planarians shifted to static culture (90). 447/741 transcripts were present in the full 13,136 LCM dataset, and 62 were intestine-enriched.

Comparison to gene expression in sorted phagocytes

28,069 unique dd_Smed_v6 transcripts were blasted (BLASTN) against 11,589 ESTs and assembled contigs from the "SmedESTs3" collection (164), which were used in microarray-based analysis of gene expression in sorted phagocytes (24). 7927 hit with length >100 bp, >90% base identity, and *E*-value less than 1×10^{-50} . Of these, 6919 of 13,136 Smed_v6 transcripts detected in LCM samples mapped to unique Smed_ESTs3 contigs or ESTs with detectable expression in the sorted phagocyte data set. 2626/6919 transcripts had FDR-adjusted *p* values <.01 for logFC in either medial or lateral intestine (vs. non-intestine, LCM data in this study), and 1498/2626 had FDR-adjusted *p* values <.05 for logFC in sorted phagocytes (vs. all other cells) (phagocyte data in (24)). 1317/2626 had logFC >1 in either medial or lateral intestine. In the sorted phagocyte data, 900/1317 had a fold-change >0 and FDR-adjusted *p* value <.05 ("high" in phagocytes), 358/1317 had an FDR-adjusted *p* value >.05 ("moderate" in phagocytes), and 59/1317 had a fold-change <0 and FDR-adjusted *p* value >.05 ("low" in phagocytes).

Comparison to gene expression in single-cell transcriptomes

We utilized data from gene expression analysis of single planarian cells from two recent studies (32, 43) to identify dd_Smed_v6 transcripts enriched in specific intestinal cell types that were also represented in our 1844 intestine-enriched transcripts (this study) and phagocyte expression data (24). For comparison to Fincher et al. (32) (Fig S4D-F), we identified 391 phagocyte-enriched transcripts (subcluster 4 or "enterocytes" in Table S2 (intestine) (32)), 21 goblet-enriched transcripts (subcluster 8 in (32)), and 30 basal-enriched transcripts (subcluster 8 or "outer intestinal cells" in (32)); transcripts found in other intestinal subclusters were excluded. For comparison to Plass et al. (43) (Fig S4G-H), we identified 92 phagocyte-enriched transcripts and 27 goblet cell-enriched transcripts that were also represented in our 1844 intestine-enriched transcripts and phagocyte expression data. For global comparison of

transcripts enriched in laser captured intestine (Fig S4I), we included all transcripts enriched in intestinal cell types (and intestinal precursors) in single cell studies (32, 43, 88), without regard to cell type or enrichment in non-intestinal cell types. For comparison to Swapna et al. (88) data, BLASTN was used to identify dd_Smed_v6 transcripts orthologous to "SmedASXL" transcripts (with >95% identity and length >100 bp).

Gene cloning and Expressed Sequence Tags

Total RNA was extracted from planarians using Trizol with two chloroform extractions and high salt precipitation. After DNase digestion, cDNA was synthesized using the iScript Kit (Bio-Rad 1708891). Genes were amplified by PCR with Platinum *Taq* (Invitrogen 10966026) using primers listed in Table S1C. Amplicons were cloned into pJC53.2 digested with Eam1105I as described (130) and sequenced to verify clone identity and orientation. For some genes, expressed sequence tags (ESTs) in pBluescript II SK(+) were utilized (164), also listed in Table S1C.

In situ hybridization and immunofluorescence

In situ hybridizations were performed as described (60), with the following adjustments: NAc (7.5%) treatment was for 15 min; 4% formaldehyde fixation was for 15 min; and animals were bleached for 3 hr. Samples were pre-incubated with tyramide solution without H₂O₂ for 10 min, then spiked with H₂O₂ (.0003% final concentration), then developed for 10 min. Immunofluorescence after FISH was conducted as described (53), using 1x PBS, 0.3% Triton-X100, 0.6% IgG-free BSA, 0.45% fish gelatin as the blocking buffer (52).

Mucin identification

S. mediterranea mucin-like genes in PlanMine (11) and SmedGD2.0 (10) were identified by TBLASTN searches with human refseq_protein and UniProt mucin sequences. Planarian

sequences were translated with NCBI ORFinder (<https://www.ncbi.nlm.nih.gov/orffinder/>), and domain searches were conducted using NCBI CD-Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) (165), Pfam 31.0 (<https://pfam.xfam.org>) (166), and SMART (<http://smart.embl-heidelberg.de>) (167). Three planarian sequences were identified that encoded three N-terminal von Willebrand factor D domains and two or more N-terminal cysteine-rich and trypsin-inhibitor-like cysteine-rich domains characteristic of human mucins (e.g., MUC-2, MUC-5AC) (168, 169), but not von Willebrand factor A or Thrombospondin type I repeats found in closely related proteins (e.g., SCO-spondin). Planarian mucin-like genes identified using this approach are: *Smed-muc-like-1* (dd_Smed_v6_17988_0_1/SMED30009111), *Smed-muc-like-2* (dd_Smed_v6_18786_0_1/SMED30002668), and *Smed-muc-like-3* (dd_Smed_v6_21309_0_1/dd_Smed_v6_38233_0_1/dd_Smed_35076_0_1/SMED30006765).

Identification of intestine-enriched transcription factors and RNA interference

Putative transcription factors were identified by extracting intestine-enriched transcripts with "DNA" and/or "transcription" Biological Process and Molecular Function GO terms, then verifying that the best Uniprot homologs regulated transcription in published experimental evidence. *Smed-gli-1* has been previously studied (108). *Smed-RREB2* was most homologous to another planarian zinc finger protein, *Smed-RREBP1* (141). However, we used the more common "RREB" abbreviation (rather than "RREBP") for the second planarian paralog.

RNAi experiments were conducted as described (170) by mixing one microgram of in vitro-synthesized dsRNA with 1 µl of food coloring, 8 µl of water, and 40 µl of 2:1 liver:water homogenate. For the primary regeneration screen, animals were fed three times over 6 days, amputated 4-5 days after the last feeding, then fixed 6 days post amputation. For further analysis of *gli-1(RNAi)* and *RREB2(RNAi)* phenotypes in uninjured animals, planarians were fed 1X/week for 6 weeks or 2X/week for 3 weeks (6 feedings total). *gli-1(RNAi)* and *RREB2(RNAi)*

animals were scored as "non-eaters" if they refused food on two successive days (7 and 8 days after the previous feeding). Non-eating and/or curling animals were excluded from FISH analysis. For experiments in regenerates, animals were fed 2X/week for 4 weeks (8 feedings total).

Image Collection

Confocal images were collected on a Zeiss 710 Confocal microscope, with the following settings: two tracks were used, one with both 405/445 (excitation/emission nm, blue) and 565/650 (red), and the other with only 501/540 (green) to minimize bleed-through between channels. Detector gains were adjusted so that no pixels were saturated. Digital offset was set to 0 or -1, to ensure that most pixel intensities were non-zero, with averaging set to 2. Whole animals were captured with a single z-plane (5 μ m section) with a 10x objective, tiled, and stitched in Zen (version 11.0.3.190, 2012-SP2). Magnified regions were captured with a 63x Oil immersion objective and a z-stack (50-100 slices, 0.31 μ m/slice) was collected from the tail and/or head of the animal. In some cases, after imaging min/max or linear best fit adjustments were made in Zen to improve contrast of final images. Profile graphs (Figs 5-7) represent raw, unadjusted pixel intensity values from a single optical section.

WISH images were collected on a Zeiss Stemi 508 with an Axiocam 105 color camera, an Olympus SZX12 dissection microscope with an Axiocam MRc color camera, or a Zeiss Axio Zoom.V16 with an Axiocam 105 camera. In some cases brightness and/or contrast were adjusted in photoshop to improve signal contrast.

Quantification of animal area and length for *gli-1* and *RREB2* knockdown animals

Animals were separated into 35mm x 10mm petri dishes in groups of two, and imaged on a Zeiss Stemi 508 prior to the first dsRNA feeding, then again seven days after the fifth feeding (animals were fed once per week). For area, images were processed with ImageJ by

first applying an Auto Threshold, using method = Intermodes, Huang, or Triangles with “White objects on black background” selected, to highlight the planarian. Analyze Particles was then run, with size = 600000-10000000 μm^2 , and “Display results” and “Include holes” selected. For length, a straight line was drawn manually from the tip of the head to the tip of the tail and then Measure was used to obtain the length.

Data availability

Raw and processed RNA-Seq data associated with this study have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE135351.

