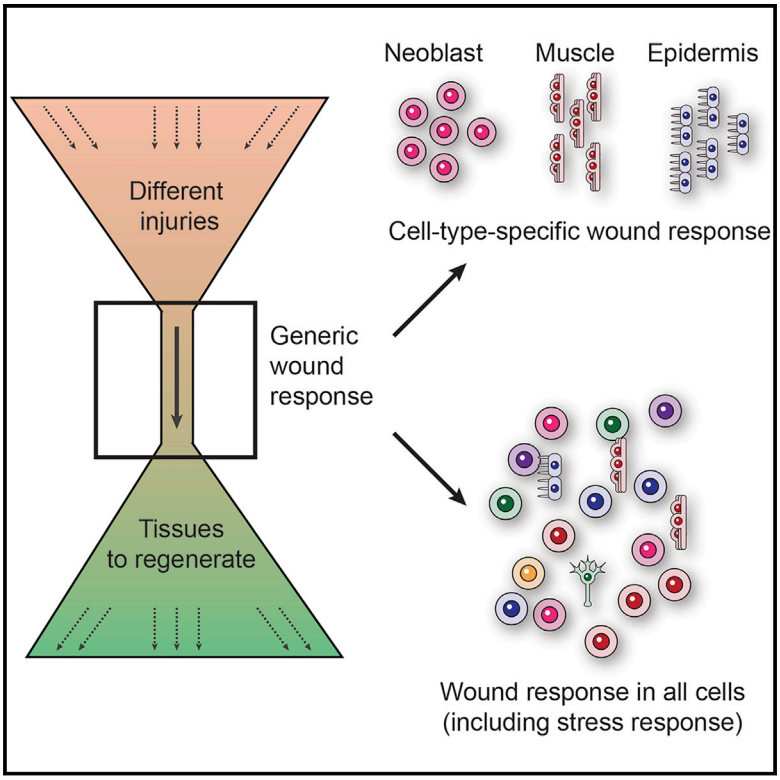
Resource

Developmental Cell

A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians

Graphical Abstract



Authors

Omri Wurtzel, Lauren E. Cote, Amber Poirier, Rahul Satija, Aviv Regev, Peter W. Reddien

Correspondence

reddien@wi.mit.edu

In Brief

A resource characterizing major planarian cell-type transcriptomes identifies 1,214 tissue-specific markers across 13 cell types. Single-cell RNA sequencing showed that wound-induced genes were expressed in nearly all cell types or specifically in one of three cell types. A generic wound response is activated with any injury regardless of regenerative outcome.

Highlights

- Injury activates a common wound-response transcriptional program
- Muscle, epidermis, and stem cells express most wound-induced genes
- A single gene, *notum*, is differentially induced at head- versus tail-facing injuries
- Injury-specific transcriptional changes emerge 24 hr after response initiation

Wurtzel et al., 2015, Developmental Cell *35*, 632–645
December 7, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.devcel.2015.11.004

CellPress

Developmental Cell
Resource

A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians

Omri Wurtzel,[1,2] Lauren E. Cote,[1,2] Amber Poirier,[1,2] Rahul Satija,[3,4,5] Aviv Regev,[2,3] and Peter W. Reddien[1,2,*]
[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]New York Genome Center, New York, NY 10013, USA
[5]Department of Biology, New York University, New York, NY 10003, USA
*Correspondence: reddien@wi.mit.edu
http://dx.doi.org/10.1016/j.devcel.2015.11.004

SUMMARY

Regeneration starts with injury. Yet how injuries affect gene expression in different cell types and how distinct injuries differ in gene expression remain unclear. We defined the transcriptomes of major cell types of planarians—flatworms that regenerate from nearly any injury—and identified 1,214 tissue-specific markers across 13 cell types. RNA sequencing on 619 single cells revealed that wound-induced genes were expressed either in nearly all cell types or specifically in one of three cell types (stem cells, muscle, or epidermis). Time course experiments following different injuries indicated that a generic wound response is activated with any injury regardless of the regenerative outcome. Only one gene, *notum*, was differentially expressed early between anterior- and posterior-facing wounds. Injury-specific transcriptional responses emerged 30 hr after injury, involving context-dependent patterning and stem-cell-specialization genes. The regenerative requirement of every injury is different; however, our work demonstrates that all injuries start with a common transcriptional response.

INTRODUCTION

Wounding leads to a series of complex responses that are necessary for recovery (Gurtner et al., 2008). Recent studies in regenerative organisms, including planarians (Wenemoser et al., 2012), sea anemones (DuBuc et al., 2014), hydra (Lengfeld et al., 2009), and axolotls (Knapp et al., 2013), have demonstrated that wounding broadly affects gene expression, including the activation of stress-response genes, tissue-patterning factors, matrix metalloproteinases, and growth factors. However, the functions of the vast majority of genes that are induced following injury remain unknown (DuBuc et al., 2014; Wenemoser et al., 2012).

Planarians are free-living flatworms with a remarkable regenerative capacity that is mediated by tissue-resident proliferative cells (neoblasts) that include pluripotent cells (Reddien and Sánchez Alvarado, 2004; Wagner et al., 2011). Following wounding, rapid gene expression changes are observed in both neoblasts and differentiated tissues (Wenemoser et al., 2012). A number of genes were shown to be activated in different wound types (Adell et al., 2009; Petersen and Reddien, 2009; Wenemoser et al., 2012), raising the possibility that a common transcriptional wound response precedes regeneration (Wenemoser et al., 2012). In contrast, it has been recently proposed that different injuries activate distinct transcriptional programs that subsequently converge to similar transcriptional programs later in regeneration (Kao et al., 2013). Determining whether wounds that will regenerate different anatomy begin with similar, identical, or very different transcriptional responses remains a central problem in understanding regeneration.

Some wound-induced genes, such as *HSP90* and *HSP70*, are associated with general stress response, but others, such as *follistatin*, are critical for initiating regeneration (Gaviño et al., 2013). In contrast, some wound-induced genes have known functions only in particular injuries. For example, wound-induced *wnt1* expression has a known role in tail but not head regeneration (Adell et al., 2009; Petersen and Reddien, 2009), despite its induction in both wound types (Petersen and Reddien, 2009).

Multiple key questions about wound responses and how they associate with regeneration of different body parts remain unresolved. First, how does the transcriptional response to wounding map onto the different cell types at the site of injury? Second, how does the transcriptional response to injury differ depending on the injury type and the eventual regenerative outcome? Finally, which transcriptional changes are specific to the regeneration of particular anatomical structures, and when do these changes appear?

We addressed these key questions by combining multiple experimental and computational approaches. We applied single-cell RNA sequencing (SCS) to 619 individual planarian cells and determined the transcriptomes of 13 distinct cell types, including all major planarian tissues, leading to the identification of 1,214 unique tissue markers. SCS from injured animals associated 49 wound-induced genes with the cell types that expressed them, revealing that major wound-induced gene classes were expressed either in nearly all cell types at the wound or specifically in one of three cell types (neoblast, muscle, or epidermis). Time course experiments on bulk RNA from injuries leading to distinct regenerative outcomes determined that a

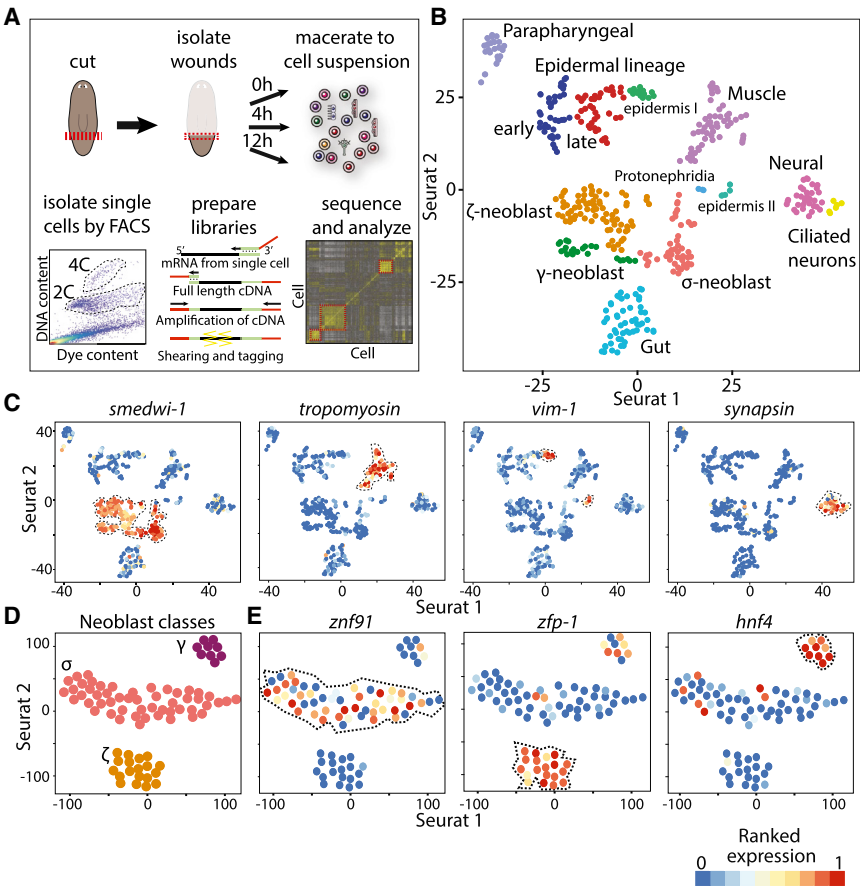


Figure 1. Unbiased Detection of Major Planarian Cell Types by SCS

(A) Illustration of SCS data generation and analysis. Animals were cut postpharyngeally (red line), and wound sites (red box) were isolated at 3 time points. Wound tissue was macerated, and dividing (4C) and non-dividing (2C) cells were isolated by FACS (Experimental Procedures; dashed line shows gates). Sequencing libraries were prepared by cDNA amplification and shearing, and libraries were sequenced and analyzed.

(B) t-SNE plot of single cells. Cells (colored dots) are grouped by density clustering and labeled on the basis of marker analysis. Cells shown are from the 2C (wounded and unwounded) and 4C (wounded) fractions.

(C) Expression of canonical cell-type markers overlaid on t-SNE plots of the single-cells (dots); low- and high-ranked expression are colored by a gradient of blue, yellow, and red.

(D) Analysis of the neoblast compartment. Shown are neoblasts (dots) from uninjured animals. Clusters are annotated on the basis of multiple neoblast markers.

(E) Expression of class-specific neoblast markers. See also Figures S1 and S2.

single conserved transcriptional program was activated at essentially all wounds, except for the differential activation of a single gene, *notum*. Over 24 hr following the peak of this generic wound response, specialized transcriptional programs emerged, specific for the body parts requiring regeneration. Our results define a generic and conserved response to wounding, identify the cell types that drive it, and describe the subsequent transcriptional changes leading to regeneration.

RESULTS

Single-Sell Sequencing of Planarian Cells

To dissect how different cell types transcriptionally respond to injuries, we used SCS, because it profiles the transcriptional responses of a cell and allows its cell type classification (Jaitin et al., 2014; Shalek et al., 2014). We isolated cells by fluorescence-activated cell sorting (FACS) (Figure 1A) from postpharyngeal wound sites that were collected from animals immediately following amputation or after a recovery period (4 or 12 hr post injury [hpi]; Experimental Procedures). In total, we sequenced RNA from 214 dividing neoblasts and 405 non-dividing cells (Table S1) and measured their gene expression by mapping the sequencing reads to the planarian transcriptome (Liu et al., 2013). On average, we detected the expression of 4,401 genes per cell (Figure S1A), with more than 91% of the cells expressing more than 1,000 genes (Supplemental Experimental Procedures).

We assessed the SCS data quality by comparing the expression of canonical neoblast markers (Guo et al., 2006; Reddien et al., 2005; Shibata et al., 1999) between sorted neoblasts and non-dividing cells. Neoblasts had a striking enrichment for these transcripts ($p < 1 \times 10^{-75}$; Figure S1B). For example, *smedwi-1* and *bruli* were overexpressed in neoblasts 217- and 140-fold, respectively, highlighting the expression data specificity.

Unbiased Assignment of Planarian Cells to Putative Cell Types

To define the cell types present at wounds, cells were clustered and analyzed according to their gene expression (Figure S1C). Initially, genes with high variance across cells were selected (dispersion $\geq$ 1.5; Figures S1D–S1F; Experimental Procedures), because their expression levels can partition cells to groups (Jaitin et al., 2014; Shalek et al., 2013). Next, we used these genes as input for the recently published Seurat algorithm (Macosko et al., 2015; Satija et al., 2015) that extends the list of genes used for clustering by finding genes with significant expression structure across principal components (Supplemental Experimental Procedures; Figure S1G). Then, cells were embedded and visualized in a 2D space by applying t-distributed stochastic neighbor embedding (t-SNE) on the genes selected by Seurat (Figure 1B; Experimental Procedures). Finally, clusters were defined by applying density clustering (Ester et al., 1996) on the 2D embedded cells. Importantly, the time point at which cells were isolated did not affect cluster assignments (Table S1), indicating that the identity of a cell had a stronger impact on cluster assignment than did transcriptional responses to wounding. This process revealed

13 cell clusters (Figure 1B), which likely represented different major planarian cell types.

Detection of the Major Planarian Cell Types

Multiple approaches were used to assign cell type identity to the clusters and to test whether cells in a cluster were of the same type. First, we plotted the expression of published cell-type-specific markers on the t-SNE plots (Figure 1C) and found that canonical tissue markers for major cell types were found exclusively in distinct clusters. This was highly suggestive of cluster identity for cell types, such as neoblast (Reddien et al., 2005), muscle (Witchley et al., 2013), neurons (Sánchez Alvarado et al., 2002), and epidermis (van Wolfswinkel et al., 2014).

Second, we identified cluster-specific genes by using a binary classifier (Sing et al., 2005) that quantified the ability of individual genes to partition cells assigned to one cluster from all other clusters by measuring the area under the curve (AUC) in a receiver operating characteristic (ROC) curve (Figure S1H; Experimental Procedures). Similarly, we searched for markers that were expressed in multiple clusters displaying expression of the same canonical markers (e.g., *smedwi-1* or *synapsin*; Figure 1C; Experimental Procedures).

In total, 1,214 genes (false discovery rate [FDR] < 0.1) were highly specific for a cluster or shared between cluster groups (Table S2). We used the multiple published anatomical markers found in this gene set to determine cluster identity for the following cell types: muscle (Witchley et al., 2013), gut (Forsthoefel et al., 2011), epidermis (van Wolfswinkel et al., 2014), early epidermal progenitors (*prog-1*) (Pearson and Sánchez Alvarado, 2010), late epidermal progenitors (*agat-1*) (Eisenhoffer et al., 2008; van Wolfswinkel et al., 2014), neoblasts including specialized neoblasts (Scimone et al., 2014; van Wolfswinkel et al., 2014), protonephridia (Scimone et al., 2011), and two neuronal types (Cowles et al., 2013; Sánchez Alvarado et al., 2002) (Figures 1B–1E and S2; Table S2).

Finally, a single cluster was unique in lacking enriched expression of genes with published expression patterns. Whole-mount in situ hybridization (WISH) using RNA probes on four of its top cluster-specific genes (*Rab-11B*, *myoferlin*, *ESRP-1*, and *anoctamin*) revealed strong parapharyngeal expression with a ventral anatomical bias (Figure S2A; Experimental Procedures). Double fluorescent in situ hybridization (dFISH) (Figure S2B) validated that single cells in the parapharyngeal region co-expressed these genes, indicating that this was indeed a cell type lacking prior molecular definition.

The clustering analysis we performed allowed detection of subpopulations of cells that appeared largely homogeneous when examined only with canonical markers. For example, two adjacent clusters (Figure 1B) were determined to be neural on the basis of specific expression of canonical neural markers, including *synapsin*, *synaptotagmin*, and *prohormone convertase 2* (*PC2*) (Figures 1C and S2D). However, one of these clusters co-expressed genes encoding known cilia components, such as *bbs1*, *bbs9* (Figure S2D), *ift88*, and *iguana* (Glazer et al., 2010), suggesting that these might be neurons with sensory cilia (Louvi and Grove, 2011). The only other cell type expressing these cilia genes was the epidermis (Figure S2D).

In the neoblast compartment, we detected three subpopulations representing the recently described σ-, ζ-, and γ-type neoblasts (van Wolfswinkel et al., 2014) (Figures 1D and 1E) and revealing multiple putative markers unique to each subpopulation (Table S2; Figure 1E), such as *znf91*, a previously undescribed gene encoding a zinc finger protein showing the highest specificity to the σ-neoblasts (AUC = 0.81, FDR = 2.6×10^{-5}) (Figure 1E; Table S2).

Importantly, the dissection of planarian cell types and their associated gene expression generated an extensive repository of cell-type-specific markers for every major cell type, including signaling molecules, receptors, and transcription factors, as well as profiles of their co-expression (available at https://radiant.wi.mit.edu/app/).

Identification of Cell-Type-Specific Wound-Induced Genes

Knowing which cell types express particular wound-induced genes is important for understanding how the wound response differs across injuries with different anatomy. However, the cell-type specificity of only a small number of wound-induced genes is known (Wenemoser et al., 2012; Witchley et al., 2013).

Because SCS data are often noisy and incomplete (Jaitin et al., 2014), we first defined a comprehensive list of wound-induced genes by RNA sequencing (RNA-seq) of bulk samples from two different injury types. We profiled the expression of anterior-facing (head removal) and posterior-facing (trunk and tail removal) wounds in the prepharyngeal region (Figure 2A) by isolating RNA, in triplicate, at four time points (0, 3, 6, and 12 hpi) (Figure 2A; Experimental Procedures).

The bulk sequencing data revealed that 128 genes were overexpressed in at least one time point compared with the 0 hpi (uninjured) samples, in at least one of the two wound types (fold change [FC] $\geq$ 2, FDR $\leq$ 0.05) (Figure 2A; Table S3; Experimental Procedures). To determine what cell types participated in the wound response, we compared the SCS expression of the 128 wound-induced genes (1) between cells isolated from uninjured animals and injured animals and (2) between different cell types using only cells isolated following wounding (Figures 2B and 2C; Experimental Procedures). In total, we detected the cell-type specificity of 49 of the 128 genes (38%). Ten of these genes were wound induced in nearly all cell types (Figure 2), with 6 of them annotated as general stress response factors, including *heat-shock protein 90* (*HSP90*), *HSP70*, and *HSP40* (Experimental Procedures). Only one of the genes encoded a transcription factor, *egr-2* (Figures 2B and 2C).