Acknowledgments

We thank Forrest Waters for initial optimization of fixation, labeling, and RNA extraction for laser microdissection. We thank Mayandi Sivaguru (Institute for Genomic Biology, University of Illinois at Urbana-Champaign) and Muralidharan Jayaraman (University of Oklahoma Health Sciences Center) for help with LCM optimization, and Alvaro Hernandez (Roy J. Carver Biotechnology Center, Illinois) and Graham Wiley (OMRF Clinical Genomics Center) for library preparation and Illumina sequencing. We thank Ben Fowler, Amanda Valdez, Julie Crane, and Alex Weddle (OMRF Imaging Core) for technical assistance with confocal microscopy, histology, and LCM slide preparation; Jonathan Wren (OMRF) (Bioinformatics and Pathways Core) for advice on reciprocal blast searches; and Stuart Glenn and the OMRF Quantitative Analysis Core for high performance computing resources. We are grateful to Ricardo Zayas and Francesc Cebrià, who originally identified *Smed-npc2* as a marker for goblet cells. We gratefully acknowledge members of the Newmark and Forsthoefel Labs, Linda Thompson, Dean Dawson, and David Jones for discussions and manuscript critique.

Funding

This work was supported by the Oklahoma Center for Adult Stem Cell Research, a program of TSET (Grant 4340 to DJF), National Institutes of Health (NIH)/Centers of Biomedical Research Excellence (COBRE) GM103636 (Project 1 to DJF), the Oklahoma Medical Research Foundation, and NIH R01 HD043403 (to PAN). The OMRF Imaging Core and Bioinformatics and Pathways Core were supported by NIH/COBRE GM103636. The OMRF Quantitative Analysis Core was supported by NIH/COBRE GM110766. LMD was supported by the Molecular Biology Shared Resource, Stephenson Cancer Center (SCC) CCSG P30CA225520 and SCC COBRE P20GM103639. PAN is an investigator of the Howard Hughes Medical Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Brockes JP, Kumar A. Comparative aspects of animal regeneration. *Annu Rev Cell Dev Biol.* 2008;24:525-49.
2. Tanaka EM, Reddien PW. The cellular basis for animal regeneration. *Dev Cell.* 2011;21(1):172-85.
3. Elliott SA, Sánchez Alvarado A. Planarians and the History of Animal Regeneration: Paradigm Shifts and Key Concepts in Biology. *Methods Mol Biol.* 2018;1774:207-39.
4. Kumar A, Godwin JW, Gates PB, Garza-Garcia AA, Brockes JP. Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. *Science.* 2007;318(5851):772-7.
5. Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW. Muscle cells provide instructions for planarian regeneration. *Cell Rep.* 2013;4(4):633-41.
6. Gemberling M, Karra R, Dickson AL, Poss KD. Nrg1 is an injury-induced cardiomyocyte mitogen for the endogenous heart regeneration program in zebrafish. *Elife.* 2015;4.
7. Mokalled MH, Patra C, Dickson AL, Endo T, Stainier DY, Poss KD. Injury-induced *ctgfa* directs glial bridging and spinal cord regeneration in zebrafish. *Science.* 2016;354(6312):630-4.
8. Tanaka EM. The Molecular and Cellular Choreography of Appendage Regeneration. *Cell.* 2016;165(7):1598-608.
9. Newmark PA, Sánchez Alvarado A. Not your father's planarian: a classic model enters the era of functional genomics. *Nat Rev Genet.* 2002;3(3):210-9.
10. Robb SM, Gotting K, Ross E, Sánchez Alvarado A. SmedGD 2.0: The Schmidtea mediterranea genome database. *Genesis.* 2015;53(8):535-46.
11. Brandl H, Moon H, Vila-Farre M, Liu SY, Henry I, Rink JC. PlanMine--a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res.* 2016;44(D1):D764-73.
12. Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, et al. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature.* 2018;554(7690):56-61.
13. Reddien PW. The Cellular and Molecular Basis for Planarian Regeneration. *Cell.* 2018;175(2):327-45.
14. Reddien PW, Sánchez Alvarado A. Fundamentals of planarian regeneration. *Annu Rev Cell Dev Biol.* 2004;20:725-57.
15. Rink JC. Stem Cells, Patterning and Regeneration in Planarians: Self-Organization at the Organismal Scale. *Methods Mol Biol.* 2018;1774:57-172.
16. Pellettieri J. Regenerative tissue remodeling in planarians - The mysteries of morphallaxis. *Semin Cell Dev Biol.* 2018.

- 1034 17. Roberts-Galbraith RH, Newmark PA. On the organ trail: insights into organ regeneration
1035 in the planarian. *Curr Opin Genet Dev.* 2015;32:37-46.
- 1036 18. Hyman LH. The Acoelomate Bilateria: Phylum Platyhelminthes. *The Invertebrates:*
1037 *Platyhelminthes and Rhynchocoela. II.* New York: McGraw-Hill; 1951. p. 52-458.
- 1038 19. Forsthoefel DJ, Park AE, Newmark PA. Stem cell-based growth, regeneration, and
1039 remodeling of the planarian intestine. *Dev Biol.* 2011;356(2):445-59.
- 1040 20. Forsthoefel DJ, Newmark PA. Emerging patterns in planarian regeneration. *Curr Opin*
1041 *Genet Dev.* 2009;19(4):412-20.
- 1042 21. Umesono Y, Tasaki J, Nishimura Y, Hrouda M, Kawaguchi E, Yazawa S, et al. The
1043 molecular logic for planarian regeneration along the anterior-posterior axis. *Nature.*
1044 2013;500(7460):73-6.
- 1045 22. Hosoda K, Morimoto M, Motoishi M, Nishimura O, Agata K, Umesono Y. Simple blood-
1046 feeding method for live imaging of gut tube remodeling in regenerating planarians. *Dev Growth*
1047 *Differ.* 2016;58(3):260-9.
- 1048 23. Barberán S, Fraguas S, Cebrià F. The EGFR signaling pathway controls gut progenitor
1049 differentiation during planarian regeneration and homeostasis. *Development.*
1050 2016;143(12):2089-102.
- 1051 24. Forsthoefel DJ, James NP, Escobar DJ, Stary JM, Vieira AP, Waters FA, et al. An RNAi
1052 screen reveals intestinal regulators of branching morphogenesis, differentiation, and stem cell
1053 proliferation in planarians. *Dev Cell.* 2012;23(4):691-704.
- 1054 25. Adler CE, Sánchez Alvarado A. PHRED-1 is a divergent neurexin-1 homolog that
1055 organizes muscle fibers and patterns organs during regeneration. *Dev Biol.* 2017;427(1):165-75.
- 1056 26. Seebeck F, Marz M, Meyer AW, Reuter H, Vogg MC, Stehling M, et al. Integrins are
1057 required for tissue organization and restriction of neurogenesis in regenerating planarians.
1058 *Development.* 2017;144(5):795-807.
- 1059 27. Bonar NA, Petersen CP. Integrin suppresses neurogenesis and regulates brain tissue
1060 assembly in planarian regeneration. *Development.* 2017;144(5):784-94.
- 1061 28. Scimone ML, Wurtzel O, Malecek K, Fincher CT, Oderberg IM, Kravarik KM, et al. *foxF-*
1062 *1* Controls Specification of Non-body Wall Muscle and Phagocytic Cells in Planarians. *Curr Biol.*
1063 2018;28(23):3787-801 e6.
- 1064 29. Willier BH, Hyman LH, Rifenburgh SA. A histochemical study of intracellular digestion in
1065 triclad flatworms. *Journal of Morphology and Physiology.* 1925;40(2):299-340.
- 1066 30. Ishii S. Electron microscopic observations on the Planarian tissues II. The intestine.
1067 *Fukushima J Med Sci.* 1965;12(1):67-87.
- 1068 31. Bowen ID, Ryder TA, Thompson JA. The fine structure of the planarian *Polycelis tenuis*
1069 *Iijima.* II. The intestine and gastrodermal phagocytosis. *Protoplasma.* 1974;79(1):1-17.

1070 32. Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. Cell type transcriptome
1071 atlas for the planarian *Schmidtea mediterranea*. *Science*. 2018;360(6391).

1072 33. Wagner DE, Wang IE, Reddien PW. Clonogenic neoblasts are pluripotent adult stem
1073 cells that underlie planarian regeneration. *Science*. 2011;332(6031):811-6.

1074 34. van Wolfswinkel JC, Wagner DE, Reddien PW. Single-cell analysis reveals functionally
1075 distinct classes within the planarian stem cell compartment. *Cell Stem Cell*. 2014;15(3):326-39.

1076 35. Labbé RM, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DD, et al. A Comparative
1077 Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians
1078 and Mammals. *Stem Cells*. 2012;30(8):1734-45.

1079 36. Zeng A, Li H, Guo L, Gao X, McKinney S, Wang Y, et al. Prospectively Isolated
1080 Tetraspanin(+) Neoblasts Are Adult Pluripotent Stem Cells Underlying Planaria Regeneration.
1081 *Cell*. 2018;173(7):1593-608 e20.

1082 37. Flores NM, Oviedo NJ, Sage J. Essential role for the planarian intestinal GATA
1083 transcription factor in stem cells and regeneration. *Dev Biol*. 2016;418(1):179-88.

1084 38. Henderson JM, Nisperos SV, Weeks J, Ghulam M, Marin I, Zayas RM. Identification of
1085 HECT E3 ubiquitin ligase family genes involved in stem cell regulation and regeneration in
1086 planarians. *Dev Biol*. 2015;404(2):21-34.

1087 39. Miller CM, Newmark PA. An insulin-like peptide regulates size and adult stem cells in
1088 planarians. *Int J Dev Biol*. 2012;56(1-3):75-82.

1089 40. Gaviño MA, Wenemoser D, Wang IE, Reddien PW. Tissue absence initiates
1090 regeneration through follistatin-mediated inhibition of activin signaling. *Elife*. 2013;2:e00247.

1091 41. Dingwall CB, King RS. Muscle-derived matrix metalloproteinase regulates stem cell
1092 proliferation in planarians. *Dev Dyn*. 2016;245(9):963-70.

1093 42. Lei K, Thi-Kim Vu H, Mohan RD, McKinney SA, Seidel CW, Alexander R, et al. *Egf*
1094 Signaling Directs Neoblast Repopulation by Regulating Asymmetric Cell Division in Planarians.
1095 *Dev Cell*. 2016;38(4):413-29.

1096 43. Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glazar P, et al. Cell type atlas and
1097 lineage tree of a whole complex animal by single-cell transcriptomics. *Science*. 2018;360(6391).

1098 44. Baguñà J, Romero R. Quantitative analysis of cell types during growth, degrowth and
1099 regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hydrobiologia*.
1100 1981;84(1):181-94.

1101 45. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al.
1102 Laser capture microdissection. *Science*. 1996;274(5289):998-1001.

1103 46. Mahalingam M. Laser Capture Microdissection: Insights into Methods and Applications.
1104 *Methods Mol Biol*. 2018;1723:1-17.

- 1105 47. Bevilacqua C, Ducos B. Laser microdissection: A powerful tool for genomics at cell level.
1106 Mol Aspects Med. 2018;59:5-27.
- 1107 48. Espina V, Wulfkühle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, et al. Laser-
1108 capture microdissection. Nat Protoc. 2006;1(2):586-603.
- 1109 49. Goldsworthy SM, Stockton PS, Trempus CS, Foley JF, Maronpot RR. Effects of fixation
1110 on RNA extraction and amplification from laser capture microdissected tissue. Mol Carcinog.
1111 1999;25(2):86-91.
- 1112 50. Gillespie JW, Best CJ, Bichsel VE, Cole KA, Greenhut SF, Hewitt SM, et al. Evaluation
1113 of non-formalin tissue fixation for molecular profiling studies. Am J Pathol. 2002;160(2):449-57.
- 1114 51. Golubeva YG, Warner AC. Laser Microdissection Workflow for Isolating Nucleic Acids
1115 from Fixed and Frozen Tissue Samples. Methods Mol Biol. 2018;1723:33-93.
- 1116 52. Forsthoefel DJ, Waters FA, Newmark PA. Generation of cell type-specific monoclonal
1117 antibodies for the planarian and optimization of sample processing for immunolabeling. BMC
1118 Dev Biol. 2014;14:45.
- 1119 53. Ross KG, Omuro KC, Taylor MR, Munday RK, Hubert A, King RS, et al. Novel
1120 monoclonal antibodies to study tissue regeneration in planarians. BMC Dev Biol. 2015;15:2.
- 1121 54. Ishikawa H. Evolution of ribosomal RNA. Comp Biochem Physiol B. 1977;58(1):1-7.
- 1122 55. Matz MV. Amplification of representative cDNA samples from microscopic amounts of
1123 invertebrate tissue to search for new genes. Methods Mol Biol. 2002;183:3-18.
- 1124 56. Winnebeck EC, Millar CD, Warman GR. Why does insect RNA look degraded? J Insect
1125 Sci. 2010;10:159.
- 1126 57. Asai S, Ianora A, Lauritano C, Lindeque PK, Carotenuto Y. High-quality RNA extraction
1127 from copepods for Next Generation Sequencing: A comparative study. Mar Genomics. 2015;24
1128 Pt 1:115-8.
- 1129 58. Umesono Y, Watanabe K, Agata K. A planarian *orthopedia* homolog is specifically
1130 expressed in the branch region of both the mature and regenerating brain. Dev Growth Differ.
1131 1997;39(6):723-7.
- 1132 59. Pearson BJ, Eisenhoffer GT, Gurley KA, Rink JC, Miller DE, Sánchez Alvarado A.
1133 Formaldehyde-based whole-mount in situ hybridization method for planarians. Dev Dyn.
1134 2009;238(2):443-50.
- 1135 60. King RS, Newmark PA. *In situ* hybridization protocol for enhanced detection of gene
1136 expression in the planarian *Schmidtea mediterranea*. BMC Dev Biol. 2013;13:8.
- 1137 61. Gurley KA, Rink JC, Sánchez Alvarado A. β -catenin defines head versus tail identity
1138 during planarian regeneration and homeostasis. Science. 2008;319(5861):323-7.
- 1139 62. Petersen CP, Reddien PW. *Smed- β catenin-1* is required for anteroposterior blastema
1140 polarity in planarian regeneration. Science. 2008;319(5861):327-30.

1141 63. Iglesias M, Gomez-Skarmeta JL, Saló E, Adell T. Silencing of *Smed-beta-catenin1*
1142 generates radial-like hypercephalized planarians. *Development*. 2008;135(7):1215-21.

1143 64. Reuter H, Marz M, Vogg MC, Eccles D, Grifol-Boldu L, Wehner D, et al. Beta-catenin-
1144 dependent control of positional information along the AP body axis in planarians involves a
1145 teashirt family member. *Cell Rep*. 2015;10(2):253-65.

1146 65. Thi-Kim Vu H, Rink JC, McKinney SA, McClain M, Lakshmanaperumal N, Alexander R,
1147 et al. Stem cells and fluid flow drive cyst formation in an invertebrate excretory organ. *Elife*.
1148 2015;4.

1149 66. Stuckemann T, Cleland JP, Werner S, Thi-Kim Vu H, Bayersdorf R, Liu SY, et al.
1150 Antagonistic Self-Organizing Patterning Systems Control Maintenance and Regeneration of the
1151 Anteroposterior Axis in Planarians. *Dev Cell*. 2017;40(3):248-63 e4.

1152 67. Gamo J, Garcia-Corrales P. Ultrastructural changes in the gastrodermal phagocytic cells
1153 of the freshwater planarian *Dugesia gonocephala* s.l. during starvation. *J Submicrosc Cytol*.
1154 1987;19(2):291-302.

1155 68. Xu M, Joo HJ, Paik YK. Novel functions of lipid-binding protein 5 in *Caenorhabditis*
1156 *elegans* fat metabolism. *J Biol Chem*. 2011;286(32):28111-8.

1157 69. Buchon N, Osman D, David FP, Fang HY, Boquete JP, Deplancke B, et al.
1158 Morphological and molecular characterization of adult midgut compartmentalization in
1159 *Drosophila*. *Cell Rep*. 2013;3(5):1725-38.

1160 70. Andre M, Ando S, Ballagny C, Durliat M, Poupard G, Briancon C, et al. Intestinal fatty
1161 acid binding protein gene expression reveals the cephalocaudal patterning during zebrafish gut
1162 morphogenesis. *Int J Dev Biol*. 2000;44(2):249-52.

1163 71. Chalmers AD, Slack JM. Development of the gut in *Xenopus laevis*. *Dev Dyn*.
1164 1998;212(4):509-21.

1165 72. Gajda AM, Storch J. Enterocyte fatty acid-binding proteins (FABPs): different functions
1166 of liver and intestinal FABPs in the intestine. *Prostaglandins Leukot Essent Fatty Acids*.
1167 2015;93:9-16.

1168 73. Jennings JB. Studies on feeding, digestion, and food storage in free-living flatworms
1169 (Platyhelminthes: Turbellaria). *Biol Bull*. 1957;112(1):63-80.

1170 74. Garcia-Corrales P, Gamo J. Ultrastructural changes in the gastrodermal phagocytic cells
1171 of the planarian *Dugesia gonocephala* s.l. during food digestion (Plathelminthes).
1172 *Zoomorphology*. 1988;108(2):109-17.

1173 75. Rindi G, Wiedenmann B. Neuroendocrine neoplasms of the gut and pancreas: new
1174 insights. *Nat Rev Endocrinol*. 2011;8(1):54-64.

1175 76. Spang A. Membrane Tethering Complexes in the Endosomal System. *Front Cell Dev*
1176 *Biol*. 2016;4:35.