Strikingly, most of the cell-type-specific genes (35 of 49 [71%]; Figure 2D) were wound induced in one of three cell types. Sixteen genes were enriched in neoblasts, including genes related to proliferation (e.g., *H2B*, *topbp1*, *rrm2b*) and neural regeneration (*runt-1*, known to be induced in neoblasts) (Sandmann et al., 2011; Wenemoser et al., 2012). In muscle cells, we found enrichment for 14 wound-induced genes, including 5 genes that were implicated in major signaling pathways, including Wnt, BMP, and TGF-β, which are essential for proper patterning of planarian tissues (Reddien, 2011; Witchley et al., 2013). Importantly, because the number of muscle cells sequenced was smaller than the numbers of many other cell types (e.g., the number of gut cells was almost twice the number of muscle cells), these results cannot be explained by an increased statistical power resulting from larger sample size.

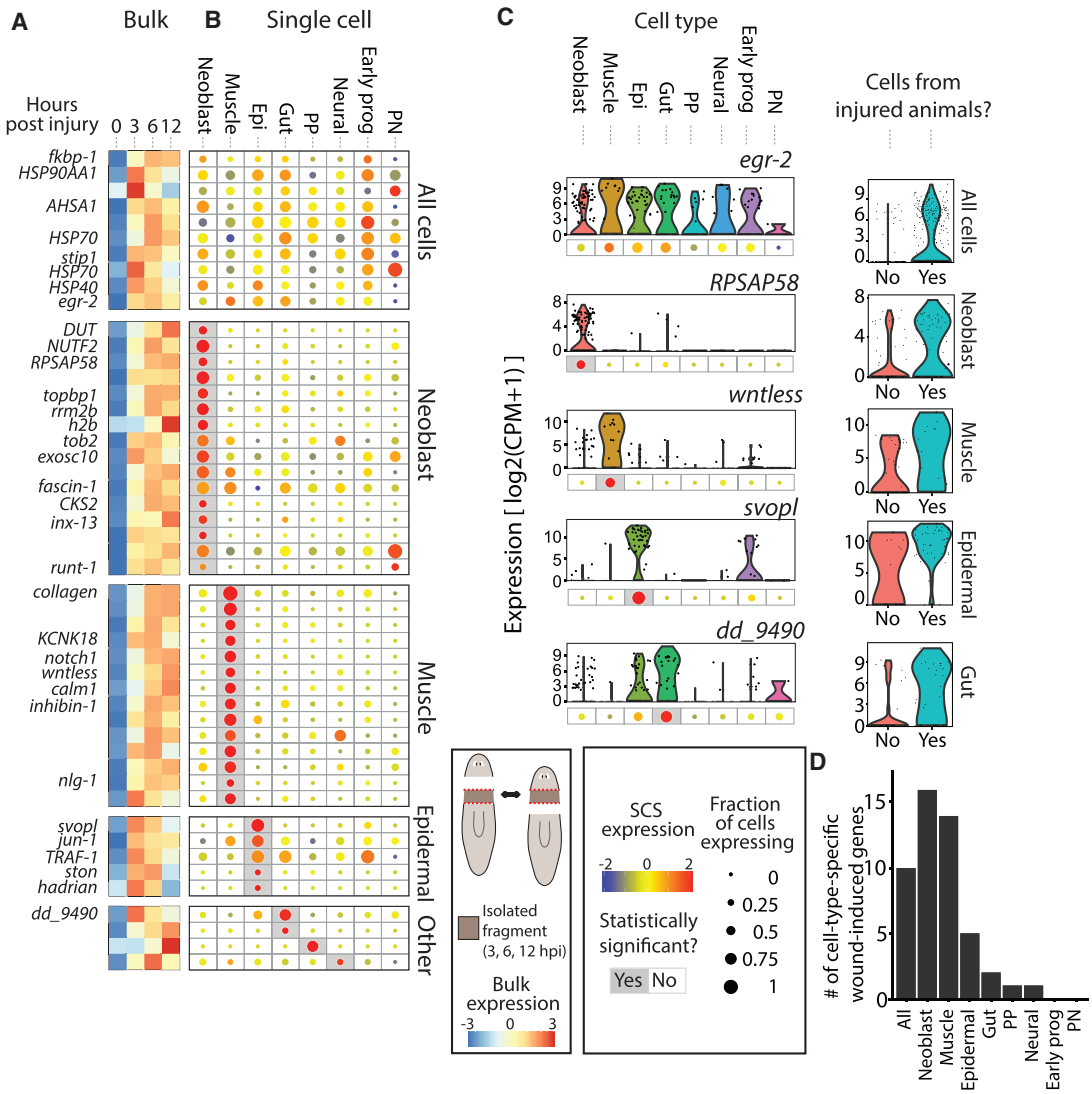


Figure 2. Cell-Type-Specific Expression of Wound-Induced Genes

(A) The expression of wound-induced genes, as detected by bulk RNA-seq, is shown at different time points (0, 3, 6, and 12 hpi). Shown is the average expression of the anterior- and posterior-facing time courses. Rows and columns represent genes and time points, respectively. Gene expression is colored according to the z-transformed expression (z score range is −3 to 3). Shown are wound-induced genes for which cell-type specificity was determined.

(B) The corresponding cell-type-specific gene expression is shown in a dot-plot map. Dot size represents the proportion of cells expressing the gene (see key; 0–1), and the color represents normalized expression in cells expressing the gene (blue to red, low to high expression). Gray background represents statistically significant enrichment in a cell type (FDR ≤ 0.01; Supplemental Experimental Procedures). Genes are ordered according to their controlled enrichment p values. Genes assigned to "All cells" were overexpressed following wounding in multiple cell types (Experimental Procedures). Early prog, early epidermal progenitors; Epi, epidermal lineage; NB, neoblasts; PN, protonephridia; PP, parapharyngeal.

(C) Left: representative genes with wound-induced expression in different cell types. Expression across cell types is shown in violin plots with corresponding dot plots beneath. Right: violin plots comparing the expression in cells of the cell type the gene was found to be enriched in between uninjured and injured animals.

(D) Summary of the detected cell-type-specific wound-induced genes.

Finally, 5 genes were enriched in epidermal lineage cells, including *Smed-jun-1* (Wenemoser et al., 2012). In addition, a small number of genes (1 or 2) were wound induced in three other cell types: gut, parapharyngeal (Figure S3A), and neural cells.

Our results are supported by two recent studies that examined the co-expression of several wound-induced genes with cell-type-specific markers. *nlg1*, *inhibin-1*, and *wntless* were found to be specifically wound-induced in muscle cells of injured animals (Witchley et al., 2013), whereas *jun-1*, *TRAF-1*, *ston*, and *hadrian* were found to be localized to the epidermis (Wenemoser et al., 2012).

We used multiple approaches to validate our results. First, we examined the co-localization of three candidates (*svopl*, *dd_9519*, and *Tob2*) with published cell-type markers. dFISH analysis found, in all cases, high specificity of expression to the identified cell type in the single-cell analysis (Figure 3A).

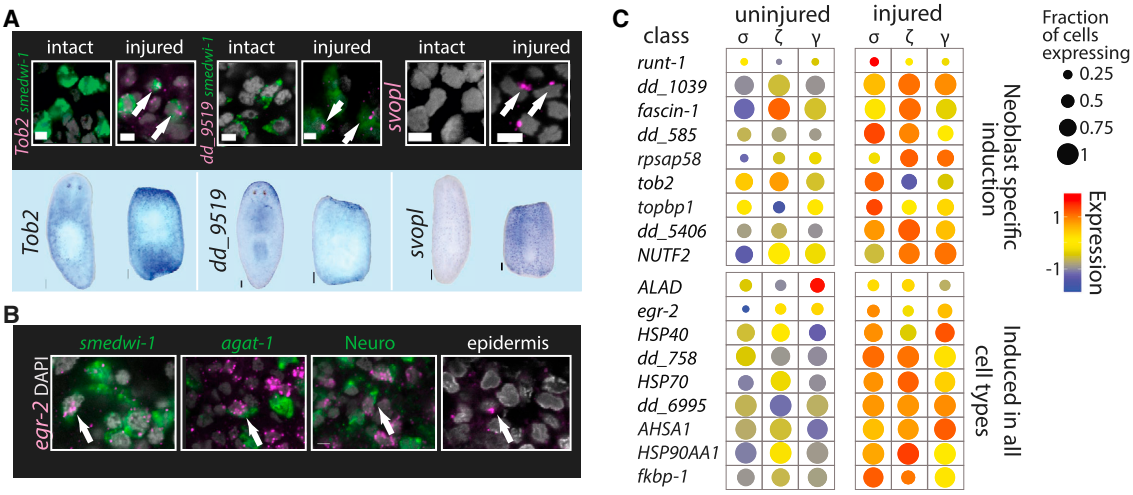


Figure 3. Analysis of Cell-Type-Specific Expression after Injury

(A) Validations of tissue-specific wound-induced genes. Top: dFISH analysis (the scale bars indicate 5 μm) of cell-type-specific wound-induced gene (magenta) and a cell-type marker (green), or imaging of the outermost layer (epidermis). Nuclei were labeled with DAPI (gray). White arrows point to co-expressing cells. Bottom: WISH analysis comparing gene expression in intact and amputated animals (the scale bars represent 100 μm).

(B) dFISH analysis of *egr-2* (magenta) with markers of multiple tissues in animals 12 hpi (green; *smedwi-1*, neoblasts; *agat-1*, epidermal progenitors; Neuro [pooled RNA probes for *PC2*, *synapsin*, *synaptotagmin*], neural tissue; epidermal cells were imaged by the outermost layer of the animals). WISH/FISH analysis was done on at least 15 fragments for each gene.

(C) Gene expression comparison of uninjured and injured neoblasts. Shown are dot plots of neoblast-specific wound-induced genes (top) and genes found to be wound induced in most or all cell types (bottom) in the different neoblast classes. Dot size represents the fraction of expressing cells (0–1); color represents the expression levels (z score) in the fraction of expressing cells.

Furthermore, we tested whether *egr-2* was indeed wound induced in multiple cell types (Figure 2) and found that it was co-localized with markers for neoblasts (*smedwi-1*), epidermal progenitors (*agat-1*), neural cells (*PC2*, *synapsin*, and *synaptotagmin*), and differentiated epidermis (outermost epidermal layer) (Figure 3B).

Next, we tested whether different neoblast subpopulations (van Wolfswinkel et al., 2014) responded differently to wounding (Figure 3C). We compared the gene expression of neoblasts representing the general neoblast pool (σ), the epidermal progenitors (ζ), and the putative gut progenitors (γ) between uninjured and injured animals. Interestingly, although some wound-induced genes were overexpressed in specific populations (e.g., *runt-1* in the σ-neoblasts), most genes changed similarly across neoblast subtypes (Figure 3C).

This analysis demonstrates that the cell-type architecture of the wound response involves (1) genes induced broadly in most or all cell types; (2) multiple genes induced in a cell-type-specific manner in one of three types of cells: neoblast, muscle, or epidermis; and (3) rare individual genes expressed in a specific cell type (gut, parapharyngeal, or neural cells).

A Single Gene, *notum*, Detectably Differentiates between Anterior and Posterior Wound Responses

How similar are the transcriptional responses to distinct injuries? The cell types that express wound-induced genes are widespread across the planarian body and, in principle, could mount a similar transcriptional response at injuries requiring the regeneration of distinct tissues.

However, the extent of similarity in wound responses between distinct injuries is yet to be resolved. To address this question, we searched for wound-induced genes that were enriched at anterior- over posterior-facing wounds, or vice versa, at any of the three time points (3, 6, and 12 hpi) (Experimental Procedures; Figures 2A and 4A; Table S3). Importantly, these two wound types had very similar tissue composition but required distinct regenerative outcomes (Figure 4A).

Of the 128 wound-induced genes, only one gene (*notum*) had a biased expression of more than 2-fold in one of the amputations compared with the other, in at least one time point (Figure 4A). Even with relaxed thresholds (FC $\geq$ 1.5, FDR $\leq$ 0.1), we found that only seven genes were overexpressed at one of the injuries compared with the other (Figure 4A). We tested the expression data predictions by WISH, and strikingly, only *notum* displayed asymmetric expression, with the six other genes having no robust differential expression in anterior and posterior wound sites (Figure 4B). The one true-positive gene, *notum*, is known to be activated at all wounds but to have stronger expression at anterior-facing compared with posterior-facing wounds (Petersen and Reddien, 2011). Importantly, *notum* is essential for establishing correct head-tail regeneration in planarians (Petersen and Reddien, 2011).

We extended this analysis by screening 218 additional genes by WISH; these genes represented a diversity of fold changes for wound induction and genes that were below threshold for significant difference between wound types. All wound-induced genes had similar expression at anterior and posterior-facing injuries (Figure S3B; Tables S3 and S4). These data strongly indicate that following anterior or posterior amputations, the same transcriptional response to wounding is immediately activated, except for higher expression of a single gene, *notum*, at anterior-facing wounds.

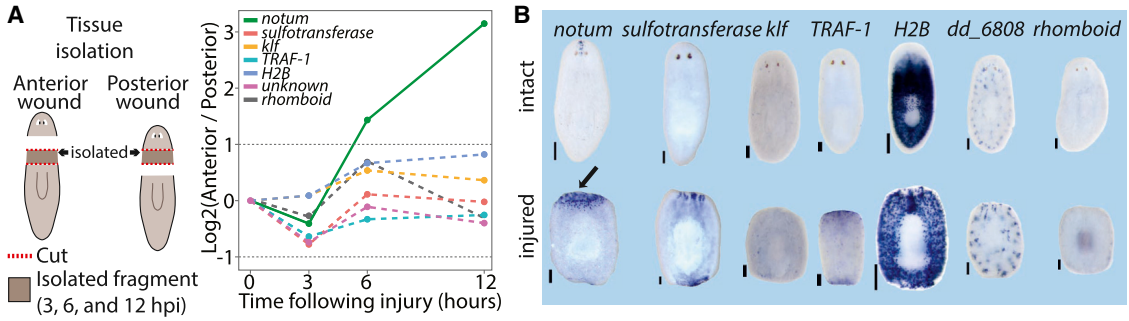


Figure 4. *notum* Is the Only Gene Detectably Induced Asymmetrically at Wounds

(A) The gene expression profiles of injuries with different wound orientation (anterior and posterior; left) are compared in time course experiments of tissues isolated from the same location. Plotted is the $\log_2$ ratio of differentially expressed genes between the two wound types (FDR $\leq$ 0.05, FC $\geq$ 1.5). Dashed lines represent genes that could not be validated by WISH and that are likely false positives.

(B) WISH validations of wound-induced genes shown in (A) (performed on at least ten animals). Top: gene expression in intact animals compared with expression in amputated trunks (bottom). Amputated animals were fixed at the time point showing peak asymmetry in expression. Only *notum* showed asymmetrical expression following wounding (black arrow). The scale bars represent 100 μm.

See also Figure S3.

Comparison of Responses to Diverse Injuries through Extended Time Course Experiments

The striking similarity in the wound response following two amputations types is consistent with the possibility that a generic wound response would be activated following any injury, even when regeneration is not required (Wenemoser et al., 2012). To test this hypothesis, we studied distinct injuries requiring regeneration of different body parts in time courses that span the wound response and extended to subsequent regenerative phases (0–120 hpi) (Figures 5A and S4A; Table S5).

At every time point, we isolated wound sites from the following injuries: (1) postpharyngeal anterior-facing, (2) postpharyngeal posterior-facing, (3) sagittal-anterior, (4) sagittal-posterior, and (5) a lateral incision, which did not require regeneration (Figures 5A and S4A; Experimental Procedures). Gene expression was measured by RNA-seq and compared with uninjured equivalent anatomical regions. In addition, a recently published head regeneration RNA-seq data set was incorporated (Liu et al., 2013).

To test if the same transcriptional response was activated in every injury, a comprehensive collection of wound-induced genes was required. We therefore determined whether the 128-gene list (described above) included the majority of wound-induced genes without detecting an abundance of false positives. WISH was performed on 225 genes (Table S4), which covered a wide range of fold changes and FDRs following wounding. We found that a threshold of FC > 2 balanced sensitivity (57%) with precision (88%). This analysis estimates that the total number of wound-induced genes, detectable with the methods used, is approximately 224 (SD = 27), an appreciably small (~1%) fraction of all planarian genes (Figures S4B–S4E; Table S4; Experimental Procedures).

A Common Response to Wounding Activated Following Diverse Injuries

To test whether a generic transcriptional program is activated at every injury, we evaluated how many of the 128 wound-induced genes were induced within 16 hr following the injuries described above. Eighty-five percent of the genes were overexpressed in at least five time courses (FC > 1.5) (Table S5; Experimental Procedures); fold changes in time courses that did not meet this threshold were often (43%) just below it. We tested by WISH whether the wound-induced genes that did not appear to be overexpressed by RNA-seq in a given time course were indeed not induced by that injury type. In all cases, the genes were actually expressed at the tested injury site (9/9 incisions) (Table S5). Furthermore, we tested 10 additional of the 128 wound-induced genes that appeared to be lowly induced in incisions (2 > FC > 1.5) and 8 genes that appeared to be lowly induced in posterior amputations (2 > FC > 1.5) and found that they were in fact induced in all cases (Table S5). By contrast, tissues far from the injury (Supplemental Experimental Procedures) showed upregulation of a fraction of the wound-induced genes (15%) (Figure S4G), with many of these genes (9 of 23) associated with stress responses.

To further validate that tissue removal was not required for activating the wound-response program, we compared the expression of 35 randomly selected wound-induced genes by WISH in intact, amputated, or incised animals at their time of peak expression (Figure 5B; Table S5; Experimental Procedures). All 35 genes were induced following amputations, and strikingly, 34 of 35 of the genes (97%) were detectably overexpressed following incisions, corroborating the time course experiments (Figure 5A; Table S5). *sulfotransferase*, which was not detectably overexpressed by WISH, was at least 2-fold overexpressed in all RNA-seq time courses. Collectively, these results strongly suggest that a single generic transcriptional program was activated at every injury. This response might include genes that are insignificant for many types of injuries but essential for the recovery from others. Consistent with this possibility, RNAi of only 8 of 62 wound-induced genes displayed a detectable phenotype (Table S3), further suggesting that many wound-induced genes are not essential for survival and recovery after injury.