1177 77. Zhang CS, Jiang B, Li M, Zhu M, Peng Y, Zhang YL, et al. The lysosomal v-ATPase-
1178 Ragulator complex is a common activator for AMPK and mTORC1, acting as a switch between
1179 catabolism and anabolism. *Cell Metab.* 2014;20(3):526-40.

1180 78. Komatsu M, Waguri S, Ueno T, Iwata J, Murata S, Tanida I, et al. Impairment of
1181 starvation-induced and constitutive autophagy in Atg7-deficient mice. *J Cell Biol.*
1182 2005;169(3):425-34.

1183 79. Sou YS, Waguri S, Iwata J, Ueno T, Fujimura T, Hara T, et al. The Atg8 conjugation
1184 system is indispensable for proper development of autophagic isolation membranes in mice.
1185 *Mol Biol Cell.* 2008;19(11):4762-75.

1186 80. Bowen ED, Ryder TA, Dark C. The effects of starvation on the planarian worm *Polycelis*
1187 *tenuis* Iijima. *Cell Tissue Res.* 1976;169(2):193-209.

1188 81. González-Estévez C, Felix DA, Smith MD, Paps J, Morley SJ, James V, et al. SMG-1
1189 and mTORC1 act antagonistically to regulate response to injury and growth in planarians. *PLoS*
1190 *Genet.* 2012;8(3):e1002619.

1191 82. Felix DA, Gutiérrez-Gutiérrez Ó, Espada L, Thems A, González-Estévez C. It is not all
1192 about regeneration: Planarians striking power to stand starvation. *Semin Cell Dev Biol.* 2018.

1193 83. Straight SW, Pieczynski JN, Whiteman EL, Liu CJ, Margolis B. Mammalian lin-7
1194 stabilizes polarity protein complexes. *J Biol Chem.* 2006;281(49):37738-47.

1195 84. Simske JS, Kaech SM, Harp SA, Kim SK. LET-23 receptor localization by the cell
1196 junction protein LIN-7 during *C. elegans* vulval induction. *Cell.* 1996;85(2):195-204.

1197 85. Kurata S. Peptidoglycan recognition proteins in *Drosophila* immunity. *Dev Comp*
1198 *Immunol.* 2014;42(1):36-41.

1199 86. Dziarski R, Gupta D. How innate immunity proteins kill bacteria and why they are not
1200 prone to resistance. *Curr Genet.* 2018;64(1):125-9.

1201 87. Kowalski EJA, Li L. Toll-Interacting Protein in Resolving and Non-Resolving
1202 Inflammation. *Front Immunol.* 2017;8:511.

1203 88. Swapna LS, Molinaro AM, Lindsay-Mosher N, Pearson BJ, Parkinson J. Comparative
1204 transcriptomic analyses and single-cell RNA sequencing of the freshwater planarian *Schmidtea*
1205 *mediterranea* identify major cell types and pathway conservation. *Genome Biol.* 2018;19(1):124.

1206 89. Xie P. TRAF molecules in cell signaling and in human diseases. *J Mol Signal.*
1207 2013;8(1):7.

1208 90. Arnold CP, Merryman MS, Harris-Arnold A, McKinney SA, Seidel CW, Loethen S, et al.
1209 Pathogenic shifts in endogenous microbiota impede tissue regeneration via distinct activation of
1210 TAK1/MKK/p38. *Elife.* 2016;5.

1211 91. Abnave P, Mottola G, Gimenez G, Boucherit N, Trouplin V, Torre C, et al. Screening in
1212 planarians identifies MORN2 as a key component in LC3-associated phagocytosis and
1213 resistance to bacterial infection. *Cell Host Microbe.* 2014;16(3):338-50.

- 1214 92. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y. Predicting
1215 function: from genes to genomes and back. *J Mol Biol.* 1998;283(4):707-25.
- 1216 93. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science.*
1217 1997;278(5338):631-7.
- 1218 94. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*
1219 2018;46(5):2699.
- 1220 95. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al.
1221 Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
- 1222 96. Rink JC, Vu HT, Sánchez Alvarado A. The maintenance and regeneration of the
1223 planarian excretory system are regulated by EGFR signaling. *Development.* 2011;138(17):3769-
1224 80.
- 1225 97. Scimone ML, Srivastava M, Bell GW, Reddien PW. A regulatory program for excretory
1226 system regeneration in planarians. *Development.* 2011;138(20):4387-98.
- 1227 98. Yen CL, Farese RV, Jr. MGAT2, a monoacylglycerol acyltransferase expressed in the
1228 small intestine. *J Biol Chem.* 2003;278(20):18532-7.
- 1229 99. Braeuning A, Ittrich C, Kohle C, Hailfinger S, Bonin M, Buchmann A, et al. Differential
1230 gene expression in periportal and perivenous mouse hepatocytes. *FEBS J.* 2006;273(22):5051-
1231 61.
- 1232 100. Pilotte L, Larrieu P, Stroobant V, Colau D, Dolusic E, Frederick R, et al. Reversal of
1233 tumoral immune resistance by inhibition of tryptophan 2,3-dioxygenase. *Proc Natl Acad Sci U S*
1234 *A.* 2012;109(7):2497-502.
- 1235 101. Lamont BJ, Visinoni S, Fam BC, Kebede M, Weinrich B, Papapostolou S, et al.
1236 Expression of human fructose-1,6-bisphosphatase in the liver of transgenic mice results in
1237 increased glycerol gluconeogenesis. *Endocrinology.* 2006;147(6):2764-72.
- 1238 102. Visinoni S, Khalid NF, Joannides CN, Shulkes A, Yim M, Whitehead J, et al. The role of
1239 liver fructose-1,6-bisphosphatase in regulating appetite and adiposity. *Diabetes.*
1240 2012;61(5):1122-32.
- 1241 103. Gordon DA, Jamil H, Sharp D, Mullaney D, Yao Z, Gregg RE, et al. Secretion of
1242 apolipoprotein B-containing lipoproteins from HeLa cells is dependent on expression of the
1243 microsomal triglyceride transfer protein and is regulated by lipid availability. *Proc Natl Acad Sci*
1244 *U S A.* 1994;91(16):7628-32.
- 1245 104. Garcia-Corrales P, Gamo J. The ultrastructure of the gastrodermal gland cells in the
1246 freshwater planarian *Dugesia gonocephala* s.l. *Acta Zool.* 1986;67(1):43-51.
- 1247 105. Garcia-Fernández J, Baguña J, Saló E. Genomic organization and expression of the
1248 planarian homeobox genes *Dth-1* and *Dth-2*. *Development.* 1993;118(1):241-53.
- 1249 106. Zayas RM, Cebrià F, Guo T, Feng J, Newmark PA. The use of lectins as markers for
1250 differentiated secretory cells in planarians. *Dev Dyn.* 2010;239(11):2888-97.

1251 107. Bueno D, Baguña J, Romero R. Cell-, tissue-, and position-specific monoclonal
1252 antibodies against the planarian *Dugesia (Girardia) tigrina*. Histochem Cell Biol.
1253 1997;107(2):139-49.

1254 108. Rink JC, Gurley KA, Elliott SA, Sánchez Alvarado A. Planarian Hh signaling regulates
1255 regeneration polarity and links Hh pathway evolution to cilia. Science. 2009;326(5958):1406-10.

1256 109. Kobayashi C, Kobayashi S, Orii H, Watanabe K, Agata K. Identification of two distinct
1257 muscles in the planarian *Dugesia japonica* by their expression of myosin heavy chain genes.
1258 Zoolog Sci. 1998;15:861-9.

1259 110. Orii H, Ito H, Watanabe K. Anatomy of the planarian *Dugesia japonica* I. The muscular
1260 system revealed by antisera against myosin heavy chains. Zoolog Sci. 2002;19(10):1123-31.

1261 111. Arnold G. Intra-cellular and General Digestive Processes in Planariae. Quart Jour Micr
1262 Science. 1909;54:207-20.

1263 112. Pedersen KJ. Some observations on the fine structure of planarian protonephridia and
1264 gastrodermal phagocytes. Z Zellforsch. 1961;53:609-28.

1265 113. Pascolini R, Gargiulo AM. Ultrastructural modifications of cells in the intestine of *Dugesia*
1266 *lugubris*. Boll Zool. 1975;42:123-31.

1267 114. Dalton AJ. A study of the golgi material of hepatic and intestinal epithelial cells with the
1268 electron microscope. Z Zellforsch Mikrosk Anat. 1952;36(6):522-4.

1269 115. Freeman JA. Goblet cell fine structure. Anat Rec. 1966;154(1):121-47.

1270 116. McCauley HA, Guasch G. Three cheers for the goblet cell: maintaining homeostasis in
1271 mucosal epithelia. Trends Mol Med. 2015;21(8):492-503.