The Response to Wounding Terminates Earlier When Regeneration Is Not Required

Whereas different injuries activated essentially the same genes, the dynamics of their expression across injuries could be different. We therefore fit the gene expression data to a quantitative model (impulse) (Chechik and Koller, 2009; Sivriver et al.,

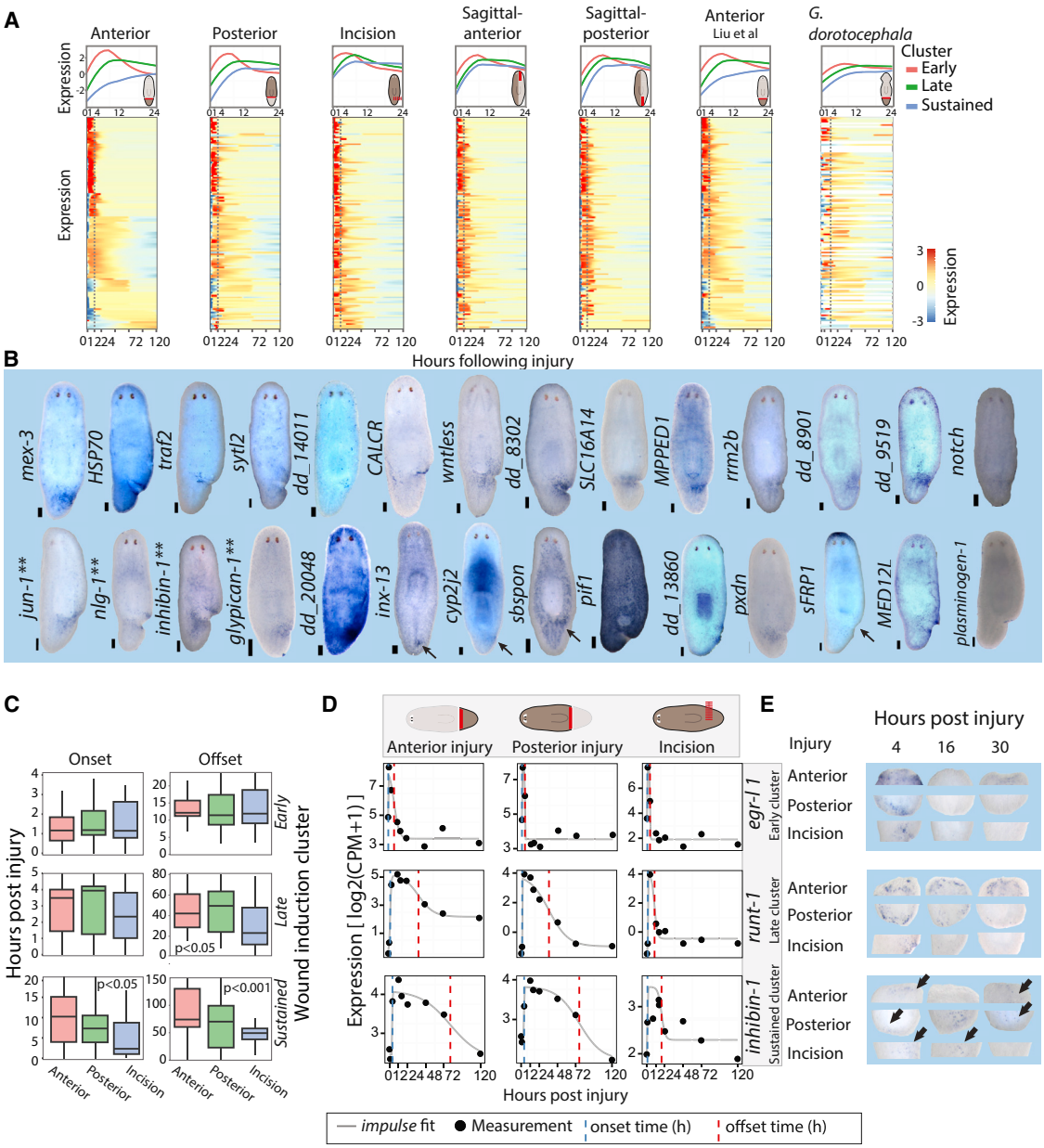


Figure 5. Time Course Analysis Reveals a Generic Response to Wounding

(A) Expression of wound-induced genes at different planarian injuries. A core set of 128 wound-induced genes is plotted in different extended time courses. Worm illustrations show the injury site and isolated tissue location (red block line). Top: the expression of different wound-induced clusters from 0–24 hpi (lines are loess fit of wound-induced gene expression in each cluster; the same genes were used in all panels). Bottom: The expression of the wound-induced genes from 0–120 hpi is shown according to fitting of individual genes to a constrained impulse model (Chechik and Koller, 2009) (shown is row z score; blue to red, low to high expression, respectively). Rightmost column: conservation of the wound response in anteriorly regenerating *G. dorotocephala*. Gene order follows orthology assignment between *G. dorotocephala* and *S. mediterranea* (Experimental Procedures; white lines represent genes with no ortholog assigned).

(B) WISH analysis of wound-induced genes. Shown are representative animals 4 or 12 hr following incision (the scale bars represent 100 μm; **genes for which WISH analysis of incision was previously published).

(C) Analysis of onset and offset times in different wound-induced genes clusters and injuries, as computed using the impulse model (ks-test).

(D) Expression of representative genes from the early (*egr-l 1*), late (*runt-1*), and sustained (*inhibin-1*) clusters (0–120 hpi) is shown in time course data. Gene expression data points (black dots) are plotted with the impulse fit function (gray line). Onset and offset times, blue and red dashed lines, respectively.

(E) WISH validation of onset and decay times for the genes shown in (D). Gene expression is shown for the three types of injuries tested (anterior, posterior, and incision). The scale bars represent 100 μm.

See also Figure S4.

2011) that extracted transcriptional parameters for every wound-induced gene (Figures 5A and 5C; Experimental Procedures), including their onset and offset times (time to reach half maximal expression and time to return to half baseline expression, respectively; Experimental Procedures). Wound-induced genes were then clustered on the basis of their fitted expression into three groups with significantly different onset and offset parameters (Figures 5A–5D). On the basis of these parameters, wound-induced clusters were labeled as early (n = 44), late (n = 53), or sustained (n = 31). Most of the wound-induced stress-response genes, such as *HSP70*, *HSP90*, and *HSP40*, were part of the early cluster, rapidly induced and fast to decay (Table S5), and our SCS data showed that they are induced in nearly all cell types (Figures 2A and 2B).

The late cluster included many cell-type-specific wound-induced genes, such as patterning factors overexpressed selectively in muscle cells following wounding (Figures 2A–2D; Table S5) (Witchley et al., 2013). Strikingly, in every injury, patterning factors were overexpressed with a median onset of less than 4 hr, even without any tissue loss. Such a rapid induction for these genes is remarkable considering that the timescale of regeneration and its associated patterning is days to more than a week (Reddien and Sánchez Alvarado, 2004).

Next, we compared the onset and offset times of wound-induced gene clusters across injuries (Figure 5C). The onset (~1 hpi) and offset (~12 hpi) of the early cluster did not differ significantly between injuries (ks-test $p > 0.05$, following Bonferroni correction). Similarly, the late cluster was already induced at ~3 hpi in each injury; however, the offset time, following an incision, was almost 20 hr earlier compared with anterior and posterior regeneration ($p < 0.05$; Figure 5C). Finally, the onset and offset of the sustained cluster were significantly earlier in the incision ($p < 0.05$), suggesting that lack of tissue was required for the response to sustain or, alternatively, that tissue fusion was sufficient to terminate it.

We tested these results by selecting candidates from each wound-induced gene cluster and performing WISH time courses (Figures 5D and 5E) on animals that suffered different injuries. Comparison between the fitted data (Figure 5D) and the in situ gene expression (Figures 5E and S4F) further validated that (1) early cluster genes (e.g., *egr-l 1*) displayed similar onset and offset times across injuries and that (2) late and sustained cluster genes (e.g., *runt-1* and *inhibin-1*) had similar expression across injuries in early time points, but their expression returned to baseline earlier at incisions. Together, these results indicated that although the same set of genes is activated at every injury, the duration of their activation is shorter when regeneration is not required.

The Generic Wound Response Is Conserved in a Related Planarian Species

To assess if the generic wound-response program described above in *Schmidtea mediterranea* is conserved in other species, we used a second planarian model, *Girardia dorotocephala* (Flickinger and Coward, 1962). We sequenced and assembled its transcriptome and found high-confidence orthologs for 95 of 128 (74%) of the wound-induced genes (Supplemental Experimental Procedures; Data S1; Table S6). RNA-seq on anterior-facing wounds revealed strong and significant correlation between the fold changes of wound-induced genes in both organisms (Pearson $r = 0.56$, $p = 5.1 \times 10^{-9}$), with genes from all three clusters of wound induction (i.e., early, late, and sustained) being upregulated. The overexpressed genes included cell-type-specific wound-induced *S. mediterranea* genes expressed in muscle (*wntless*, *notum*), neoblasts (*runt-1*, *Tob2*, *inx-13*), and epidermis (*jun-1*, *ston*). Furthermore, both gut- and parapharyngeal-specific genes were induced following injury. In total, 61% (58 of 95) of the *S. mediterranea* wound-induced genes were detectably overexpressed following wounding in *G. dorotocephala* (Table S6). The activation of orthologous stress-response, patterning, and proliferation-related genes further highlights key conserved components of the generic wound response.

The Generic Wound Response Is Followed by a Specific Regenerative Response

The response to wounding was nearly identical in different injuries, despite preceding regeneration of very different anatomy. We therefore used our extended time course data to search for the onset of injury-specific gene expression. We compared the expression of known head-enriched genes (n = 43) (Gurley et al., 2010; Reddien, 2011; Scimone et al., 2014; van Wolfswinkel et al., 2014; Vogg et al., 2014) between tail fragments that regrew heads and incisions that did not require regeneration (Figures 6A and 6B). Fitting the gene expression of regenerating animals (Figure 6; Experimental Procedures) revealed that they had a wide range (>90 hr) of onset values, which was significantly later than the wound-induced genes (ks-test $p = 9.2 \times 10^{-11}$).

Genes were categorized on the basis of previously suggested functions to three groups (1) tissue-patterning factors, which were previously associated with expression in muscle (Witchley et al., 2013); (2) genes associated with specialized neoblasts (Scimone et al., 2014; van Wolfswinkel et al., 2014); and (3) markers of differentiated anterior tissues. All three groups were highly upregulated during anterior regeneration, but they were separable into two distinct phases (Figure 6A). During the first phase, genes enriched in specializing neoblasts (34 hpi) and anteriorly expressed patterning genes (39 hpi) were upregulated. Subsequently, almost 40 hr later, genes enriched in differentiated head cell types were upregulated (ks-test $p = 4.4 \times 10^{-4}$; 77 hpi; Figure 6A). Similar phases were found for orthologous genes in *G. dorotocephala* (Figure 6C). Importantly, both regenerative phases were separable from the generic wound-response onset by over 24 hr (ks-test $p = 9.2 \times 10^{-11}$; Figure 6D). By contrast, in animals suffering incisions we could not detect significant expression changes in any of the genes associated with regeneration (Figure 6B), which prohibited fitting to the impulse model, indicating that these were indeed part of a specific regenerative response.

Hierarchal clustering of samples from the anterior regeneration and incision time courses, using wound-induced gene expression, further supported the conclusion that gene expression changes are sustained only when tissue is missing (Figure 6E). Samples from early time points (0, 1, and 4 hpi) from incisions and anterior amputations formed a cluster, because of similarities in early wound response. However, starting at 12 hpi, the wound-induced gene expression at incisions was largely eliminated (Figures 5A–5C), and these samples clustered with 72 and 120 hpi samples from anterior-regenerating fragments.

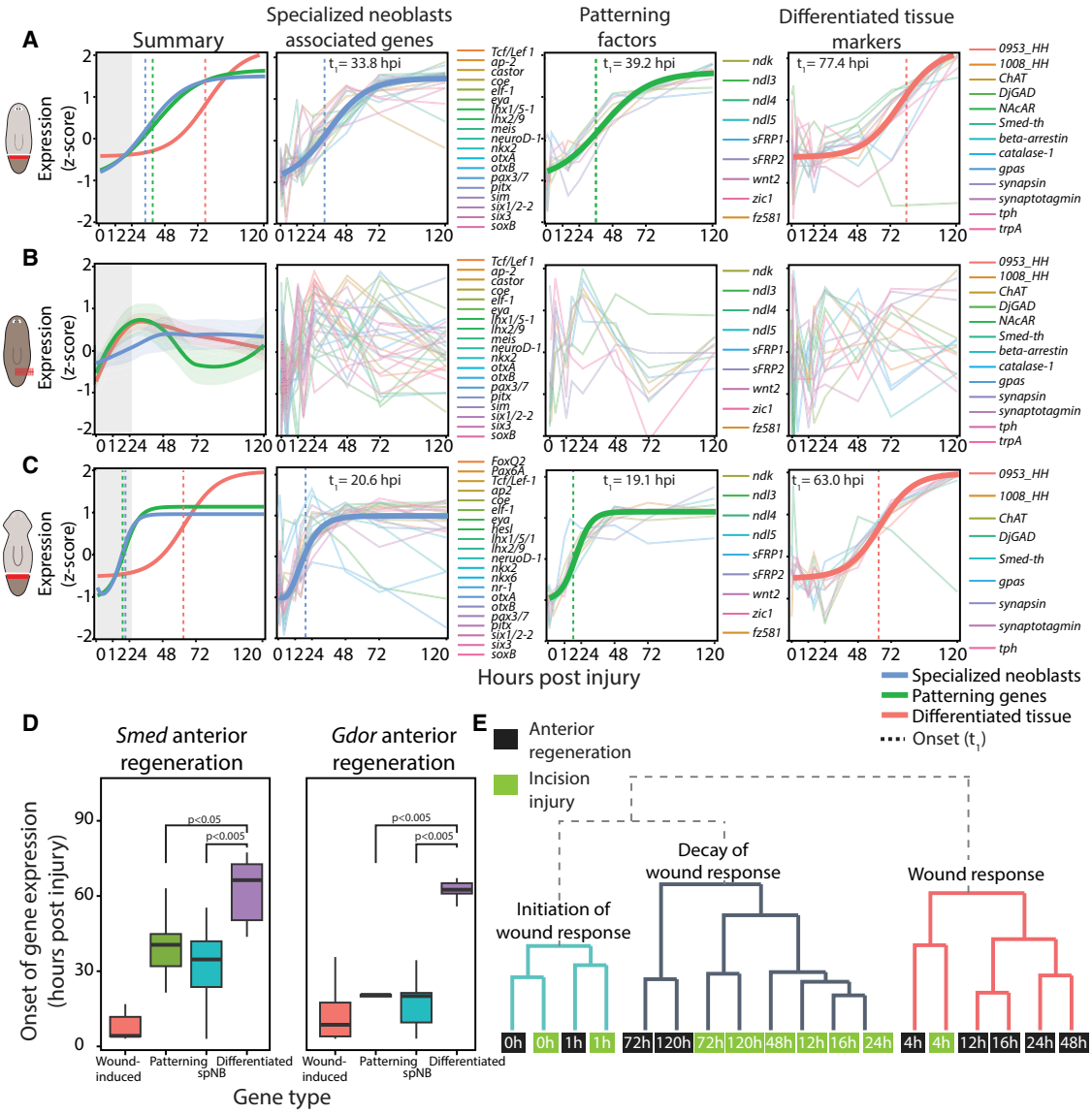


Figure 6. Injury-Specific Regeneration Occurs in a Temporally Defined Order

(A) Summary panel: shown is a fit of the normalized median expression of neoblast specialization-associated genes, injury-specific patterning factors, and terminally differentiated tissue markers (blue, green, and red, respectively). Matching colored vertical lines mark the onset times of the corresponding group of genes. Gray box highlights the wound-response phase. Other panels: bold lines represent impulse model fit of the genes used for modeling the dynamics of the group; thin lines represent individual genes. Onset time is marked by a vertical dashed line.

(B) The genes used for (A) were plotted with the incision time course data in which there was no missing tissue. Shown is a loess fit (bold lines) and confidence interval of the z scores for each class of genes (lightly colored area) because the data could not be fit to the impulse model. Individual panels show a non-specific response following wounding.

(C) A similar analysis performed on anteriorly regenerating *G. dorotocephala* revealed a similar order of events to amputation in *S. mediterranea*.

(D) Box plot showing the onset time of different groups of genes following amputation. Boxes represent the interquartile range, thick lines are the median. Statistical significance was tested by a ks-test.

(E) Dendrogram illustrating the similarity of gene expression of wound-induced genes in samples from the anterior regeneration and the incision time courses. Each node represents a sample (0–120 hpi; green and black nodes, incision and anterior samples, respectively). Annotations on the tree represent the interpretation of samples in clade.

Our results support a model of a sequentially activated regenerative program starting with the generic wound response (0–24 hpi), followed by the expression of injury-specific patterning factors and specialized neoblast genes (~30 hpi), and finally with the appearance of differentiated tissues (~70 hpi).

DISCUSSION

The ability of planarians to regenerate from almost any injury, combined with the wide array of methods established for their study, makes them a unique system for studying regeneration

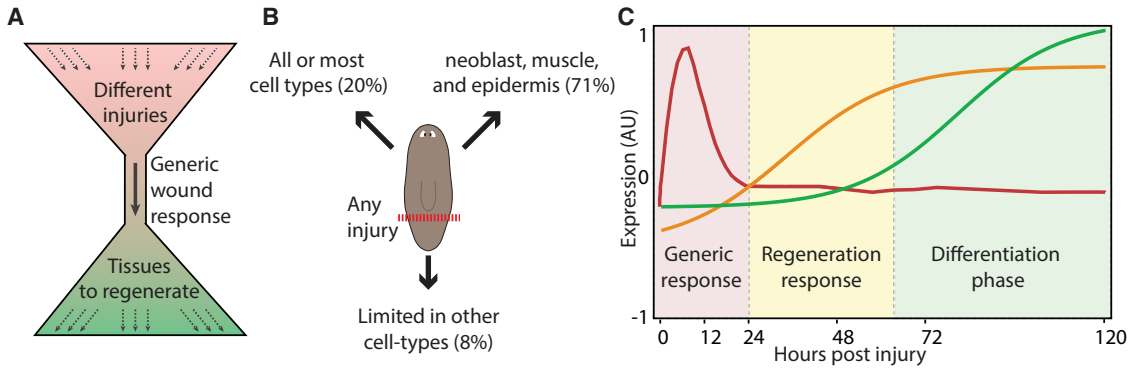


Figure 7. Model for Planarian Wound Response and Initiation of Regeneration

(A) Planarians regenerate from almost any injury through a single transcriptional response.

(B) Transcriptional changes following the wound response are divided into three cellular components.