1272 117. Birchenough GM, Johansson ME, Gustafsson JK, Bergstrom JH, Hansson GC. New
1273 developments in goblet cell mucus secretion and function. Mucosal Immunol. 2015;8(4):712-9.

1274 118. Knoop KA, Newberry RD. Goblet cells: multifaceted players in immunity at mucosal
1275 surfaces. Mucosal Immunol. 2018.

1276 119. Glazer AM, Wilkinson AW, Backer CB, Lapan SW, Gutzman JH, Cheeseman IM, et al.
1277 The Zn finger protein Iguana impacts Hedgehog signaling by promoting ciliogenesis. Dev Biol.
1278 2010;337(1):148-56.

1279 120. Yazawa S, Umesono Y, Hayashi T, Tarui H, Agata K. Planarian Hedgehog/Patched
1280 establishes anterior-posterior polarity by regulating Wnt signaling. Proc Natl Acad Sci U S A.
1281 2009;106(52):22329-34.

1282 121. Currie KW, Molinaro AM, Pearson BJ. Neuronal sources of hedgehog modulate
1283 neurogenesis in the adult planarian brain. Elife. 2016;5.

1284 122. Jennings JB. Further studies on feeding and digestion in Triclad Turbellaria. Biol Bull.
1285 1962;123(3):571-81.

1286 123. Prassas I, Eissa A, Poda G, Diamandis EP. Unleashing the therapeutic potential of
1287 human kallikrein-related serine proteases. *Nat Rev Drug Discov.* 2015;14(3):183-202.

1288 124. Schachter M, Longridge DJ, Wheeler GD, Mehta JG, Uchida Y. Immunocytochemical
1289 and enzyme histochemical localization of kallikrein-like enzymes in colon, intestine, and
1290 stomach of rat and cat. *J Histochem Cytochem.* 1986;34(7):927-34.

1291 125. Grün D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, et al. De
1292 Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell.*
1293 2016;19(2):266-77.

1294 126. Stadnicki A. Intestinal tissue kallikrein-kinin system in inflammatory bowel disease.
1295 *Inflamm Bowel Dis.* 2011;17(2):645-54.

1296 127. Kontos CK, Mavridis K, Talieri M, Scorilas A. Kallikrein-related peptidases (KLKs) in
1297 gastrointestinal cancer: mechanistic and clinical aspects. *Thromb Haemost.* 2013;110(3):450-7.

1298 128. Pauls D, Chen J, Reiher W, Vanselow JT, Schlosser A, Kahnt J, et al. Peptidomics and
1299 processing of regulatory peptides in the fruit fly *Drosophila*. *EuPA Open Prot.* 2014;3:114-27.

1300 129. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JW. PCSK1 Mutations and Human
1301 Endocrinopathies: From Obesity to Gastrointestinal Disorders. *Endocr Rev.* 2016;37(4):347-71.

1302 130. Collins JJ, 3rd, Hou X, Romanova EV, Lambrus BG, Miller CM, Saberi A, et al. Genome-
1303 wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS*
1304 *Biol.* 2010;8(10):e1000509.

1305 131. Chang SK, Dohrman AF, Basbaum CB, Ho SB, Tsuda T, Toribara NW, et al.
1306 Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human
1307 normal intestine and colon cancer. *Gastroenterology.* 1994;107(1):28-36.

1308 132. Johansson ME, Larsson JM, Hansson GC. The two mucus layers of colon are organized
1309 by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc*
1310 *Natl Acad Sci U S A.* 2011;108 Suppl 1:4659-65.

1311 133. Martindale MQ, Hejnal A. A developmental perspective: changes in the position of the
1312 blastopore during bilaterian evolution. *Dev Cell.* 2009;17(2):162-74.

1313 134. Telford MJ, Budd GE, Philippe H. Phylogenomic Insights into Animal Evolution. *Curr*
1314 *Biol.* 2015;25(19):R876-87.

1315 135. Gagné-Sansfaçon J, Allaire JM, Jones C, Boudreau F, Perreault N. Loss of Sonic
1316 hedgehog leads to alterations in intestinal secretory cell maturation and autophagy. *PLoS One.*
1317 2014;9(6):e98751.

1318 136. Liang R, Morris P, Cho SS, Abud HE, Jin X, Cheng W. Hedgehog signaling displays a
1319 biphasic expression pattern during intestinal injury and repair. *J Pediatr Surg.*
1320 2012;47(12):2251-63.

1321 137. Ray SK, Nishitani J, Petry MW, Fessing MY, Leiter AB. Novel transcriptional potentiation
1322 of BETA2/NeuroD on the secretin gene promoter by the DNA-binding protein Finb/RREB-1. Mol
1323 Cell Biol. 2003;23(1):259-71.

1324 138. Tian A, Shi Q, Jiang A, Li S, Wang B, Jiang J. Injury-stimulated Hedgehog signaling
1325 promotes regenerative proliferation of Drosophila intestinal stem cells. J Cell Biol.
1326 2015;208(6):807-19.

1327 139. Baechler BL, McKnight C, Pruchnicki PC, Biro NA, Reed BH. Hindsight/RREB-1
1328 functions in both the specification and differentiation of stem cells in the adult midgut of
1329 Drosophila. Biol Open. 2015;5(1):1-10.

1330 140. Wang IE, Lapan SW, Scimone ML, Clandinin TR, Reddien PW. Hedgehog signaling
1331 regulates gene expression in planarian glia. Elife. 2016;5.

1332 141. Mihaylova Y, Abnave P, Kao D, Hughes S, Lai A, Jaber-Hijazi F, et al. Conservation of
1333 epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. Nat
1334 Commun. 2018;9(1):3633.

1335 142. González-Sastre A, De Sousa N, Adell T, Saló E. The pioneer factor Smed-gata456-1 is
1336 required for gut cell differentiation and maintenance in planarians. Int J Dev Biol. 2017;61(1-
1337 2):53-63.

1338 143. Wurtzel O, Cote LE, Poirier A, Satija R, Regev A, Reddien PW. A generic and cell-type-
1339 specific wound response precedes regeneration in planarians. Dev Cell. 2015;35(5):632-45.

1340 144. Wurtzel O, Oderberg IM, Reddien PW. Planarian Epidermal Stem Cells Respond to
1341 Positional Cues to Promote Cell-Type Diversity. Dev Cell. 2017;40(5):491-504 e5.

1342 145. Wu AR, Wang J, Streets AM, Huang Y. Single-Cell Transcriptional Analysis. Annu Rev
1343 Anal Chem (Palo Alto Calif). 2017;10(1):439-62.

1344 146. Sonesson C, Robinson MD. Bias, robustness and scalability in single-cell differential
1345 expression analysis. Nat Methods. 2018;15(4):255-61.

1346 147. Okano D, Ishida S, Ishiguro S, Kobayashi K. Light and electron microscopic studies of
1347 the intestinal epithelium in *Notoplana humilis* (Platyhelminthes, Polycladida): the contribution of
1348 mesodermal/gastrodermal neoblasts to intestinal regeneration. Cell Tissue Res.
1349 2015;362(3):529-40.

1350 148. Ortiz-Pineda PA, Ramírez-Gómez F, Pérez-Ortiz J, González-Díaz S, Santiago-De
1351 Jesús F, Hernández-Pasos J, et al. Gene expression profiling of intestinal regeneration in the
1352 sea cucumber. BMC Genomics. 2009;10:262.

1353 149. Sun L, Chen M, Yang H, Wang T, Liu B, Shu C, et al. Large scale gene expression
1354 profiling during intestine and body wall regeneration in the sea cucumber *Apostichopus*
1355 *japonicus*. Comp Biochem Physiol Part D Genomics Proteomics. 2011;6(2):195-205.

1356 150. Zhang X, Sun L, Yuan J, Sun Y, Gao Y, Zhang L, et al. The sea cucumber genome
1357 provides insights into morphological evolution and visceral regeneration. PLoS Biol.
1358 2017;15(10):e2003790.

1359 151. Zattara EE, Bely AE. Evolution of a novel developmental trajectory: fission is distinct
1360 from regeneration in the annelid *Pristina leidyi*. *Evol Dev*. 2011;13(1):80-95.

1361 152. Takeo M, Yoshida-Noro C, Tochinal S. Morphallactic regeneration as revealed by
1362 region-specific gene expression in the digestive tract of *Enchytraeus japonensis* (Oligochaeta,
1363 Annelida). *Dev Dyn*. 2008;237(5):1284-94.

1364 153. Kaneko N, Katsuyama Y, Kawamura K, Fujiwara S. Regeneration of the gut requires
1365 retinoic acid in the budding ascidian *Polyandrocarpa misakiensis*. *Dev Growth Differ*.
1366 2010;52(5):457-68.