(C) A temporal model of planarian regeneration. Every injury triggers a prototypical generic response (red box; red line). If regeneration is not required following the injury, the response will decline. Otherwise, the expression of an injury-specific response emerges (yellow box; yellow line). These responses involve patterning molecules and neoblast-associated fate specialization genes. About 3 days following the injury, expression of differentiated tissue markers appears in association with the emergence of the newly regenerated structures (green box; green line).

initiation. Here, we took an SCS approach, combined with bulk tissue sequencing from multiple distinct wound types, to characterize the transcriptional responses associated with planarian regeneration initiation. Our data support a model in which a generic transcriptional program is activated by wounding to accommodate the regeneration of diverse tissue types depending on the nature of the injury (Figure 7A). How can a generically activated transcriptional program be activated if every injury involves different combinations of cell types at unpredictable wound sites? We found that the generic wound response includes stress-related responses in all cell types and cell-specific responses in neoblasts, muscle, and epidermis that are distributed throughout the planarian body (Figure 7B). Finally, following the generic wound response, injury-specific transcription is activated, including patterning and stem cell specialization genes, that precedes the appearance of differentiated tissue markers by ~40 hr (Figure 7C). Together, these results link a common transcriptional wound response with divergent regenerative outcomes.

Wound-Response Polarity Is Likely Determined by a Single Gene, *notum*

To find genes activated at wounds associated with different regenerative outcomes, we performed RNA-seq on two wound types that regenerate different tissues, heads or tails. Strikingly, only one gene, *notum*, a Wnt-pathway inhibitor (Gerlitz and Basler, 2002), demonstrated a strong bias in expression (more than 2-fold) to one of the two injuries. *notum* was previously shown to be preferentially expressed at anterior-facing wounds over posterior-facing wounds and to be required for the head-versus-tail regeneration decision (Petersen and Reddien, 2011). However, whether other genes showed similar expression asymmetry was unknown. We tested more than 200 additional genes that appeared to show any expression bias to one of the two injuries but found none that were clearly preferentially induced in one wound type over the other. Other subtle transcriptional differences could exist between these wounds, but were undetectable by RNA-seq and WISH. Therefore, our analyses suggest that *notum* is the only gene with a transcriptional response distinguishing anterior and posterior-facing wounds up to 24 hpi, which is striking given that these wounds will initiate completely different regenerative programs.

A Generic, Conserved Response to Wounding Precedes Regeneration

Several planarian genes were previously shown to be induced following wounding, even without tissue loss, suggesting that they are generically induced by the injury (Petersen and Reddien, 2011; Wenemoser et al., 2012). Interestingly, a few of these genes, such as *wnt1*, are important planarian patterning genes (Petersen and Reddien, 2009). Using time course experiments from different anatomical positions, we rigorously tested the hypothesis that a common transcriptional program is activated at every type of wound. We found that indeed all wound responses start the same, regardless of the eventual regenerative outcome. We estimated that the generic response involves the upregulation of 224 genes in the first 12 hr following injury. When there was no missing tissue to regenerate, the wound response initiated largely normally but decayed earlier.

We propose that the generic wound response acts as a funnel between the varied injuries an organism might suffer and subsequent varied regenerative outcomes (Figure 7). As such, the generic response includes all the necessary components for promoting survival and allowing regeneration of any tissue. The generic response is modified with time to achieve the necessary regenerative outcome. In parallel to the transcriptional wound response, massive neoblast proliferation (Wenemoser and Reddien, 2010) and apoptosis (Pellettieri et al., 2010) take place following any injury, even at injuries that will not require substantial regeneration, such as following needle puncture. Strikingly, these processes appear to be interconnected: following the initial generic wound response a sequence of events involving the activation of context-dependent transcriptional programs (Lapan and Reddien, 2012; Scimone et al., 2011), mitosis (Wenemoser and Reddien, 2010), and apoptotic (Pellettieri et al., 2010) responses are observed only when the injury requires regeneration.

Cell-Type-Specific Wound-Response Genes

How could activation of the same transcriptional program be accommodated by diverse wound locations (injuries through the brain versus tail, for instance), where different cell types juxtapose the wound?

Analysis of some genes activated by wounding showed that multiple tissues are involved, including the epidermis (Wenemoser et al., 2012) and muscle (Witchley et al., 2013), although it remained unclear to what extent these results are generalizable. We compiled a list of wound-induced genes through time course experiments and assessed their expression in single cells from wounds. Our results demonstrated that the response to wounding has three components (Figure 7B): (1) a non-specific component, with genes expressed in nearly all cell types following wounding, including multiple stress-response genes, (2) a specific component, including 71% of the cell-type-specific genes, with preferential expression in one of three cell types: neoblast, muscle, or epidermis (this component included multiple patterning factors [Witchley et al., 2013], transcription factors, and genes associated with proliferation); and (3) individual wound-induced genes were expressed in gut, parapharyngeal cells, and neurons, reflecting unique physiological responses in these tissues following wounding. The architecture of the wound response, composed of genes activated in any cell type at the wound and cell-type-specific genes activated in cells widespread in the body, enables the same genes to be activated at essentially all wounds.

Several lines of evidence support the accuracy of wound-induced expression cell type assignments. First, wound-induced expression was much lower before injuries (RNA-seq and WISH); therefore, cells with the strongest SCS expression are the best candidates to explain wound-induced expression. Second, in most cases, SCS expression was mostly limited to a single cell type. Third, dFISH validated cell-type assignments for a set of tested genes. Finally, direct comparison of neoblasts isolated from intact and injured animals was in agreement with the SCS analysis.

The Onset of Regeneration and the Pruning of the Wound Response

Through extended time course experiments, we found that 24 hr following the peak of wound response, patterning genes associated with regeneration (Reddien, 2011; Witchley et al., 2013) were overexpressed, hand in hand, with transcription factors associated with neoblast specialization (Scimone et al., 2014). Upregulation of these genes emerged almost 40 hr before the upregulation of differentiated tissue markers. We therefore suggest that regeneration can be modeled by three components of gene expression changes (Figure 7C): (1) activation of a generic wound response (~224 genes), which allows the animal to mount a regenerative response to essentially any injury (0–16 hpi); (2) expression of patterning factors and neoblast specialization genes, specific to the identity of tissues being regenerated (~36 hpi); and (3) expression of differentiated tissue markers associated with functional new tissue (72 hpi) (Figure 7C).

A Unique Repository of Cell-Type-Specific Expression

This work presents the first application of SCS to planarians. Therefore, many of the profiled cell types were not previously studied at the molecular level in detail. This analysis therefore generated a unique repository, including 1,214 unique cell-type-specific markers, including signaling molecules, receptors, and transcription factors. We developed an online resource that allows accessing the transcriptome of every cell from all identified cell types, available at https://radiant.wi.mit.edu/app/.

Previous studies profiled the gene expression of several planarian cell types through the application of specially developed cell isolation techniques (Forsthoefel et al., 2012). Although successful in studying the targeted tissue, such approaches are not readily applicable to every cell type. Furthermore, as these methods are applied to cell populations, they do not reveal cell-to-cell heterogeneity or gene co-expression in individual cells (Shalek et al., 2013). By contrast, the single-cell expression data allowed us to generate comprehensive co-expression profiles in every profiled cell type, as well as their cell-type expression heterogeneity (online resource).

Conclusions

Our analysis suggests a simple and unifying model for the planarian wound response. SCS data indicate that a large component of this response is driven specifically by three abundant tissues (Figure 7B) that allow the response to take place regardless of the anatomy and location of the wound site. Instead of tailoring the response for the desired outcome, the response logic operates in an "act-first" mechanism: activating a program that is sufficient for recovery from any injury. This program is subsequently replaced with an injury-specific response appropriate for regeneration from a specific injury (Figure 7).

EXPERIMENTAL PROCEDURES

Planarian Culture

Clonal lines of asexual *S. mediterranea* (CIW4) and *G. dorotocephala* were maintained as previously described (van Wolfswinkel et al., 2014).

Single-Cell Library Construction

Libraries were prepared using the SmartSeq2 method, as previously described (Picelli et al., 2013, 2014). Briefly, RNA from single cells was reverse transcribed with a poly-dT anchored oligo and a template-switching oligo. cDNA was then amplified. Sequencing libraries were prepared using the Nextera XT kit (Illumina).

Sequencing Reads Mapping

Sequencing reads were mapped to the *S. mediterranea* dd_Smed_v4 assembly (http://planmine.mpi-cbg.de; Liu et al., 2013) using Novoalign version 2.08.02 with parameters [-o SAM -r Random] and were converted to BAM using samtools version 1.1 (Li et al., 2009). Read count, for every sample, was calculated with bedtools version 2.20.1 (Quinlan and Hall, 2010). Read counts were normalized by edgeR (Robinson et al., 2010). *G. dorotocephala* libraries were similarly mapped to a de novo transcriptome assembly (Data S1).

Single-Cell Data Clustering

An expression matrix for all cells was prepared for analysis in R version 3.1.1. Samples expressing less than 1,000 or more than 9,000 genes were discarded from further analysis. Genes that were used for t-SNE representation and density-based clustering (Ester et al., 1996) were selected by identifying principal components that contribute to the variance using the Seurat method (Macosko et al., 2015; Satija et al., 2015) (Supplemental Experimental Procedures).

Detection of Cluster-Specific Genes

Cluster-specific genes were detected by enrichment analysis (McDavid et al., 2013) on genes displaying at least 2-fold enrichment in a cluster compared to

all other clusters. Controlled p values, for each gene, were calculated using the Seurat package (Satija et al., 2015). Then, a binary classifier was used on every cell-type-specific gene (FDR < 0.1) (Sing et al., 2005). The classifier quantified, for each of the genes tested, its ability to partition the cells it was enriched in from all other cells. For every gene, the true positive rate (sensitivity) and false positive rate (1 − specificity) were calculated, and a ROC curve was generated (Figure S1H).

WISH using RNA Probes

WISH was performed as previously described (Pearson et al., 2009).

Gene Cloning

Genes were amplified from planarian cDNA using gene-specific primers (Supplemental Experimental Procedures) and cloned into a pGEM vector (Promega).

Gene Annotation

Previously undescribed genes were annotated by best-BLAST hit ($e < 1 \times 10^{-5}$) against a sequence database including planarian, human, mouse, fly, and *C. elegans* sequences. If BLAST hits were not found, the contig identifier from the transcriptome assembly (Liu et al., 2013) was used. See Supplemental Experimental Procedures for a list of all annotations used in the figures and their corresponding contig identifiers in the assembly.

Double-Stranded RNA Synthesis

Double-stranded RNA was synthesized as previously described (Petersen and Reddien, 2008). RNA was quantified by Nanodrop (Thermo Fisher Scientific) to have at least 5 μg/μl.

Illumina Library Preparations for Anterior and Posterior Time Courses

Prepharyngeal fragments were isolated in biological triplicates and placed in TRIzol reagent (0 hpi). Anterior-facing or posterior-facing wounds were amputated as prepharyngeal fragments at 3, 6, and 12 hpi in biological triplicates. RNA was purified according to manufacturer's instructions (Life Technologies), and sequencing libraries were prepared with a TruSeq RNA sample preparation kit V2 (Illumina).

Illumina Library Preparations for Extended Time Courses

Wound tissues were isolated and put in TRIzol. Tissues were lysed with Qiagen TissueLyser II, and RNA was extracted according to the manufacturer's instructions. Libraries were prepared as previously described (Engreitz et al., 2014; Schwartz et al., 2014) (Supplemental Experimental Procedures).

Detection of Differentially Expressed Genes and Genes with Putative Asymmetric Wound Expression

Wound-induced genes were called using triplicate time course experiments by using the edgeR exactTest function to compare expression at every wounding time point to 0 hr. Genes called as wound induced met the following thresholds in at least one time point: FDR $\leq$ 0.05, FC $\geq$ 2, minimal expression of reads per kilobase of transcript per million mapped reads = 6 in at least 2 of 21 libraries). Putative asymmetric expression was detected by comparing anterior and posterior wound-induced gene expression from matched time points using exactTest. All genes with FDRs $\leq$ 0.05 and FCs $\geq$ 1.5 were tested by WISH analysis, as well as 218 other genes not meeting these thresholds (Table S4).

Single-Cell Isolation and FACS

Cells from postpharyngeal wound sites were isolated and sorted (Hayashi et al., 2006) into 96-well microplates containing 5 μl Buffer TCL (Qiagen) plus 1% 2-mercaptoethanol.

Detection of Onset and Offset of Wound Induction

To extract onset and offset parameters of genes, expression data from each time course were used for fitting by the impulse model (Chechik and Koller, 2009; Chechik et al., 2008) using a MATLAB implementation (Sivriver et al., 2011) with constraint parameters (retries = 100, $t1 \geq 0$, $t2 \geq 0$, $h0 \geq 0$, $h1 \geq 0$, $h2 \geq 0$, $\beta1 \geq 0$, $\beta2 \leq 0$).

ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is SRA: PRJNA276084.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, six tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.devcel.2015.11.004.

AUTHOR CONTRIBUTIONS

O.W. and P.W.R. conceived and designed the overall study. O.W., L.E.C., and A.P. designed and performed experiments. O.W. analyzed sequencing data with feedback from R.S. and A.R. O.W. and P.W.R. wrote the manuscript with comments from all authors.

ACKNOWLEDGMENTS

We thank Travis R. Rogers for experimental support, David Gennert for SCS protocol help, and Naomi Habib for MATLAB implementation of the impulse algorithm. We thank Ben Kleaveland, Schraga Schwartz, Eran Mick, Sarit Edelheit, Guy Bushkin, and Ophir Shalem for critical reading and the Reddien lab members for manuscript comments. O.W. was supported by a European Molecular Biology Organization long-term fellowship and is the Howard Hughes Medical Institute Fellow of The Helen Hay Whitney Foundation. We acknowledge NIH support (grant R01GM080639). P.W.R. is a Howard Hughes Medical Institute Investigator and an associate member of the Broad Institute of MIT and Harvard. A.R. acknowledges the Klarman Cell Observatory. R.S. was supported by NIH grant F32 HD075541. A.R. is a member of the scientific advisory boards at Thermo Fisher Scientific and Syros Pharmaceuticals and is a consultant for Driver Genomics.

Received: June 5, 2015
Revised: October 2, 2015
Accepted: November 6, 2015
Published: December 7, 2015

REFERENCES

Adell, T., Salò, E., Boutros, M., and Bartscherer, K. (2009). *Smed-Evi/Wntless* is required for beta-catenin-dependent and -independent processes during planarian regeneration. Development *136*, 905–910.

Chechik, G., and Koller, D. (2009). Timing of gene expression responses to environmental changes. J. Comput. Biol. *16*, 279–290.

Chechik, G., Oh, E., Rando, O., Weissman, J., Regev, A., and Koller, D. (2008). Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. Nat. Biotechnol. *26*, 1251–1259.

Cowles, M.W., Brown, D.D., Nisperos, S.V., Stanley, B.N., Pearson, B.J., and Zayas, R.M. (2013). Genome-wide analysis of the bHLH gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. Development *140*, 4691–4702.

DuBuc, T.Q., Traylor-Knowles, N., and Martindale, M.Q. (2014). Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. BMC Biol. *12*, 24.

Eisenhoffer, G.T., Kang, H., and Sánchez Alvarado, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. Cell Stem Cell *3*, 327–339.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell *159*, 188–199.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd International Conference on Knowledge Discovery

and Data Mining (KDD-96), http://www.dbs.ifi.lmu.de/Publikationen/Papers/KDD-96.final.frame.pdf.

Flickinger, R.A., and Coward, S.J. (1962). The induction of cephalic differentiation in regenerating Dugesia dorotocephala in the presence of the normal head and in unwounded tails. Dev. Biol. *5*, 179–204.

Forsthoefel, D.J., Park, A.E., and Newmark, P.A. (2011). Stem cell-based growth, regeneration, and remodeling of the planarian intestine. Dev. Biol. *356*, 445–459.

Forsthoefel, D.J., James, N.P., Escobar, D.J., Stary, J.M., Vieira, A.P., Waters, F.A., and Newmark, P.A. (2012). An RNAi screen reveals intestinal regulators of branching morphogenesis, differentiation, and stem cell proliferation in planarians. Dev. Cell *23*, 691–704.

Gaviño, M.A., Wenemoser, D., Wang, I.E., and Reddien, P.W. (2013). Tissue absence initiates regeneration through follistatin-mediated inhibition of activin signaling. eLife *2*, e00247.

Gerlitz, O., and Basler, K. (2002). Wingful, an extracellular feedback inhibitor of Wingless. Genes Dev. *16*, 1055–1059.

Glazer, A.M., Wilkinson, A.W., Backer, C.B., Lapan, S.W., Gutzman, J.H., Cheeseman, I.M., and Reddien, P.W. (2010). The Zn finger protein Iguana impacts Hedgehog signaling by promoting ciliogenesis. Dev. Biol. *337*, 148–156.

Guo, T., Peters, A.H., and Newmark, P.A. (2006). A Bruno-like gene is required for stem cell maintenance in planarians. Dev. Cell *11*, 159–169.

Gurley, K.A., Elliott, S.A., Simakov, O., Schmidt, H.A., Holstein, T.W., and Sánchez Alvarado, A. (2010). Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. Dev. Biol. *347*, 24–39.

Gurtner, G.C., Werner, S., Barrandon, Y., and Longaker, M.T. (2008). Wound repair and regeneration. Nature *453*, 314–321.

Hayashi, T., Asami, M., Higuchi, S., Shibata, N., and Agata, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. Dev. Growth Differ. *48*, 371–380.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Kao, D., Felix, D., and Aboobaker, A. (2013). The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. BMC Genomics *14*, 797.

Knapp, D., Schulz, H., Rascon, C.A., Volkmer, M., Scholz, J., Nacu, E., Le, M., Novozhilov, S., Tazaki, A., Protze, S., et al. (2013). Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. PLoS ONE *8*, e61352.

Lapan, S.W., and Reddien, P.W. (2012). Transcriptome analysis of the planarian eye identifies *ovo* as a specific regulator of eye regeneration. Cell Rep. *2*, 294–307.

Lengfeld, T., Watanabe, H., Simakov, O., Lindgens, D., Gee, L., Law, L., Schmidt, H.A., Ozbek, S., Bode, H., and Holstein, T.W. (2009). Multiple Wnts are involved in Hydra organizer formation and regeneration. Dev. Biol. *330*, 186–199.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, S.Y., Selck, C., Friedrich, B., Lutz, R., Vila-Farré, M., Dahl, A., Brandl, H., Lakshmanaperumal, N., Henry, I., and Rink, J.C. (2013). Reactivating head regrowth in a regeneration-deficient planarian species. Nature *500*, 81–84.

Louvi, A., and Grove, E.A. (2011). Cilia in the CNS: the quiet organelle claims center stage. Neuron *69*, 1046–1060.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

Pearson, B.J., and Sánchez Alvarado, A. (2010). A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. Development *137*, 213–221.

Pearson, B.J., Eisenhoffer, G.T., Gurley, K.A., Rink, J.C., Miller, D.E., and Sánchez Alvarado, A. (2009). Formaldehyde-based whole-mount in situ hybridization method for planarians. Dev. Dyn. *238*, 443–450.