1367 154. Goodchild CG. Reconstitution of the intestinal tract in the adult leopard frog, *Rana*
1368 *pipiens* Schreber. *J Exp Zool*. 1956;131:301-27.

1369 155. O'Steen WK. Regeneration of the intestine in adult urodeles. *J Morphol*.
1370 1958;103(3):435-77.

1371 156. Ayyaz A, Kumar S, Sangiorgi B, Ghoshal B, Gosio J, Ouladan S, et al. Single-cell
1372 transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature*.
1373 2019;569(7754):121-5.

1374 157. Sánchez Alvarado A, Newmark PA, Robb SM, Juste R. The *Schmidtea mediterranea*
1375 database as a molecular resource for studying platyhelminthes, stem cells and regeneration.
1376 *Development*. 2002;129(24):5659-65.

1377 158. Roberts-Galbraith RH, Newmark PA. Follistatin antagonizes Activin signaling and acts
1378 with Notum to direct planarian head regeneration. *Proc Natl Acad Sci U S A*. 2013;110(4):1363-
1379 8.

1380 159. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010
1381 [Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

1382 160. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.
1383 2012;9(4):357-9.

1384 161. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1385 Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.

1386 162. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
1387 expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40.

1388 163. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess
1389 overrepresentation of gene ontology categories in biological networks. *Bioinformatics*.
1390 2005;21(16):3448-9.

1391 164. Zayas RM, Hernandez A, Habermann B, Wang Y, Stary JM, Newmark PA. The
1392 planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of
1393 ESTs from the hermaphroditic strain. *Proc Natl Acad Sci USA*. 2005;102(51):18491-6.

1394 165. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD:
1395 NCBI's conserved domain database. *Nucleic Acids Res*. 2015;43(Database issue):D222-6.

1396 166. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam
1397 protein families database: towards a more sustainable future. *Nucleic Acids Res.*
1398 2016;44(D1):D279-85.

1399 167. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic*
1400 *Acids Res.* 2018;46(D1):D493-D6.

1401 168. Lang T, Hansson GC, Samuelsson T. Gel-forming mucins appeared early in metazoan
1402 evolution. *Proc Natl Acad Sci U S A.* 2007;104(41):16209-14.

1403 169. Lang T, Klasson S, Larsson E, Johansson ME, Hansson GC, Samuelsson T. Searching
1404 the Evolutionary Origin of Epithelial Mucus Protein Components-Mucins and FCGBP. *Mol Biol*
1405 *Evol.* 2016;33(8):1921-36.

1406 170. Rouhana L, Weiss JA, Forsthoefel DJ, Lee H, King RS, Inoue T, et al. RNA interference
1407 by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics.
1408 *Dev Dyn.* 2013;242(6):718-30.
1409
1410
1411
1412

Figure Legends

Figure 1. Laser-capture microdissection coupled with RNA-Seq identifies 1844 intestine-enriched transcripts. (A) Schematic of microdissection workflow. Planarians were fixed and cryosectioned, then sections were stained with Eosin Y. Medial intestine, lateral intestine, and non-intestine tissue were then laser captured, followed by RNA extraction and RNA-Seq. **(B)** Images of an eosin-stained section as tissue is progressively removed (left) and captured (right), yielding three samples with medial intestine, lateral intestine, and non-intestinal tissue. **(C)** Pie chart of RNA-Seq results: of 28,069 total transcripts, 13,136 were detected. Of these, 1844 were upregulated significantly in either medial or lateral intestine. **(D)** Venn diagram showing overlap of medially and laterally enriched intestinal transcripts. **(E)** Schematic of the number of transcripts with enrichment in medial or lateral intestine, expressed as a ratio of Fold Changes (FC) in each region. Dark green, FC-medial/FC-lateral > 2x. Green, FC-medial/FC-lateral = 1.5x-2x. Dark blue, FC-medial/FC-lateral = 1x-1.5x. Blue, FC-lateral/FC-medial = 1x-1.5x. Orange, FC-lateral/FC-medial = 1.5x-2x. Dark orange, FC-lateral/FC-medial > 2x. **(F)** Examples of transcripts expressed in the intestine (WISH) in medial (top), mediolateral (middle), and lateral (bottom) regions. Color outlines correspond to the color bar in panel F. Detailed gene ID information is available in Table S1 and in Results. Scale bars, 100 μ m (B), 200 μ m (F).

Figure 2. Transcripts involved in metabolism, transport, organelle physiology, transport, and stimuli responses are enriched in the intestine. (A) Biological Process Gene Ontology terms significantly over-represented in intestine-enriched transcripts. Bubble size indicates fold enrichment relative to all transcripts detected in laser-microdissected tissue, while position on the x-axis indicates FDR-adjusted significance. Numbers in parentheses indicate the number of intestine-enriched transcripts annotated with each term. **(B)** Examples of intestinal expression

(WISH) of transcripts involved in biological processes indicated. Detailed gene ID information is available in Table S1 and in Results. Scale bars, 200 μ m.

Figure 3. Homologs of planarian intestinal transcripts are enriched in human digestive tissues. (A) 5,583 *S. mediterranea* and *Homo sapiens* UniProt transcripts hit each other in reciprocal TBLASTX queries. (B) 5,561/5,583 human UniProt RBH homologs of planarian transcripts were present in the Human Protein Atlas. (C) 699/5,561 RBH homologs were enriched in the planarian intestine. (D) Enrichment of RBH homologs in human tissues. The first number in parentheses is the number of planarian intestine-enriched RBH homologs (of 699) and the second number in parentheses is the number of all RBH homologs (of 5,561) for each human tissue. Histogram bars represent percentage of planarian transcripts with RBH homologs enriched in human tissues. For example, 19/699 (2.72%) UniProt RBH homologs of planarian intestine-enriched transcripts were tissue-enriched, tissue-enhanced, or group-enriched in the duodenum, while only 41/5561 (0.74%) of all UniProt RBH homologs of planarian transcripts were similarly enriched. (E) Fold enrichment (planarian intestine/non-intestine) of RBH homologs for each human tissue. Histogram bars were calculated as a ratio of percentages in panel D. For example, 2.72% of 699 planarian intestinal RBH homologs and 0.74% of all 5,561 planarian RBH homologs were enriched in the duodenum, yielding fold enrichment of $2.72/0.74 = 3.68X$. (F) Intestinal expression (WISH) of transcripts with RBH homologs enriched in human digestive organs: *hnf-4* (duodenum, small intestine); *slc10a2* (duodenum, kidney, small intestine); *mogat2* (colon, duodenum, liver, rectum, small intestine); *slc5a1* (duodenum, small intestine); *plb1* (small intestine); *uroc1* and *tdo2* (liver only); *sult1c14* (gallbladder only); *ggt1*, *upp2* (kidney only). With the exception of *Smed-hnf-4*, gene names are based on human Uniprot best hit IDs. Detailed gene ID information is available in Table S1 and in Results. Scale bars, 200 μ m.

Figure 4. Identification of transcripts enriched in specific intestine cell types and regions.

(A) Log₂ fold-changes for laser-microdissected medial (left) and lateral (right) intestinal tissue (relative to non-intestinal tissue) are plotted on the x-axis, while log₂ fold-changes for sorted/purified intestinal phagocytes (compared to all other cell types) are plotted on the y-axis. Transcripts in the upper right quadrant (colorized according to the legend in Fig 1 and at the bottom of this figure) are expressed preferentially in laser captured intestine (fold-change>2, FDR-adjusted *p* value<.01) and preferentially in sorted phagocytes (fold-change>2, FDR-adjusted *p* value<.05) (phagocytes: “high”). Most of these transcripts are not medially or laterally enriched in LCM transcriptomes. **(B)** Whole-mount in situ hybridizations on uninjured planarians showing examples of expression patterns for transcripts in (A). Borders are colorized according to the mediolateral legend in Fig 1 and at the bottom of the figure. Expression patterns are mostly uniform and ubiquitous in the intestine, consistent with phagocyte-specific expression, with the exception of *fh13* (top left), which is medially enriched. **(C)** Plots as in (A), but with colorized transcripts expressed preferentially in laser-captured intestine, but not significantly up- or down-regulated in sorted phagocytes (FDR-adjusted *p* value>.05) (phagocytes: “moderate”). Some of these transcripts are medially or laterally enriched in LCM transcriptomes. **(D)** Examples of gene expression for transcripts in (C). A variety of intestine expression patterns is observed. **(E)** Plots as in (A), with transcripts in the lower right quadrant enriched in laser microdissected intestine, but significantly downregulated (fold-change <2, FDR-adjusted *p* value<.05) in sorted phagocytes relative to non-phagocytes (phagocytes: “low”). Many of these transcripts are enriched in medial or lateral LCM transcriptomes. **(F)** Examples of gene expression patterns for transcripts in (E). A majority of these transcripts are enriched in goblet or basal cells, sometimes in medial or lateral subpopulations. Detailed gene ID information is available in Table S1 and in Results. Scale bars, 200 μm.