Pellettieri, J., Fitzgerald, P., Watanabe, S., Mancuso, J., Green, D.R., and Sánchez Alvarado, A. (2010). Cell death and tissue remodeling in planarian regeneration. Dev. Biol. *338*, 76–85.

Petersen, C.P., and Reddien, P.W. (2008). *Smed-betacatenin-1* is required for anteroposterior blastema polarity in planarian regeneration. Science *319*, 327–330.

Petersen, C.P., and Reddien, P.W. (2009). A wound-induced Wnt expression program controls planarian regeneration polarity. Proc. Natl. Acad. Sci. U S A *106*, 17061–17066.

Petersen, C.P., and Reddien, P.W. (2011). Polarized *notum* activation at wounds inhibits Wnt function to promote planarian head regeneration. Science *332*, 852–855.

Picelli, S., Björklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Reddien, P.W. (2011). Constitutive gene expression and the specification of tissue identity in adult planarian biology. Trends Genet. *27*, 277–285.

Reddien, P.W., and Sánchez Alvarado, A. (2004). Fundamentals of planarian regeneration. Annu. Rev. Cell Dev. Biol. *20*, 725–757.

Reddien, P.W., Oviedo, N.J., Jennings, J.R., Jenkin, J.C., and Sánchez Alvarado, A. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science *310*, 1327–1330.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sánchez Alvarado, A., Newmark, P.A., Robb, S.M., and Juste, R. (2002). The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. Development *129*, 5659–5665.

Sandmann, T., Vogg, M.C., Owlarn, S., Boutros, M., and Bartscherer, K. (2011). The head-regeneration transcriptome of the planarian *Schmidtea mediterranea*. Genome Biol. *12*, R76.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162.

Scimone, M.L., Srivastava, M., Bell, G.W., and Reddien, P.W. (2011). A regulatory program for excretory system regeneration in planarians. Development *138*, 4387–4398.

Scimone, M.L., Kravarik, K.M., Lapan, S.W., and Reddien, P.W. (2014). Neoblast specialization in regeneration of the planarian *Schmidtea mediterranea*. Stem Cell Reports *3*, 339–352.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014).

Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., and Agata, K. (1999). Expression of vasa(vas)-related genes in germline cells and totipotent somatic stem cells of planarians. Dev. Biol. *206*, 73–87.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. Bioinformatics *21*, 3940–3941.

Sivriver, J., Habib, N., and Friedman, N. (2011). An integrative clustering and modeling algorithm for dynamical gene expression data. Bioinformatics *27*, i392–i400.

van Wolfswinkel, J.C., Wagner, D.E., and Reddien, P.W. (2014). Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. Cell Stem Cell *15*, 326–339.

Vogg, M.C., Owlarn, S., Pérez Rico, Y.A., Xie, J., Suzuki, Y., Gentile, L., Wu, W., and Bartscherer, K. (2014). Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. Dev. Biol. *390*, 136–148.

Wagner, D.E., Wang, I.E., and Reddien, P.W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. Science *332*, 811–816.

Wenemoser, D., and Reddien, P.W. (2010). Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. Dev. Biol. *344*, 979–991.

Wenemoser, D., Lapan, S.W., Wilkinson, A.W., Bell, G.W., and Reddien, P.W. (2012). A molecular wound response program associated with regeneration initiation in planarians. Genes Dev. *26*, 988–1002.

Witchley, J.N., Mayer, M., Wagner, D.E., Owen, J.H., and Reddien, P.W. (2013). Muscle cells provide instructions for planarian regeneration. Cell Rep. *4*, 633–641.

Developmental Cell

Supplemental Information

A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians

Omri Wurtzel, Lauren E. Cote, Amber Poirier, Rahul Satija, Aviv Regev, and Peter W. Reddien

Supplementary figures

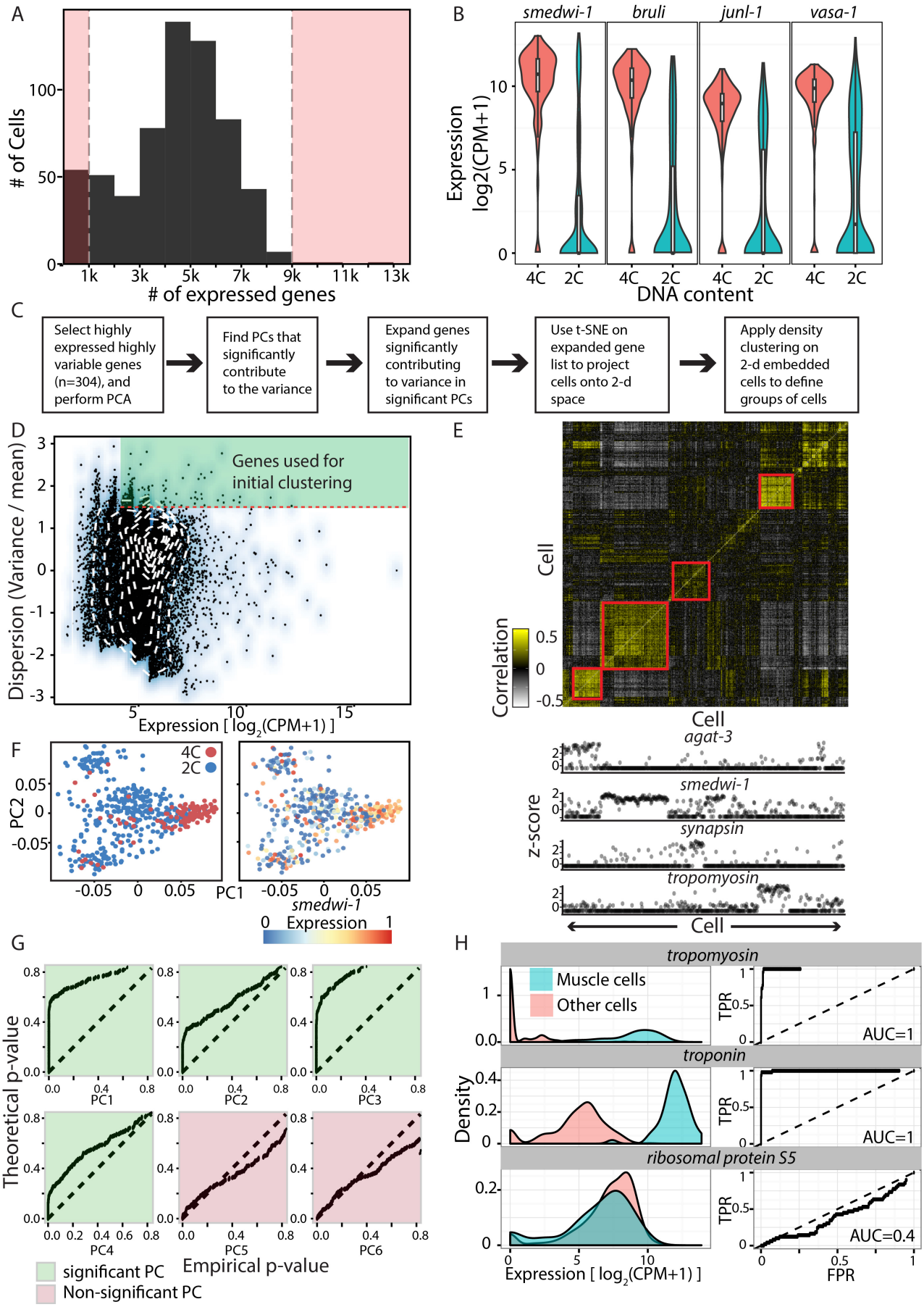


Figure S1, related to Figure 1. Assessment of single-cell sequencing. (A) Histogram of the number of genes expressed (CPM > 0) in the single cells used in this analysis. Shaded red regions represent thresholds used for eliminating cells with extremely low or high number of expressed genes. (B) Violin plots of canonical neoblast markers (Wagner et al., 2012) between cells that were collected using FACS from the 4C gate (Methods) compared to the 2C gate. White rectangle represents the interquartile range, black bar is the median. (C) Schematic describing the different steps of analysis performed for assigning cells to clusters. (D) Genes selected for initial clustering exhibited high dispersion and expression across the sequenced-cells (2C: 0h, 4-6, and 12-14 hpi, 4C: 4-6, 12-14 hpi). Black dots with blue hue represent the mean expression of a gene, white dashed contours represent the density of the dots as obtained by 2d kernel density estimation with 25 bins (kde2d). Green shaded rectangle outlines the selection of genes for the initial clustering. (E) Upper panel: Correlation matrix generated based on the initial set of genes. Cell-order was determined by hierarchal clustering of the cells based on the initial set of genes used for clustering. Red rectangles represent cells co-expressing canonical markers for several cell types: *smedwi-1* for neoblasts, *tropomyosin* for muscle, *synapsin* for neural cells, and *agat-3* for *late epidermal lineage* cells. (F) Left panel: PCA projection of individual sequenced cells (dots), based on the initial set of genes (n=304) used for clustering. PC1 separates the dividing cell fraction from cells that are not dividing (red and blue; 4C and 2C DNA content, respectively; DNA content determine by Hoechst dye analysis during cell isolation with FACS; Methods). Right panel: The same PCA projection is shown with the cells colored based on their rank of expression of the canonical neoblast marker *smedwi-1* (blue, yellow, and red correspond to low, medium and high). Most of the *smedwi-1* expressing cells are separated by PC1. (G) Testing the significance of different principal components through Jackstraw analysis (Chung and Storey, 2015). Each subplot is a quantile-quantile plot (qqplot) of gene p-values in the principal component, as determined by a jackstraw analysis compared to theoretical p-values based on sampling from uniform distribution (Extended experimental procedures). Empirical values near the dashed lines fit a uniform distribution and hence are not considered for further testing (in this case genes were selected from principal components 1 through 4). Green and red backgrounds represent PCs found to be significant and non-significant, respectively, through this analysis. (H) Example of classification of genes to clusters. Shown is the cluster we subsequently determined to be muscle. Left panel: For every cluster, a list of genes that are highly expressed compared to all other clusters was assembled. Shown is the expression of the canonical muscle markers *tropomyosin* and *troponin*, and a negative control *ribosomal protein s5* (top, middle, and

bottom, respectively; blue and red area, muscle cluster, and all other clusters, respectively). Right panel: The ability of individual genes to partition the cells to the tested cluster is plotted by the true positive rate (TPR; sensitivity) and false positive rate (FPR; 1 - specificity) of the assignments, and the area under the curve. The diagonal (dashed black line; AUC=0.5) represents random assignment to the cluster, such as observed for the negative control.

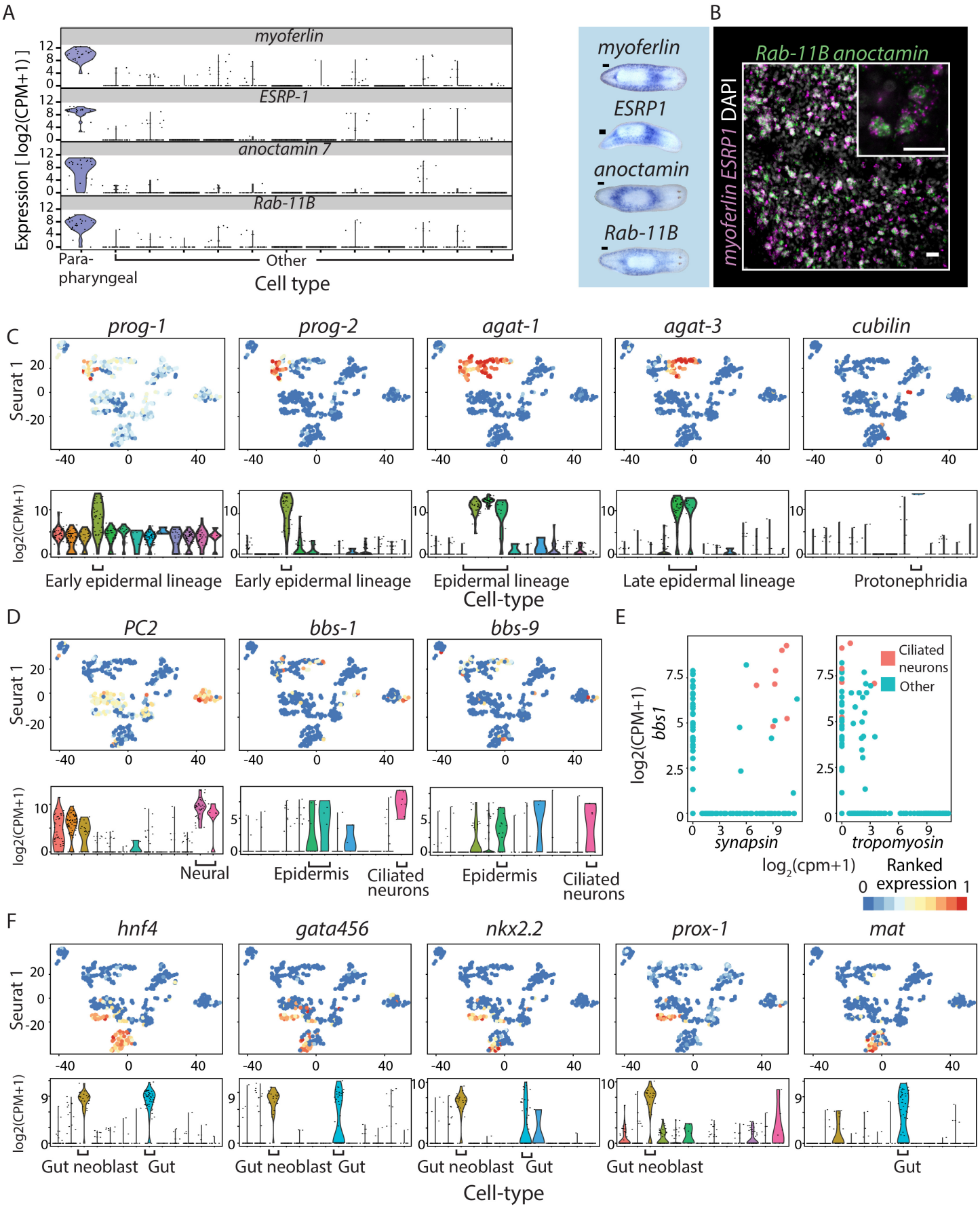

A
myoferlin
ESRP-1
anoctamin 7
Rab-11B
Expression [log2(CPM+1)]
Para-pharyngeal
Other
Cell type
B
myoferlin
ESRP1
anoctamin
Rab-11B
Rab-11B anoctamin
myoferlin ESRP1 DAPI
C
prog-1
prog-2
agat-1
agat-3
cubilin
Seurat 1
log2(CPM+1)
Early epidermal lineage
Early epidermal lineage
Epidermal lineage
Late epidermal lineage
Protonephridia
Cell-type
D
PC2
bbs-1
bbs-9
Seurat 1
log2(CPM+1)
Neural
Epidermis
Ciliated neurons
Epidermis
Ciliated neurons
E
log2(CPM+1) bbs1
synapsin
tropomyosin
Ciliated neurons
Other
log2(cpm+1)
Ranked expression
F
hnf4
gata456
nkx2.2
prox-1
mat
Seurat 1
log2(CPM+1)
Gut neoblast
Gut
Gut neoblast
Gut
Gut neoblast
Gut
Gut neoblast
Gut
Cell-type

Figure S2, related to Figure 1. Single cell gene expression planarian resource. (A) Left panel: violin plots show high specificity to a single cluster (violet; black dots represent single-cell expression). Right panel: WISH analyses of the genes reveal, in all cases, a parapharyngeal localization (scale = 100 μm). (B) dFISH of genes enriched in the parapharyngeal cluster. Pooled probes for *myoferlin* and *ESRP-1* (magenta) and *Rab-11B* and *anoctamin* (green) were used for coexpression analysis (scales = 20 μM; DAPI in gray). (C) Upper panel: Expression of cell type specific markers plotted on *Seurat* maps showing the specificity of genes to cell types (Cells represented by dots; color is the ranked expression of the gene in cells. blue to red, low and high ranked expression, respectively). Lower panels: Violin plots of gene expression across cell-types. X-axis annotation highlights cell types enriched for the plotted gene. (D) Upper panels: *Seurat* maps of the canonical neural marker PC2, and two canonical cilia components (*bbs-1* and *bbs-9*). The components are expressed almost exclusively in differentiated epidermal cells and in a subset of the neural cells (*PC2+/synapsin+)*. Lower panels: violin plots of neural (*PC2*) and cilia (*bbs1* and *bbs9*) related genes. (E) Left panel: Co-expression plots of *bbs-1*, a cilia component, and *synapsin*, a canonical neural marker, shows that a subset of the cells expressing high levels of *synapsin* also express *bbs1*. Right panel: Co-expression of *bbs-1* with *tropomyosin*, a canonical muscle marker, shows that there are no cells highly expressing both genes (cells represented by dots, red and blue colors are cells determined to be ciliated neurons, and other cells, respectively). (F) Upper and lower panels: *Seurat* maps and violin plots of putative gut neoblasts markers, including transcription factors and gut markers. The expression of *hnf4*, *gata4/5/6*, *nkx-2.2* appears in both the differentiated gut cluster, and the gut (γ) neoblasts. The transcription factor *prox-1* is expressed the gut neoblasts cluster, but not in the differentiated gut cells. *mat*, a planarian gut marker, is expressed exclusively in the differentiated gut marker.