Figure 5. Double fluorescence in situ hybridization reveals three major cell types in the planarian intestine. (A) Confocal images of *npc2* (green) and *ctsla* (magenta) in situ hybridization. Left to right, whole animal (yellow box indicates magnified region in right panels), zoomed area in green, magenta, and merge (yellow arrow indicates profile line in right-most panel), and a graph showing pixel intensity in each color from tail to head of the yellow profile arrow. *ctsla* is the top phagocyte-specific gene in the phagocyte microarray dataset, while *npc2* is enriched in goblet cells. **(B)** *slc22a6* (green) and *ctsla* (magenta). *slc22a6* mRNA is restricted to the basal region of the intestine, and shows minimal overlap with the phagocyte marker *ctsla*. The white box represents the cropped region shown below with DAPI labeling nuclei, indicating that these riboprobes label distinct cells. **(C)** *npc2* (green) and *slc22a6* (magenta). *npc2* is enriched in goblet cells while *slc22a6* is enriched in basal cells, with minimal overlapping signal. **(D)** *tmi-4* (green) and *slc22a6* (magenta). *tmi-4* is expressed by visceral muscles, while *slc22a6* is found in basal cells. **(E)** Anti-muscle antibody (6G10, green) and *slc22a6* (magenta). Detailed gene ID information is available in Table S1 and in Results. Scale bars, whole animals 200 μ m; magnified images, 10 μ m, magnified crop of basal cell (B), 2 μ m.

Figure 6. Transcripts expressed by intestinal subpopulations and multiple cell types. (A) Confocal images of *cg7631* (green) and *pi16* (magenta) in situ hybridization. Left to right, whole animal (yellow box indicates magnified region in right panels), zoomed area in green, magenta, and merge (yellow arrow indicates profile line in right-most panel), and a graph showing pixel intensity in each color from tail to head of the yellow profile arrow. *cg7631* is enriched in medial goblet cells, while *pi16* is found in all goblet cells. **(B)** *spint3* (green) and *npc2* (magenta). *spint3* is enriched in the lateral goblet cell population, while *npc2* is expressed by all goblet cells. **(C)** *cg7631* (green) and *spint3* (magenta). *cg7631* is enriched in medial goblet cells; *spint3* is enriched in lateral goblet cells. Only rarely do these two markers label the same cell, indicated with a white arrow. **(D)** *zgc:172053* (green) and *slc22a6* (magenta). *zgc:172053* is enriched in a

subset of basal cells, while *slc22a6* is more ubiquitously enriched in most basal cells. **(E)** *slc22a6* (green) and *fabp7* (magenta). *slc22a6* is a basally enriched gene, while *fabp7* is expressed by both basal cells and more apical cells (phagocytes). Blue arrows indicate apical gene expression where *slc22a6* is absent. **(F)** *ctsla* (green) and *fabp7* (magenta). *ctsla* expression is enriched in phagocytes, while *fabp7* is found in both phagocytes and basal cells. **(G)** *npc2* (green) and *fabp7* (magenta). *npc2* is enriched in goblet cells, and overlaps minimally with *fabp7* in phagocytes and basal cells. Detailed gene ID information is available in Table S1 and in Results. Scale bars, whole animals 200 μm ; magnified images, 10 μm .

Figure 7. *gli-1* and *RREB2* regulate goblet cell numbers. **(A)** In 7-day tail regenerates, *gli-1* knockdown dramatically reduces goblet cells (*npc2*+) at the midline in regenerating intestine (yellow arrow), while *RREB2* knockdown reduces goblet cells in old tissue (yellow arrow). Phagocytes (*ctsla*+) appear normal in all conditions. **(B)** Basal cells (*slc22a*+) are unaffected in *gli-1(RNAi)* and *RREB2(RNAi)* regenerates, while goblet cells are reduced similar to A. **(C)** In uninjured animals, *gli-1* RNAi causes moderate goblet cell loss, while *RREB2* RNAi results in severe goblet cell loss. **(D)** Phenotypes in *gli-1(RNAi)* and *RREB2(RNAi)* planarians during 6 dsRNA feedings (once per week). Animals refuse food and undergo lysis and death with increasing frequency over the RNAi time course. Total sample size was $n \geq 55$ for each condition; data were pooled from 3 independent biological replicates of $n \geq 18$ each. **(E)** DAPI labeling of tail regenerates shown in B. White arrows indicate normal regeneration of new brain and pharynx. Animals in A, B, and E were fed dsRNA 8 times (twice per week), starved 7 days, amputated, then fixed 7 days later. Animals in C and D were fed dsRNA 6 times (once per week), starved 7 days, then fixed for FISH. Detailed gene ID information is available in Table S1 and in Results. Scale bars, 200 μm .

Supplementary Material Legends

Figure S1. Optimization of fixation and histological staining for laser microdissection. (A) RNA yields (left table) and Bioanalyzer/TapeStation analysis of RNA integrity (right panels) from whole planarian samples (15 planarians, ~20-25 mg total tissue) treated as shown. As in some other invertebrates (54-57), heat denaturation prior to analysis causes 28S rRNA to co-migrate as two bands with 18S rRNA (indicated with red arrows). Samples are representative of at least three independent replicates. RNA from FA-fixed samples had low yields and low 260/280 ratios. **(B)** Analysis of degraded RNA from FA-fixed samples using high-sensitivity (HS) TapeStation kit (for low concentration samples). Degradation of rRNA is indicated with red arrows. **(C)** Hematoxylin and Eosin Y labeling of transverse (cross) cryosections from methacarn-fixed planarians. Yellow arrows indicate intestine labeling in magnesium-treated samples. Orange arrows indicate compromised intestinal morphology in HCl- and NAc-treated samples. Boxed regions for Eosin Y-labeled sections are magnified to the right. **(D)** RNA yields (left table) and HS TapeStation analysis of RNA integrity (right panels) for all LCM samples used in this study (BR, biological replicates, 16-20 tissue sections per replicate). TapeStation lanes (middle panel) were assembled from multiple runs. Intensity plots (right panel) for BR1 are representative of other replicates. Mg, MgCl₂ treatment. HCl, 2% HCl treatment. NAc, 7.5% N-Acetyl-L-Cysteine treatment. FA, 4% formaldehyde fixation. MCN, methacarn fixation. Scale bars: 100 μ m (C).

Figure S2. Expression of transcripts enriched in laser captured intestinal tissue.

Whole-mount in situ hybridizations performed during this study, organized by mediolateral ratio. 144/162 total in situs conducted are shown (18 omissions were due to very low or no detectable expression; *helz2/dd_12219* was detected only in peripharyngeal cells). Each image also shows

the phagocyte enrichment value (Log_2FC in Table S5), as well as the Uniprot best hit ID (Table S1) and the Dresden v6 Transcriptome GeneID (dd_Smed_v6). Several transcripts are named to be consistent with previous publications (*hnf-4/dd_1694*, *PTF1A/dd_6869*, *gli-1/dd_7470*, *slc22a6/dd_1159*), with human Uniprot IDs used in our comparison to the Human Protein Atlas (*sult1c4/dd_7766*, *tdo2/dd_3550*), or to distinguish paralogs (*apob-1/dd_636*, *apob-2/dd_194*). Detailed gene ID information is available in Table S1. Scale bars, 200 μm .

Figure S3. Innate immunity-related transcripts enriched in the intestine. (A) Intestine-enriched transcripts detected by LCM (pink) compared to *S. mediterranea* transcripts that were up- or downregulated upon transfer of planarians from continuous flow to static culture in Arnold et al., 2016 (90) (blue). (B) *S. mediterranea* intestine-enriched transcripts detected by LCM (pink) compared to *D. japonica* homologous transcripts that were upregulated in response to *L. pneumophila* or *S. aureus* infections in Abnave et al., 2014 (91) (blue).

Figure S4. Validation of cell-type specific mRNA expression by WISH and correlation with single cell analysis. (A-C) Transcripts with WISH patterns indicating enrichment in phagocytes (A), goblet-like cells (B), or basal cells (C), mapped onto plots of phagocyte (y-axis) vs. LCM medial/lateral (x-axis) fold-change expression (log_2FC). (D-F) Transcripts found to be highly enriched in phagocytes (“enterocytes”) (D), goblet cells (E), or basal cells (“outer intestinal cells”) in Fincher et al., 2018 (32), mapped onto phagocyte vs. LCM intestine expression plots. (G-H) Transcripts found to be highly enriched in phagocytes (G) or goblet cells (H) in Plass et al., 2018 (43), mapped onto phagocyte vs. LCM intestine expression plots (basal cells were not described). (I) Venn Diagram showing overlap between intestine-enriched transcripts in three recent single cell studies (32, 43, 88) and our LCM data. Overlaps of two or fewer genes are not displayed. Examples of 23/28 transcripts detected by LCM whose intestinal expression was

validated by WISH. Detailed gene ID information is available in Table S1 and the Results. Scale bars, 200 μ m.