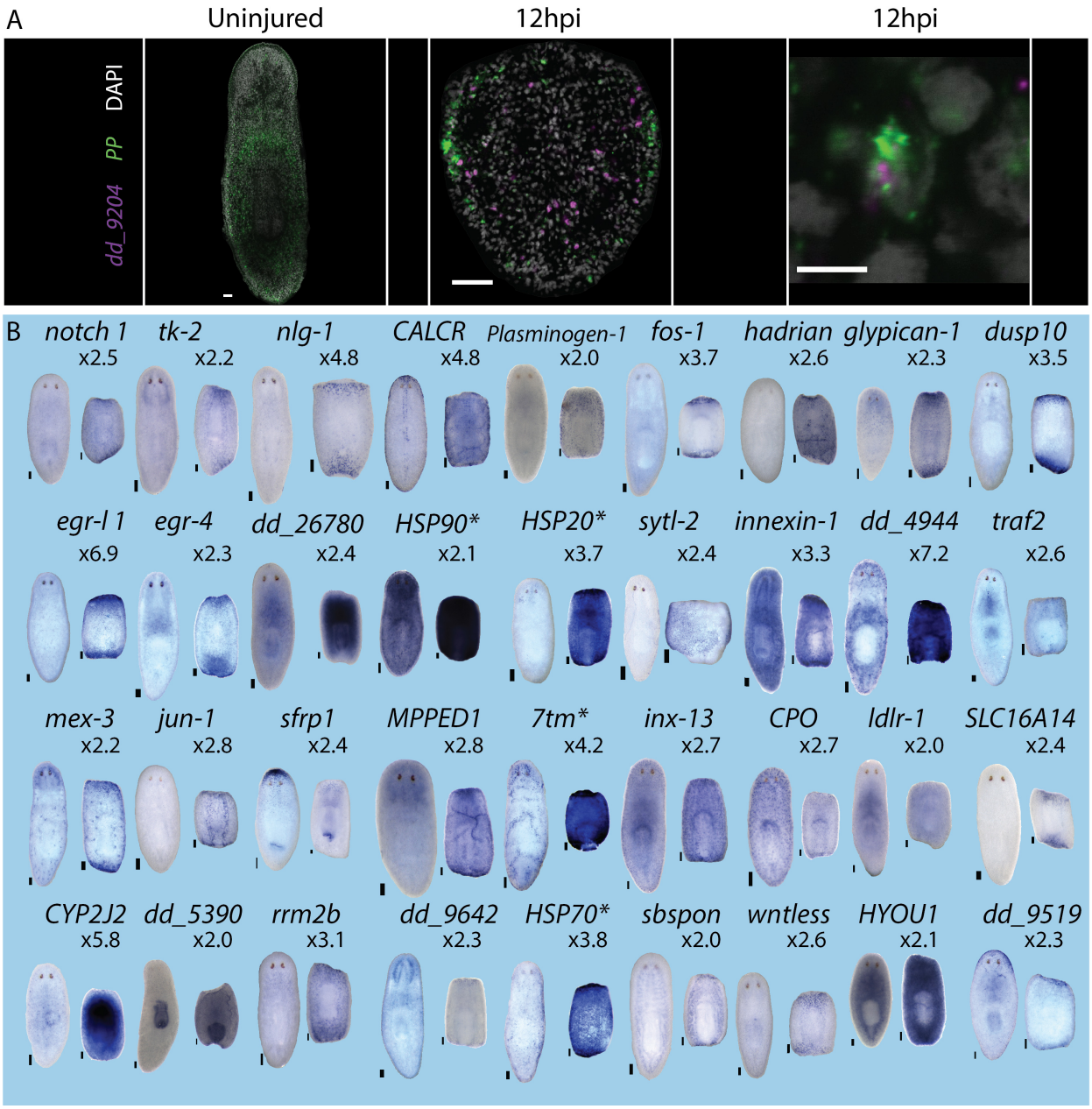


Figure S3, related to Figure 4. Wound induced gene expression. (A) dFISH validation of parapharyngeal-specific gene expression of *dd_9204* (magenta) with a parapharyngeal probe pool (*myoferlin* and *ESRP1*; green) and DAPI (gray) in intact and injured (12 hpi) animals. Scale = 50 µm; right panel scale = 5 µm. (B) WISH analysis of 36 additional genes tested for asymmetry in expression of wound-induced genes. Shown are intact animals and trunks. * denotes annotation based on protein family domains (PFAM; Methods). Scale=100 µm.

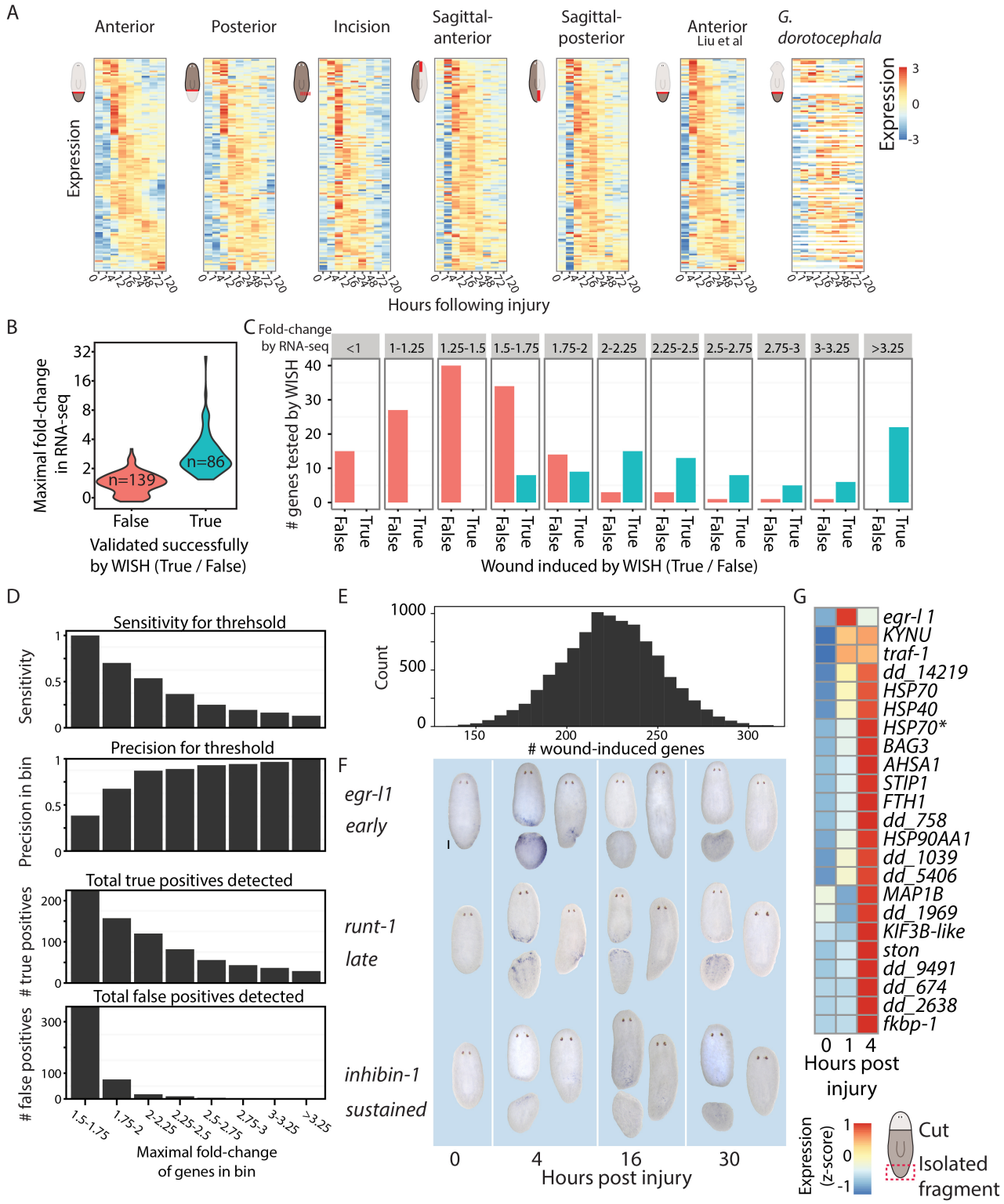


Figure S4, related to Figure 5. Extended time course analyses of distinct injuries. (A) Shown is the expression of 128 wound-induced genes in extended time-courses from multiple injuries. Each row represents a gene, and columns represent the time of isolation (hours post injury; 0-120 h). The colors are z-transformed value (minimal and maximal range was set to -3 and 3, respectively; blue, yellow and red colors correspond to low, medium, and high

expression, respectively). The ordering of the genes is identical in all heatmaps to facilitate comparisons, and furthermore, it is identical to the gene order in figure 5A. The rightmost heatmap presents regeneration timecourse from *G. dorotocephala*. Gene ordering of the orthologs found in *G. dorotocephala* for the 128 wound-induced genes, was retained. In case no ortholog was found, a blank line was plotted. (B) Violin plot summarizing the WISH analyses performed to estimate the sensitivity and precision of RNA-seq for detecting wound-induced genes (n represents the number of WISH analyses in the group it is plotted in). (C) Shown are bar-plots summarizing the number of true positive and false positive found through WISH analyses grouped according to their maximal change in expression (top title, linear scale) up-to 12 hours from amputation (Sampling performed at 0, 3, 6, 12 hpi in anterior and posterior amputations; Methods). (D) Summary of key parameters of false-discovery, sensitivity, and precision, obtained through comparisons of RNA-seq and WISH analyses. Shown are bar plots comparing different groups of genes tested by WISH that were binned by their maximal expression induction following wounding. Shown are estimates for the total number of true and false positives in each bin, through multiplying the sensitivity and precision by the total number of significantly overexpressed genes (FDR < 0.05). (E) Estimation of total number of wound-induced genes in the planarian genome by resampling analysis (n=10,000; extended experimental methods). (F) Shown are the full WISH images corresponding to the fragments displayed in Fig 5E. (G) Heatmap of wound-induced genes that were found by analyzing tissues far from the wound site (row z-score; Table S4).

Extended experimental procedures

Gene cloning and transformation

Genes were amplified from planarian cDNAs using gene-specific primers and cloned into pGEM vector according the manufacturer's protocol (Promega). Vectors were transformed into *E. coli* DH10B by the heat-shock method as follows: 20-100 µl of bacteria were mixed with 10 µl of pGEM vector cloned products and incubated on ice for 30 minutes, and then put at 42°C for 1 minute. The mixtures were then supplemented with 100 ul of SOC medium and following 1 h incubation at 37°C, were plated on agarose plates containing 1:500 carbacyclin, 1:200 Isopropylthio-b-D-galactoside (IPTG), and 1:625 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). Colonies were grown overnight at 37°C and white colonies were screened by colony PCR using primer sequences M13F (GTAAAACGACGGCCAGT) and M13R (CAGGAAACAGCTATGAC) with the following PCR program: a. 5 minutes at 95°C; b. 34 cycles of 45 sec at 95°C, 60 sec at 55°C, and 2:30 minutes at 72°C; c. 10 minutes at 72°C; d. hold at 4°C. Reactions were analyzed by gel-electrophoresis and for each gene a colony showing the correctly sized gene product was transferred to Luria Broth media (LB) supplemented with 1:500 carbacyclin for overnight incubation at 37°C. Plasmids were purified from overnight cultures with the QIAprep Spin Miniprep Kit (CAT #27106; Qiagen). Cloned genes sequences were validated by Sanger-sequencing (Genewiz, Inc.).

Double-stranded RNA synthesis for RNAi experiments

Double stranded RNA (dsRNA) was synthesized as previously described (Petersen and Reddien, 2008). Briefly, PCR templates of sequences for the forward and reverse of the target genes were prepared with a 5' flanking T7 promoter (TAATACGACTCACTATAGGG). The forward and reverse templates (4 ul) were mixed, in separate reactions, with 16 ul of 10 mM rNTPs (Promega); 1 ul of 100 uM dithiothreitol (DTT; Promega); 1 ul of thermostable Inorganic Pyrophosphatase (TIPP; New-England Biolabs); 0.5 ul of RNasin (Promega); 1.5 ul

of T7 polymerase; and 6 ul of 5x Transcription optimized buffer (Promega). Reactions were incubated for 4-12 h at 37°C and then supplemented with RNase-free DNase for 45 minutes. RNA was purified by phenol extraction followed by ethanol precipitation, and finally resuspended in 30 ul of MiliQ H2O. RNA was analyzed on 1% agarose gel, and quantified by Nanodrop (Thremoscientific) to have at least 5 ug/ul. RNA for forward and reverse strands were combined and annealed by heating the reactions in a thermo-cycler to 90°C and lowering gradually the temperature to 20°C.

Planarian dsRNA feedings

Animals were starved for at least 10 days prior to the first feeding. dsRNA was mixed 1:3 with 100% homogenized beef liver, and supplemented with 1 ul of red food coloring. Animals were kept in dark for at least 2 h before feeding, and were then taken out of the dark and fed the dsRNA-liver mix for at least 2 h. Animals uptake of the food was evaluated by the red coloring of the gut branches. Following a feeding, the culture plates and water were replaced and worms were kept in the dark; water in plates was replaced the day following a feeding as well, and every 3 days, unless another feeding was done.

RNAi feeding protocol

Worms were fed with liver containing dsRNA every three or four days. Three days following the fourth feeding animals were cut to three fragments, and the trunks were immediately soaked in 100 ul of planarian water supplemented with dsRNA against the target gene for 6 hours in the dark. Then, animals were washed and trunks were kept in the dark for 9 days before being fed with liver containing dsRNA against the target gene (booster). Then, 3 days following the booster feeding, trunks were cut to 3 fragments and soaked in planarian water containing dsRNA for 6 hours. Regenerating fragments were screened for defects every other day.

Illumina library preparation for anterior and posterior time-courses

Control prepharyngeal fragments (0 hpi) were isolated in biological triplicates and placed in TRIzol Reagent (Life Technologies) within 5 minutes from tissue isolation. Anterior-facing or posterior-facing wounds were amputated as pre-pharyngeal fragments as follows: A first cut was done either in the anterior or posterior end of the pre-pharyngeal region, and, at a given time-point (3, 6, or 12 hpi), a second cut was done to the opposite end of the pre-pharyngeal region. Prepharyngeal fragments were placed in biological triplicates in TRIzol (Life Technologies). Total RNA was purified following manufacturer's instructions (Life Technologies), followed by a second chloroform extraction to remove residual phenol contamination. Libraries (total 21) were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina) and were sequenced on Illumina HiSeq 2000 sequencer (Illumina).

Illumina library preparation for far libraries

Animals were amputated prepharyngeally and were placed in planarian water. Following a recovery period (0, 1, or 4 h) tail fragments were isolated and put immediately in TRIzol (Life Technologies). Total RNA was purified and sequencing libraries were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina).

Illumina library preparation for extended time courses

Tissues were isolated and placed in TRIzol (Life Technologies) for RNA extraction as previously described (Liu et al., 2013). Briefly, animals were cut as: postpharyngeally for (1) anterior- and (2) posterior- regeneration time courses; sagittally for (3) anterior and (4) posterior sagittal time-course; a postpharyngeal incision for the (5) incision time-course; and postpharyngeal amputation on *G. dorotocephala* for (6) anterior regeneration. Then, animals were put in planarian water for recovery. At each of the time points (1, 4, 12, 16, 24, 48, 72, 120 hpi) at least 8 animals were killed in 1% HCl for 1 minute, followed by 2 washes in

phosphate buffered saline (PBS). Animals were then put in RNALater (Life Technologies) and wound-sites were isolated on a cold block and put immediately in TRIzol as previously described (Liu et al., 2013). Uninjured fragments were isolated similarly, with the exception that the animals were killed and put in RNALater before isolation of the desired fragment. RNA was extracted according to manufacturers' instructions following tissue lysis in TRIzol with TissueLyser II (Qiagen; 2 minutes at 20Hz, followed by 2 minutes at 30Hz). RNA concentration was measured with Qubit 2.0 Fluorometer (Life technologies). At least 500 ng of purified RNA was used for strand-specific Illumina RNA-sequencing library construction as previously described (Engreitz et al., 2014; Schwartz et al., 2014). Briefly, for each sample poly-adenylated RNA was purified by two rounds of polyA selection (Dynabeads mRNA Purification Kit; Life Technologies) and eluted in 18 ul of H_2O. RNA was then fragmented with RNA Fragmentation reagent for 2 minutes (AM8740; Ambion) and purified on paramagnetic beads (Dynabeads Silane MyOne; Life Technologies 37002D). Then, RNA was incubated with 2U of Turbo DNase (Life Technologies) for 30 minutes followed by addition of FastAP for 10 minutes (Life Technologies). Then, RNA was ligated with an RNA oligo corresponding to a truncated 3' Illumina adapter (AGAUCGGAAGAGCACACGUC; IDT) using T4 RNA ligase 1 (36 units; NEB), and reverse transcribed with a specific primer (AGACGTGTGCTCTTCCG; IDT) with AffinityScript reverse transcriptase (Agilent). Following cDNA synthesis, primers were removed by adding ExoSAP-IT (Affymetrix) directly to the mix. RNA was degraded by adding NaOH, and cDNA was isolated and eluted in H_2O by paramagnetic beads clean-up. The 3' of the cDNA was ligated with a truncated 5' adapter (AGATCGGAAGAGCGTCGTGTAG; IDT), and the cDNAs were amplified for 12 cycles with barcoded Illumina primers.

Sequencing read mapping

Sequencing reads from each library were mapped to the *S. mediterranea* dd_Smed_v4 transcriptome assembly (Liu et al., 2013) with 5 additional sequences (sequences listed below) using Novoalign v2.08.02 with parameters [-o SAM -r Random]. The resulting

Sequence Alignment/Map files (SAM) were converted to sorted BAM format with the samtools v1.1 (Li et al., 2009) command [samtools view -T dd_Smed_v4 -bS IN| samtools sort – OUT] where dd_Smed_v4 is the fasta file of the assembly (http://planmine.mpi-cbg.de/), IN is the name of the SAM file, and OUT is the name of the sorted BAM file. Mapping statistics for each library were calculated by the samtools flagstat, and were examined manually. Read count per contig was calculated by bedtools v2.20.1 (Quinlan and Hall, 2010) using the coverageBed command [-abam IN -b BED > OUT] where IN is an input sorted BAM file; BED is a bed formatted file with all contigs in the assembly and their lengths; and OUT is the resulting read-counting coverage file.

Detection of differentially expressed wound-induced genes

Coverage files from the high-resolution wound-response time courses (0, 3, 6, and 12 h following anterior or posterior amputation) were consolidated to a read count matrix. The expression matrix was filtered for contigs longer than 450 base-pairs (bp). Following TMM data-transformation with edgeR v3.6.8 (Robinson et al., 2010), low expressing contigs were filtered with a cut-off of CPM of 6 in at least 2 out of 21 libraries. Next, differentially expressed genes were determined by the exactTest function by comparing each time point in the two time courses, separately, to the expression at time 0. Following hypothesis testing the p-values were corrected for multiple testing with false discovery rate (FDR). FDR smaller or equal to 0.05 and a fold-change of 2 or more were set as thresholds for determining wound induction. Genes that were found to be upregulated in at least one time point were included in the wound-induced genes list, except for contigs dd_Smed_v4_9491_0_1, dd_Smed_v4_14725_0_1, and dd_Smed_v4_1071_0_1 that were not validated by WISH (Table S3).

Detection of genes with putative asymmetric expression

Expression levels of wound-induced genes were compared between matching time points in the anterior and posterior time courses using the edgeR exactTest function. Genes with

corrected FDR of 0.1 or less, or exhibiting a fold change of at least 1.5 were selected for WISH validation, as well as 50 additional wound-induced genes exhibiting smaller differences in expression or less significant FDR between injuries. WISH validation of the genes was done on (i) intact animals and (ii) trunks of amputated animals that were cut and subsequently fixed at the time-point exhibiting the largest difference in expression between anterior and posterior wound sites.

Estimating the sensitivity of the wound-induced gene detection

Recent surveys of wound-induced gene expression in planaria yielded very partially overlapping results (Kao et al., 2013; Sandmann et al., 2011; Wenemoser et al., 2012), reflecting different instrumentation, analytical methods, and experimental setup. WISH was performed on 225 different genes on intact and amputated animals that covered a wide range of expression changes and FDR following wounding, including 46 negative controls (fold change; FC; and FDR, 0 to 28.6 and 0 to 1, respectively; Fig S3B; Table S5). 38% (86/225) of all of the tested genes were detectibly wound-induced by WISH. None of the genes with maximal FC less than 1.5 (n=82) could be validated by WISH regardless of their FDR. Furthermore, genes with maximal FC between 1.5 and 2 could be called with a precision of only 26% (17/65; Table S4; Fig S4B-D). By contrast, 88% (69/78) of the genes with FC > 2 could be validated by the WISH analysis (Table S4-5; Fig S4B-D). Therefore, 2-fold overexpression in at least one time point following wounding was used as a threshold for calling wound-induced expression. This threshold balanced sensitivity (57%) with precision (88%) compared to alternative thresholds (Fig S4C-D). Estimation of the total number of wound-induced genes was done by sampling 50% of the differentially-expressed genes according to thresholds [FC> 1.5; FDR < 0.05; minimal CPM 6 in at least 2 libraries] 10,000 times. For each sample the total number of wound-induced genes was estimated by multiplying the number of genes in an expression bin by the fraction of genes that were wound induced, as detected by the WISH validated genes in the sample. Estimations were

multiplied by 2 to correct for the sample size. The total number of wound-induced genes was the average of the individual estimations.

Single-cell isolation and Fluorescence-Activated Cell Sorting

Wound sites were collected from post-pharyngeally amputated animals 4-6h or 12-14h following an amputation; control cells were collected from the same region in intact animals and were processed immediately. Cell suspension was prepared and was subjected to FACS as previously described (Hayashi et al., 2006; Reddien et al., 2005). Briefly, isolated tissues were put in 450 ul of CMFB (calcium magnesium free buffer + 1% BSA) with 50 ul of collagenase and incubated at room temperature while gently pipetting the samples. Samples were then filtered through a 40 um filter into CMFB. Samples were span down and re-suspended in 500 ul of CFMB. To each sample, 20 ul of Hoechst was added and incubated in the dark for 45 minutes, followed by addition of 1 ul of propidium iodide. Negative controls devoid of either Hoechst, PI, or both were prepared in parallel. Single cells were sorted to 96-well microplates containing 5 ul Buffer TCL (Qiagen) + 1% 2-mercaptoethanol. Plates were incubated for 5 minutes at room temperature and were then placed on dry ice.

Single-cell sequencing library construction

RNA-sequencing libraries were prepared from Single sorted-cells as previously described (Picelli et al., 2013; Picelli et al., 2014) with few modifications. Each well in a 96-well microplate was supplemented with x2.2 (11 ul) of Ampure XP beads (Agencourt) and incubated for 10 minutes at room temperature, and then put on a 96-well magnet plate (Dynamag 96-side magnet; Life Technologies) for 5 minutes. Supernatant was removed and beads were washed twice with 100 ul of 80% EtOH. EtOH was removed and beads were air-dried for 10 minutes before elution of the beads in a mixture of 1 ul of reverse transcription primer (5′-AAGCAGTGGTATCAACGCAGAGTACT(30)VN-3′, IDT DNA), 1 ul of dNTP mix (10 mM), 0.1 ul of SUPERase RNase-inhibitor (40 U/ul; Life Technologies #AM2696), and 1.9 ul of H2O. The plate was incubated at 72°C for 3 minutes and placed immediately on ice. Each

well was supplemented with 7 ul of a mixture consisting of 1.65 ul H_2O, 2 ul of 5x Maxima reverse-transcription buffer (Thermo-Fischer), 0.9 ul MgCl2 (100mM Sigma-Aldrich; M1028), 2 ul of Betaine (5M; Sigma-Aldrich; B0300-5VL), 0.25 ul of SUPERase RNase-inhibitor (40 U/ul), 0.1 ul of Maxima RNase H- RT (200 U/µL; Thermo-Fischer, EP0753), and 0.1 template switching-oligo (Exiqon; 100 uM; AAGCAGTGGTATCAACGCAGAGTACrGrG+G; r and "+" denote RNA and LNA bases, respectively). Plate was briefly span-down and incubated as follows: 42°C for 90 minutes, followed by 10 cycles of (50°C for 2 minutes, 42°C for 2 minutes), followed by 70°C for 15 min. Following reverse-transcription a pre-amplification mix of 14 ul was added to each well [1 ul of H2O; 0.5 ul of PCR primer (10 uM; 5'-AAGCAGTGGTATCAACGCAGAGT-3′), and 12.5 ul of KAPA HiFi HotStart ReadyMix (Kapa Biosystems; KK2601)]. The cDNA was amplified using the following program: 98°C for 3 min; 20 cycles of (98°C for 15 sec, 67°C for 20 sec, 72°C for 6 min); 72°C for 5 min; hold at 4°C. Following pre-amplification PCR products were purified using x0.8 Ampure XP beads, and eluted in 20 uL of H2O. Amplified cDNA concentrations were measured using Qubit HS-DNA reagents (Life Technologies). Samples were diluted to 0.2 ng/ul, and sequencing libraries were prepared using the Nextera XT library kit (Illumina). For each sample, 1.25 ul of amplified cDNA was combined with 2.5 ul of tagmentation DNA buffer and 1.25 ul of the amplicon tagmentation mix. Samples were mixed and put at 55°C for 10 minutes. Samples were chilled on ice, and 1.25 ul of neutralize tagment buffer was added for 5 minutes incubation at room temperature. An amplification mix was added as follows: 3.75 ul of Nextera PCR mastermix and 1.25 ul of two barcoded amplification primers. The samples were amplified with the following PCR program: 72°C for 3 minutes; 95°C for 30 seconds; 12 cycles of (95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 1 minute); 72°C for 5 minutes; hold 4°C. Following amplification, 2.5 ul of each sample were pooled in groups of 32-96 samples, purified with x0.9 Ampure XP beads, and eluted from beads in 25-50 ul of H_2O.

Single-cell data mapping and clustering

Following Illumina sequencing, cells were eliminated from further analysis if they were found to be contaminated by more than 25% of non-planarian DNA, as detected by mapping to human, mouse, rat, yeast, and *E. coli* genomes. The sequencing reads from all other cells were mapped to the dd_Smed_v4 assembly (Liu et al., 2013) with Novoalign v2.08.02 and the number of reads for each contig was calculated as described above. Following mapping, reads mapped to contigs dd_Smed_v4_10881_0_1 and dd_Smed_v4_5614_0_1 were excluded, as they represented misalignments of primer amplification sequences to the planarian transcriptome. Samples having reads mapped to less than 1,000 or more than 9000 contigs were eliminated from subsequent analyses. Data was analyzed using the single-cell analysis Seurat method (Satija et al., 2015). Briefly, genes showing CPM expression of more than 2^4 and a dispersion of 1.5, were selected for initial PCA. Next, principal components that significantly separated cells were determined by a jackstraw analysis (Chung and Storey, 2015) by running the function jackstraw [num.pc = 15, num.replicate = 100, prop.freq = 0.03]. Based on the analysis PCs 1 through 4 were selected, and the list of genes used for the Seurat analysis was expanded by using the function pca.sig.genes [pcs.use = c(1:4), pval.cut = 1e-5]. Then, a second round of PCA was performed with the expanded list of genes, and a jackstraw analysis determined the significant genes in each PC by running the jackStraw function [num.pc = 15, num.replicate = 100]. The top 50 genes contributing to the variance in PCs 1 through 15 were examined manually to identify technical biases (Satija et al., 2015). PC 3 was eliminated from gene selection because most of its highly variable genes consisted of ribosomal proteins, which indicated that it represented technical difference between cells. Dimensional reduction was performed by t-SNE using the run_tnse function [pcs.use = c(1:2, 4:12), max_iter=500, perplexity=20]. Cells were clustered together by the Mclust_dimension function with parameters [reduction.use="tsne", G.use=3.2, set.stat=TRUE, MinPts=3]. Clusters 5 and 14 were

eliminated since they grouped cells suffering from low complexity, based on number of expressed genes. Cell-specific markers were found by using the find_all_markers function with parameters [thresh.test=4, test.use='roc', return.thresh=0.7]. Clusters having the same markers were merged [9 and 24; 4, 5, and 20; 7 and 8; 12, 19 and 23; 13, 16, and 10; 19, 21, 22, 25, and 27; 2, 6 and 26. Cluster 6 cells showing high expression (CPM > 1024) *prog1* and *prog2* (Eisenhoffer et al., 2008) were consolidated with cluster 18].

Detection of differentially expressed genes between clusters

Differentially expressed genes in clusters were detected by running the Seurat find_all_markers function [thresh.test=2] and area-under-the-curve for each gene was calculated with find_all_markers [thresh.test=2 , test.use='roc', return.thresh=0.5]. In addition, clusters expressing the same canonical cell-type markers (e.g., *smedwi-1* or *synapsin*) were temporarily merged, as they might reflect functional relationships of different clusters (e.g. subtypes of a major class of cells). Genes enriched in the merged clusters were found by comparing them to all other clusters by running find.markers [stat.1 = MERGED, stat.2 = OTHER] where MERGED stands for the merged clusters and OTHER stands for cells that were not in the merged clusters. P-values were corrected using the R function p.adjust with default parameters.

Detection of cell-type-specific wound-induced genes

Cell-type-specific wound-induced genes were determined by three analyses. First, the gene expression of wound-induced genes from cells derived from intact animals was compared with the gene expression of cells from wounded-animals using the bimodal expression hypothesis testing with parameters [FDR ≤ 1E-7] (Shalek et al., 2014). Second, gene expression of cells from the wounded time-points was contrasted between a cell type and all other cells with the following parameters pairs using the bimodal expression hypothesis (McDavid et al., 2013) [log FC ≥ 2, FDR ≤ 0.001; FDR ≤ 1E-7]. Finally, adjacent clusters on the t-SNE plot were combined, and the hypothesis testing was repeated. Particularly, late

epidermal lineage and epidermal cell-types were combined; neuronal types; and neoblast subpopulations.

Clustering of unwounded 4C isolated cells

Expression matrix from uninjured neoblasts (n=90) was generated. The *Seurat* method was applied with the following parameters [min.cells = 10, min.genes = 4000, calc.noise=FALSE, is.expr=0.01, do.scale = TRUE]. Cells expressing more than 9000 genes were discarded from further analysis. Gene selection was performed as previously described with the following parameters: mean.var.plot [y.cutoff = 1.5, x.low.cutoff = 5]; jackstraw [num.pc = 6, num.replicate = 100, prop.freq = 0.03]; pca.sig.genes [pcs.use = 1, pval.cut = 1e-3]; run_tsne [pcs.use = c(1,2,3), max_iter=500, perplexity=23]; Mclust_dimension [G.use = 25].

Transcriptome assembly of *G. dorotocephala*

Sequencing reads from all samples from *G. dorotocephala* were combined. Adapter sequences were trimmed with trimmomatic (v0.30) (Bolger et al., 2014) with the following parameters [LEADING:3 TRAILING:3 ILLUMINACLIP:TruSeq2-PE.fa:2:40:15 SLIDINGWINDOW:4:15 MINLEN:30]. Reads containing long stretches of polyA or polyT (>20 nt) were removed using a grep command. Assembly was performed with trinity (release r20131110) (Haas et al., 2013) with the following parameters [--seqType fq --JM 100G --output dor_PE --CPU 6 --min_contig_length 200 --SS_lib_type FR]. Redundant sequences were detected and removed by running cd-hit-est (v4.5.4) (Fu et al., 2012) on the output file with default parameters. Orthologs with *S. mediterranea* were identified with proteinortho (Lechner et al., 2011) with parameters [-p=blastp+ -e=1e-7 -cov=0.35 -pairs –singles].

Detection of onset and offset of wound induction

Expression data from each time course were used for fitting by the *impulse* model (Chechik and Koller, 2009; Chechik et al., 2008) using a published Matlab implementation (Sivriver et al., 2011). Fit for each wound-induced gene was produced for function fit_impulse_params_constrained [expression [log2(CPM+1], retries=100, time points = (0,

1, 4, 12, 16, 24, 48, 72, 120) for all time courses except for the head regeneration time course by Liu et al (Liu et al., 2013) that lacked 1 hpi time point and was called with (0,4,12,16,24,48,72,120)]. Internally, the fit function was called with constraint parameters [t1 ≥ 0; t2 ≥ 0; h0 ≥ 0; h1 ≥ 0; h2 ≥ 0; β1 ≥ 0; β2 ≤ 0]. Expression values for every time point within the 0-120 hpi range was extracted using the impulse function [fit parameter output, time point 0-120]. Genes used for onset of anterior regeneration analysis were collected from previous publications (Gurley et al., 2010; Reddien, 2011; Scimone et al., 2014; van Wolfswinkel et al., 2014; Vogg et al., 2014), and filtered for extremely lowly expressed genes [minimal expression 2 and standard deviation of 0.3]. A median group fit was produced by using median z-score values in each class [Genes associated with specialized neoblasts, patterning factors, and differentiated tissue markers] and a using baseline value for the three classes at 0 h time point as -1.5; to allow convergence 4 hpi samples were corrected by subtraction of 0.05-0.1, as for regeneration related genes their expression was almost identical to the 12 hpi gene expression. Fit function was called with constraint parameters [t1 ≥ 0; t2 ≥ 0; h0 ≥ 0; h1 ≥ 0; h2 ≥ 0; β1 ≥ 0; β2 ≤ 0]. Following filtering the following contigs were used for *S. mediterranea*: dd_Smed_v4_11372_0_1, dd_Smed_v4_11521_0_1, dd_Smed_v4_13056_0_1, dd_Smed_v4_13215_0_1, dd_Smed_v4_13898_0_1, dd_Smed_v4_14611_0_1, dd_Smed_v4_14633_0_1, dd_Smed_v4_15104_0_1, dd_Smed_v4_15144_0_1, dd_Smed_v4_15178_0_1, dd_Smed_v4_15253_0_1, dd_Smed_v4_15516_0_1, dd_Smed_v4_16375_0_1, dd_Smed_v4_17385_0_1, dd_Smed_v4_17726_0_1, dd_Smed_v4_17731_0_1, dd_Smed_v4_21717_0_1, dd_Smed_v4_21801_0_1, dd_Smed_v4_856_0_1, dd_Smed_v4_9774_0_1, dd_Smed_v4_9893_0_1, dd_Smed_v4_11285_0_1, dd_Smed_v4_12674_0_1, dd_Smed_v4_13487_0_1, dd_Smed_v4_13985_0_1, dd_Smed_v4_15531_0_1, dd_Smed_v4_19866_0_1, dd_Smed_v4_5102_0_1, dd_Smed_v4_6604_0_1, dd_Smed_v4_8832_0_1, dd_Smed_v4_11968_0_1, dd_Smed_v4_12112_0_1,

dd_Smed_v4_12647_0_1, dd_Smed_v4_12653_0_1, dd_Smed_v4_14207_0_1, dd_Smed_v4_16476_0_1, dd_Smed_v4_16581_0_1, dd_Smed_v4_17854_0_1, dd_Smed_v4_20433_0_1, dd_Smed_v4_29533_0_1, dd_Smed_v4_3135_0_1, dd_Smed_v4_6710_0_1, dd_Smed_v4_8392_0_1. Orthologs were identified for *G. dorotocephala* as described above and the following contigs were used: comp14905_c0_seq1, comp25657_c0_seq1, comp28223_c0_seq1, comp28241_c0_seq1, comp28262_c0_seq1, comp28562_c0_seq1, comp28762_c0_seq1, comp29894_c0_seq1, comp29915_c0_seq1, comp30125_c0_seq1, comp30289_c0_seq1, comp31066_c0_seq1, comp31293_c0_seq1, comp31342_c0_seq1, comp31414_c0_seq1, comp32324_c0_seq1, comp37945_c0_seq1, comp3844_c0_seq1, comp4224_c0_seq1, comp5212_c0_seq2, comp8621_c0_seq1, comp8817_c0_seq1, comp27033_c0_seq1, comp27221_c0_seq1, comp27470_c0_seq1, comp28896_c0_seq1, comp3788_c0_seq1, comp4439_c0_seq2, comp5124_c0_seq1, comp5348_c0_seq1, comp7983_c0_seq1, comp17302_c0_seq2, comp25468_c0_seq1, comp29782_c0_seq1, comp29980_c0_seq1, comp31706_c0_seq1, comp32106_c0_seq1, comp43392_c0_seq1, comp5522_c0_seq1, comp6277_c0_seq1.