Figure S5. Additional cell-type specific transcripts revealed by double fluorescent in situ hybridization. (A) Phagocyte-enriched genes *ctsl.1* and *apob-2* (green), in combination with a phagocyte, goblet, or basal marker shown in Fig 5 (magenta). Yellow box indicates magnified region. White arrows indicate regions of considerable overlapping expression, while yellow arrows indicate regions of minimal overlap. (B) *si:dkey-241* (green) is enriched in a subset of *slc22a6*-positive basal cells. (C) *gaa* (green) is expressed in both goblet cells (high) and phagocytes (moderate), but is largely absent in basal cells. (D) *oys* is expressed in both phagocytes and basal cells. Detailed gene ID information is available in Table S1. Scale bars, whole animals 200 μ m; magnified images, 10 μ m.

Figure S6. Expression of intestine-enriched transcription factors. Fluorescent in situ hybridization expression patterns of 22 intestine-enriched transcription factors identified in the laser capture dataset. Organized by dd_Smed_v6 ID in ascending order. Detailed gene ID information is available in Table S1. Scale bars, 100 μ m.

Figure S7. Goblet cells are reduced in *gli-1* and *RREB2* 7-day regenerates. Yellow arrows indicate regions of reduced or missing goblet cells. (A) Expression of a second, additional pan-goblet cell marker, *pi16*, in 7-day tail regenerates, which also shows dramatic reduction in goblet cell numbers (arrows). (B) *npc2* expression in 7-day regenerate heads, indicating severe reduction of goblet cells in both *gli-1* and *RREB2* knockdowns (arrows). (C) Trunk fragments also show goblet cell loss. *gli-1* RNAi causes missing or reduced *npc2* expression in blastema

regions, while *RREB2* reduces *npc2* labeling in pre-existing intestine (arrows). Detailed gene ID information is available in Table S1 and in Results. Scale bars, 200 μ m.

Figure S8. Effects of *gli-1* and *RREB2* knockdown on medial and lateral goblet cell

subpopulations. (A) Uninjured animals showing the expression of *cg7631* (medial goblet) and *spint3* (lateral goblet) in control, *gli-1*, or *RREB2* knockdown animals. *gli-1(RNAi)* animals preferentially lose lateral goblet cells. *RREB2(RNAi)* animals lose both goblet cell subtypes, although more lateral cells remain compared to *gli-1* knockdown. **(B)** 7-day tail regenerates with medial and lateral goblet cells labeled in control, *gli-1*, or *RREB2* knockdowns. *gli-1* RNAi again reduces lateral goblet cells throughout the fragment, as well as medial populations in the anterior regenerating intestine, while *RREB2(RNAi)* regenerates lack all goblet cells in pre-injury intestine (e.g., posterior in tail fragments). Detailed gene ID information is available in Table S1. Scale bars, 200 μ m.

Figure S9. Expression patterns of *gli-1* and *RREB2*. (A) Global expression of *gli-1* in

uninjured animals, along with images of mRNA expression in phagocytes (*ctsla*), basal cells (*slc22a6*), goblet cells (*npc2*), and muscle cells (*tni-4*). **(B)** Expression of *RREB2* in uninjured animals, along with mRNA co-localization with the same phagocyte, basal, goblet, and muscle cell markers as in A. Detailed gene ID information is available in Table S1. Scale bars, whole animals and tails 200 μ m magnified images, 10 μ m; digital zooms, 5 μ m.

Figure S10. Additional genes whose knockdown affects goblet cell numbers in 7-day

regenerates. (A) Control RNAi regenerates with normal numbers of goblet cells. **(B)** *Lhx2b*

RNAi reduces numbers of goblet cells in pre-existing tissue, but not in the blastema. **(C)** *PTF1A*

RNAi mildly reduces levels of lateral goblet cells. All in situ hybridizations to detect *npc2*-positive goblet cells (green). Detailed gene ID information is available in Table S1. Scale bars, 200 μ m.

Figure S11. Phenotypes, area, and length of *gli-1* and *RREB2* knockdowns. (A) Example images of viability phenotypes observed during dsRNA feeding. **(B)** Area measurements immediately before the first dsRNA feeding and seven days after the fifth dsRNA feedings (one feeding per week). No significant differences were observed between pre-treatment area sizes (n=60 each), and no difference is observed between 5x fed control (n=57) and 5x fed *gli-1* (n=41, p-val = 0.4174), but a small difference is observed between 5x fed control and 5x fed *RREB2* (n=36, p-val = 0.0258). Non-eating animals and dead animals were not included in size analysis. **(C)** Length measurements from the same animals in **(B)** before and after dsRNA feeding. No significant differences between the control and the experiments were observed (5x fed control vs 5x fed *gli-1* p-value = 0.5010, 5x fed control vs. 5x fed *RREB2* p-value = 0.0769). Thick dashed line indicates the median, the thin dashed lines indicate the upper and lower quartiles.

Figure S12. Gene expression by goblet cells suggests multiple physiological roles. (A) Whole-mount in situ hybridizations for *carboxypeptidase A2* (*cpa2*), *gastric triacylglycerol lipase* (*lipf*), *kallikrein-13* (*klk13*), and *peptidoglycan recognition protein LB* (*pgrp-lb*). **(B)** Neuroendocrine convertase 1 (*pcsk1*) mRNA expression. WISH, top left. dFISH, top right and magnified views (yellow box). *pcsk1* (green) co-expression with the pan-goblet marker *npc2* (magenta) demonstrates *pcsk1* is expressed in goblet cells as well as peripharyngeal secretory cells. **(C)** Expression of three *mucin*-like genes in secretory cells surrounding the pharynx (*muc-like-1*/dd_Smed_v6_17988_0_1, *muc-like-2*/dd_18786_0_1, and *muc-like-3*/dd_21309_0_1), and mAb 2C11 labeling peripharyngeal secretory cells and their projections into the pharynx

(magenta) (bottom row). Detailed gene ID information is available in Table S1. Scale bars:
whole animals 200 μm ; magnified images (B), 10 μm .

Figure S13. Biological Process GO terms enriched among medially and laterally enriched transcripts. (A) Pie chart showing the ratio of transcripts with higher fold changes in medial and lateral intestinal tissue. **(B)** Venn diagram showing the number of Biological Process GO terms enriched in medial, lateral, or both groups of transcripts. **(C)** Examples of GO terms over-represented in medial and lateral transcripts.

Table S1. Transcripts enriched in the planarian intestine.

Table S2. Gene ontology biological process term enrichment for intestinal transcripts.

Table S3. Intestine-enriched transcripts with potential roles in innate immunity.

Table S4. Planarian transcripts with reciprocal best hit orthologs enriched in human tissues.

Table S5. Comparison of transcripts enriched in laser-captured intestine and sorted intestinal phagocytes.

Table S6. RNA interference results for intestine-enriched transcription factors and mediolaterally enriched transcripts.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure S1

A RNA yields and integrity - whole planarians

sample	yield	OD 260/280
live	33.5 µg	2.09
Mg only	30.9 µg	2.10
HCl only	18.3 µg	2.13
NAc only	7.0 µg	2.00
FA only	1.0 µg	1.33
MCN only	32.8 µg	2.15
Mg+FA	0.85 µg	1.36
Mg+MCN	28.6 µg	2.03
HCl+FA	0.63 µg	1.21
HCl+MCN	42.0 µg	2.20
NAc+FA	0.58 µg	1.22
NAc+MCN	34.4 µg	2.18

B RNA integrity - whole planarians (high sensitivity analysis of FA samples)

C Morphology and Staining Optimization

D RNA yields and integrity - LCM samples

sample	yield	OD 260/280
BR1 - medial intestine	42.5 ng	2.04
BR1 - lateral intestine	133.1 ng	1.67
BR1 - non-intestine	151.4 ng	2.35
BR2 - medial intestine	395.8 ng	1.25
BR2 - lateral intestine	126.5 ng	1.19
BR2 - non-intestine	148.25 ng	1.46
BR3 - medial intestine	61.5 ng	1.97
BR3 - lateral intestine	156.9 ng	1.53
BR3 - non-intestine	126 ng	2.02
BR4 - medial intestine	101.25 ng	2.00
BR4 - lateral intestine	186 ng	1.72
BR4 - non-intestine	241.5 ng	1.87

Figure S2

WISH-validated genes sorted by Medial/Lateral Ratio

Figure S3

Figure S4

Figure S5

Figure S6

Figure S7

Figure S8

Figure S9

A

B

Figure S10

Figure S11

Figure S12

Figure S13