Primers used in this study

Contig	Primer A	Primer B	Primer C
dd_Smed_v4_10259_0_1	ACGCAGAGGCTTGCAGTT	TTGGTCTGTGTGCAGCCA	GCCACAAATTGCACCGCA
dd_Smed_v4_10337_0_1	AAAAGACGCGATGAGGCA	TGTCCTTTGCAATTTATTCGCGA	CAGCCAGGTCACAGTGGC
dd_Smed_v4_1039_0_1	TGTTTCGATTTCTAGACGAACCG	TTGGCCGGAATATTCTCATCA	GCATCACCACTTTCCACAGG
dd_Smed_v4_1054_0_1	CCGGAATTCACGGGCCAA	TGTAGAATGACTCGAATCTCGGA	TTGAGTGTCCGCTGCTCG
dd_Smed_v4_10569_0_1	CGCGTTCCCAATGACAGC	TGAAGGCGGTGTTCCTGAC	ACAGATAACCCTGCAAGATCCT
dd_Smed_v4_10584_0_1	CCGCCGTACAGTATCATGGA	ACCAATAGAGACAGTTCAGCCA	ACGAAATTGACAACGCTAGTGA
dd_Smed_v4_10624_0_1	ACGAGCCAATGTCCAGCC	TATGTGTTTACGAGTGCGATTTT	CACCGGGTGACGCATGAA
dd_Smed_v4_1071_0_1	ACGGGTCGACGTCAGTTG	TGCAACACAAATCGTAAACAGA	GTCCTGACGCACGAGGAA
dd_Smed_v4_10716_0_1	CGGTGAGCGGTGTGTGAT	TCGATTTCAGTTGCATTTGTGGA	TCCCGGCATACAAGAGCAC
dd_Smed_v4_10730_0_1	GCAATCAGCCAACTCGGG	TTTATTAAAGAACCCGAAAGCGT	TGGGGTGCCGGATACAGT
dd_Smed_v4_10776_0_1	GACATTTGGCAGTCCTTCCTG	CGAACTTGCTCCCGGACA	GGGGTATCTGATTATGACTGAGC
dd_Smed_v4_10868_0_1	TTGGGCTGCGGGATTTGG	GGAGCATTGATAAGTTGTTCTGT	TCGGCAACAAACTCCTCGA
dd_Smed_v4_1087_0_1	ACCAGAACCGGAAGCTCC	TGTCGCTTTCAATAAAGGCAAA	TGTTTGCTCACGTCCTCTCC
dd_Smed_v4_10927_0_1	ACAACGAATGGCAGAGTGAGT	TTTTGGAGTGTGTATGCATGAGA	CAACGCAGAGTTCTGTCAAAA
dd_Smed_v4_10930_0_1	ACCAAATTCTATGCAAAGTCGTT	ACACAGTGTTTTGGTTTCCACC	TGCGGCATTATATTTGCGGA
dd_Smed_v4_11074_0_1	CCGGCTGGTTCTGTCGAG	TCAATGAACATTATGGTCCCACC	CTCCCCGCATCGAAAGCA
dd_Smed_v4_11115_0_1	TGCCTAGAGACGACTGCTCT	TGCATTGAAATTCTGCCTTTGGT	TGCGGTGCTTGCTCATGA
dd_Smed_v4_11134_0_1	GGCCTTCTTAGCGATGCGA	ACTCTGCTCCACCACACAG	CTGGCGCTGACAATCCGA
dd_Smed_v4_11142_0_1	AGGCTTCACTGTCGGTTCG	TGTCCATGTGTTCACCAGTCA	TGTGACTCTGCGCTGACG
dd_Smed_v4_11216_0_1	CTCGAGCTGACGCGGAAA	TGACTGCGTCCATAGTGTTGA	TCTCCAAGGGGGTGCAGT
dd_Smed_v4_11220_0_1	AGGAACTTGAGGACATTTCCGC	GTTCTTCGGATAATTGTCCACCT	CAAATTTTCAATCCATCCCGACA
dd_Smed_v4_11254_0_1	TTCAATTTCATTCACGCATGTGG	TGACATTTTCGATCGTTGCGT	ATATCCTTGGCTTTGTACACTGA
dd_Smed_v4_11501_0_1	TGTCGCTCAATATGCAGGCT	TCGTGCTAACTTCCAGGGA	AATTCGACTTGCGGTGCC
dd_Smed_v4_11512_0_1	AACTCGTCTGTGCTGCGA	TCCCAGCGACATGATTGGT	TGGTGGGACATTCATAATGGC
dd_Smed_v4_11561_0_1	TGGGCAACTGCATTGGGA	CAACGAAAATCCCTCTAGCTCC	TCAAAGCTGCTTCGGGGG
dd_Smed_v4_11608_0_1	GGCCGATCAGTGCACCTT	ACGGAGAAATGTCCCCAGG	CGACTTGATGGGCCCACA
dd_Smed_v4_11629_0_1	TGCTTCCTCATTGGCGGA	GCTCCACATCCAAATGGGC	CCACATGCCATAAACACCCG
dd_Smed_v4_11635_0_1	GAGTGATCTAGCGATTTGATTGG	TCCTCGATGCCTATGGAAACT	ATTTTGCAATAGGCCCATCAGT
dd_Smed_v4_11693_0_1	CAGTGGATGGTTGCCGGT	AGCTGATCCAGAAATGCCTAA	TAGACGGGCTGTTCGGGT
dd_Smed_v4_11824_0_1	TGCTCTGTGGCACTGACG	TGTGAGAAACGCTACGATCAA	ATGTCGCTTCCCACCGTC
dd_Smed_v4_11858_0_1	TCACAGAAAACCCAGTCCCC	TGCAGTTTCAACAAAAGATTCCT	ACTATTTCGTTCAATGGACGACA
dd_Smed_v4_11943_0_1	ACACCATTCCATACGCCGA	TCCAATAACTCGAGCAATATGGT	TTGATTGAGGCCGCTGCA
dd_Smed_v4_12081_0_1	TGGAAACCAGGGGGCTTT	TGTCATCGTTTACTGTGGCT	ACGGTAAATGTGCGATGAACG
dd_Smed_v4_12210_0_1	TCGGACGCAGATTCAGAAACT	CCAATACACAAGCTTATGACACG	GGAATGGCTGTTCCGGGT
dd_Smed_v4_12467_0_1	GCAGTTTGCGATCTGTATTGC	TTGGAATCGACTGACGGAAG	TGGGTTTGCTGTAATTGGCA
dd_Smed_v4_12472_0_1	CCGTTCGATTATGCGGCC	CTCTGTACGGATATTCCCAATCA	CGTCACGCAATTCGACCG
dd_Smed_v4_12575_0_1	CCCCTCTACGAGAAATTGCTGT	TGGGCTAGCTTAATACTTTGCA	TCGGAGAAGGCGAATTCGG
dd_Smed_v4_12619_0_1	AGCATGTCAGGAGCTCGA	ACAATTACCACATCAATGGGACA	GGCTTTGGTTTAGGCTTTGGT
dd_Smed_v4_12634_0_1	GCAGGTCTTGAGGCAGCT	CCTGTCCATATAACACTGGAACA	TGTATCAGGGCAAACGAGTT
dd_Smed_v4_12695_0_1	CCATCGAGACCGCGTTGA	CAAATCGGTTTCGGAAAGTTTCA	TCGGCTGCTGTTTGCTGT
dd_Smed_v4_13056_0_1	ACAGTGGGCGATTTTCTCCT	TCTATGGATTCCCCGAAGTCC	TGACACCAAGGTTGAGGCA
dd_Smed_v4_13061_0_1	TGCAAAACAATACTAGCCAATGC	TGCGAAAGTTGTATCAATCCGT	TCGATGAAGTCATATTTCCCGT

dd_Smed_v4_13186_0_1	TCCCTGCCATTAGTACGACA	AATAGATCCGGATGAATTGCTTG	AGGAAAAGGGGGAGGGCC
dd_Smed_v4_13188_0_1	ATTGAAATTTCTTCACTGACGCT	TGTACTCGTCTATCGCTTGCA	GACTCTAAAATGGATGCCGAGC
dd_Smed_v4_13216_0_1	AAACTGCCGCGACGAAGA	TGTTTGGTGAAATGTTAGAGCAA	CGGCGGACTATGACCTCG
dd_Smed_v4_13318_0_1	CAAGTGGTGTTACATTTTCAGCA	TCAAAGGCCAAATTCTGCCT	TGACATCAATTAGCCCTGGAAA
dd_Smed_v4_13356_0_1	TCCAACTTGAACCATGTCGGA	GTCCAATTCGATTGTGAACCGA	TGTTGCAGTGGGGCTCAG
dd_Smed_v4_13468_0_1	TCCAAGTGGATTCGGGCA	TGGACGAAAATGACAATTCTCCT	AGGAGCATTGTCGTTGGCA
dd_Smed_v4_13487_0_1	ACGCGTGACTGAGTTGGT	TCGGACTACCCCATTTGCAG	TGATTGTTGAGATTGGCGAGT
dd_Smed_v4_13835_0_1	TGACTGCCAGTGTGTTATCAGA	ACACGAATTGGTTGGATCAAACC	TCCACAGAATTGCGAATCCCA
dd_Smed_v4_13843_0_1	GTAAACGGGACCTCGCCA	AGAAAGTTCAACGCAAGATCAGT	TGTCGAATCTTGCGCCCA
dd_Smed_v4_13860_0_1	CGGTTGATCTGCAATACCGC	CGTTCTCGATTGTGATAGAAAGG	TGTTGGTCAGATACAAGTGCGA
dd_Smed_v4_13985_0_1	TGACCAAGATTTTTCCCCTAAGT	TCATTGGAGATTGGCAAGCA	GGCAGACCGATTGTTGGGT
dd_Smed_v4_14011_0_1	ACTTCTCAACTGTTCAAAATGCA	TTCACTTCGGCATTTGCAACT	AGGTTTAAAACAAAAGCTGCCT
dd_Smed_v4_14068_0_1	TTTGGAACATTTTACGAGAACCG	ACTATAGCGGAAGTTTAATCGGA	TCTTAACAGCTACATGTGCAAGA
dd_Smed_v4_14158_0_1	GCCGAATGTTCATTCAAACCG	TGTCATTTTCAGTAAAAACGGCA	TCGAAAAATTTGCCGACAAGA
dd_Smed_v4_14199_0_1	GCCTTAATCGACGTGTTTGGA	CGGTTCCTCAGATTCCGAGA	TCTTGTTCAAAACGGAGGAACA
dd_Smed_v4_14370_0_1	TGATGCGGCTATTGTTGATTTT	TGCGCTTCCATTTTACCAGC	ACTGTTACGCAACAAAATAAGGT
dd_Smed_v4_14391_0_1	GGCTTCAAAGGCCACGGT	ACCTTTGCTGACAGGAGATGG	CCTCGTCATCAAGTCGTCGA
dd_Smed_v4_14392_0_2	TGTCTCAAACAGAAGTTCGTCAG	TCGTCGATTGAAAGAAATGACCT	GATGGGCGGCCGTATGAA
dd_Smed_v4_14656_0_1	TCGACCCGAAAATGTGTTTGC	TGTTCAGACCCAAGCTACCG	ACCATTTGAAACGTTCAGAAGTT
dd_Smed_v4_14711_0_1	TCAGACTGGATATACCCCATTGC	TGCCGGGAATTCATGAATCG	ATGATTTTGTCTGAAATGTCGCA
dd_Smed_v4_14725_0_1	CCCATTGTCTTTATTGCAAGGCT	CAGAAAATGCAGGAGCTCTGA	GCCAGCCATTTCAGCGAC
dd_Smed_v4_15035_0_1	CGCTGATTCCCAAGCGGA	TGCACTCACTAAAGGTACAGAA	ACAACACGAATTTGTGCAAAACA
dd_Smed_v4_15386_0_1	CGGCCGAAAGAGTCTCCA	CCGATTGACAGTGCGTATTCA	CGCTGTCGGTGTTGTCGA
dd_Smed_v4_15499_0_1	TGGTTTAGATGCGGTTCCAT	GCCCTGTTAGAAATTTATCCCGA	TGCTTCGCAGCCTACGTC
dd_Smed_v4_15531_0_1	GTTGGCCTCTCATCCAGCA	TCCGACAATTATCCGCCTGA	CCCTGTTACCGAGCCTGAC
dd_Smed_v4_15647_0_1	TCACTTATAAAGGCCGCCCA	TTTGCTTCTAGATGAGGTCTGCA	CAAAGCCCACCACTCGAGA
dd_Smed_v4_15715_0_1	TGTGAAACTGTAACCTTGTTCTG	TGATTCTCCATCTCTAGACTCCA	AAACCACTACGTTCCCAAACA
dd_Smed_v4_15787_0_1	GCCATCCCAGATGCCTCC	TGCCAGCATTACCACAGATT	ACGGCTGCTTTGACCTCC
dd_Smed_v4_158_0_1	TGCTGCAACTTCTTCGCA	GCCTCTTCAATAACTTCAGCAGC	CTCCGCTGATCAATCACCGA
dd_Smed_v4_1580_0_1	GTTGGTGAAGGCCATCCAGA	AGTGATGCCATTCTAGATGCACA	TGAGGCACTTGCTGAACGT
dd_Smed_v4_1581_0_1	CTCGGACTTGGGTCTGCC	AGGAAACGATCGTGGATGACT	GGTCACACTCTCTGCACGT
dd_Smed_v4_16092_0_1	TGCCGAAAAACGCAAGCA	TGCAGTAGACTCGAAACCAAA	ACCAAAGCAGGAGAGGAAGG
dd_Smed_v4_16209_0_1	TTTGCAGGCTTCGACCAA	CTGTTTGGATTTCTGTGGCGA	GTCCTCGACCGCAACACA
dd_Smed_v4_16222_0_1	CCAGCGATTAATTGTGTCGAACA	CGGTTCAACGGTTTCAGCA	TGATTTCTTTTACGGGGCTCCT
dd_Smed_v4_16227_0_1	GGTCGGTTTTTCCATCGTGG	AGCTCTCAACCTCAAGATCTACA	CGTCGACGTCTTGTGAGGT
dd_Smed_v4_16605_0_1	TTGGCTTTACGTTGGCATTTCT	CTTTCATGTGTATTGGCTGTGAT	AGTCGAAGTGGTCAACGCA
dd_Smed_v4_16842_0_1	AGCGTCCTTTGCGAGACA	CCTCAACTCCAAATGCTAAAACA	GGACCAGCTCATGACCCG
dd_Smed_v4_17385_0_1	TGGAACGCTATAAGTCGGTGA	TGGCGGTTCACATTTCCA	TCGGACCGATTGAAGCGT
dd_Smed_v4_17402_0_1	CGGATAGCGAATACAATTGATGC	ACTCACACAAATAATTGATGCCA	CCATCGGGAAAGCAATTGTCC
dd_Smed_v4_1771_0_1	TTCCTTTACACCGTCCTTTGT	TTGTCACCACAATGGATATCCCG	TCCATATGTTATGAATGGAGGCA
dd_Smed_v4_17726_0_1	GCAAGAAAACCGGCAGGG	CGAGTGATCCTGGAAACATTGC	ACTCCGGAGCGAGACCAT
dd_Smed_v4_1846_0_1	ATGGAACCGCAGCAAGCT	TCAAATGTGGCATGGATTTTCGT	GTCGACAGGGCCACTTGT
dd_Smed_v4_18818_0_1	GCGCTTGTTAATCTGGTCCC	AAGAGTGAAATCAAAATCGCGT	TGGAAAAACCAGCTACAATTCCA
dd_Smed_v4_1921_0_1	TTATCGGCAGTGTCGCCC	TCCTTTATTTTGGCGAGGCA	ACTATGGAGCAATACGCAGGA

dd_Smed_v4_19428_0_1	CCGAAGACGATTTGCAACGT	TGCCATCGGAATTACAGGCT	ACAGTTAGGCCATACTCAAATGA
dd_Smed_v4_19592_0_1	ACTCGGGTTTAAATGCACCAC	ACCAGTGTGACTATCTTTTGTGC	CGGCGATTGGCTGCTTCT
dd_Smed_v4_19826_0_1	CGACAATCGGCCTGAGGT	TGACATATTCGAAAACCAACCTC	AATGGGAATCACGGCGCA
dd_Smed_v4_1986_0_1	GCCGCTGGATCTTTTTGCA	TCTGCATAGCGGGATCACT	AGATCCGCGGCTTTTTGT
dd_Smed_v4_1999_0_1	TGATCGCCACTCCGAACG	CCTGATCGAAGCAGTTCCAGA	TGTCGTAGGAGGACGCCA
dd_Smed_v4_20048_0_1	TCATCGGAAAATCACCTGCT	ATCAGAAACCTGTCCAATGGT	TGTCAGGCTGAATGGTCGG
dd_Smed_v4_20122_0_1	ATTACTTCCGCCGAGAGAAGT	TCATTGGAAATCGACATGAGACA	AGTCATTTTCAACATGAACGGCG
dd_Smed_v4_20133_0_1	CGGCCGATCTCAGCCAAT	GGATTGAAAGCCGCGAAATCA	GCTTCAACAACGCGTCCA
dd_Smed_v4_20318_0_1	TGAATGCCCAATGGTCGCA	TCGAAGAGAGAGTAGAACGAGC	TGGACGCAAGCACTGTCC
dd_Smed_v4_21069_0_1	TGTGGCAATTGCATGGTGT	TGGCTGAAACAAGTCAAATCCG	CGACAAGTCGCAACATTTGT
dd_Smed_v4_21717_0_1	TGACCACTTCATCTGTTGACA	AGGGCCAAAGAAGAAGCCG	AGTGCACATGGAAATGGACCT
dd_Smed_v4_2172_0_1	AGAAGGAATCGGACTGTTTGGA	TGAGAGACCAAGTGACAAAGAA	TGGAATGGCCAAGGCAGA
dd_Smed_v4_22031_0_1	TCGTTTCTTGGGCAGTCGA	ACTCTCTCAGCAATTTTGAGTGA	TGCGGCTGCTGGGTAAAA
dd_Smed_v4_22061_0_1	AGATTTTGACATATGTTGCCTCG	TCGATGTCTCCTTCATCAGACG	AACTTTGACACAACCACAAGAGA
dd_Smed_v4_22479_0_1	TCACAGCGATGTGGAAGACA	AGCAACAATCCAGAACTCGA	GGAGCGGAAGGGAGGAGA
dd_Smed_v4_22918_0_1	TCAAGTTGCGAGGCCTTGT	TGCCAAATATGTACAGCAACGA	AGCCTAATGAATGAGTCGAAAGT
dd_Smed_v4_2324_0_1	GCGCCACCACTGTATCGA	CAGCTATCAGATGGTCAAAGTCA	GTGTTGCTGGACCCTGCT
dd_Smed_v4_23420_0_1	TCCAACTGTGTAAATGGGGTGA	TTTCTTGAAAGTTGCGTCCCG	TCCTCACAACAAGAAAACGGA
dd_Smed_v4_23666_0_1	TCTCCAACAATCTCCATCCGT	TCGGCTTTGGAAAACCGA	AGGAATCTACCGAAATCCTTCAA
dd_Smed_v4_2394_0_1	TGGAATGCCAACATTTTCTCAA	GAACCCTTCAAGATACCATGACA	GCGAATAAAAGGAAGTACTGAGC
dd_Smed_v4_24180_0_1	TGAATGATCCGCAATCCAGT	AACGTTCGCTGCAATGACG	TGAGATACCCAACGATTTCGCA
dd_Smed_v4_2442_0_1	GCTCACTGAGTTTGCGTATGC	TGACAAGTCTTCCCAGAATTCCT	GGTTTCAATGATCATTGTTGGCG
dd_Smed_v4_246_0_1	TGCACACAACCTCATGAGCA	GGTAGATCGTTCTGCAATGCA	AGTTGAACCTCCAGACAACACA
dd_Smed_v4_2575_0_1	TGGAAATTCGCATTGTTGTTGT	TCTGCAGTTCTCGCCGAT	TGTCAATCATCCAAGTCTGACA
dd_Smed_v4_2582_1_1	TCCAAGGAGGGAATGGTGGA	TGTACACGAACTGGGCGG	ACGACAAGATAACCGCTCACA
dd_Smed_v4_2588_0_1	ATGGCAGCCGGTGATGTT	AGCTATGCGAGGAAACTATTGA	TCAAATCCCAATCCTGATCGT
dd_Smed_v4_26705_0_1	TGCCTTCTTTTTCGGTGGA	TCATGTTTGTCTTTTGTCAACGA	GGTACTTAATGACAGTTGCAACT
dd_Smed_v4_26780_0_1	TCGAGTTTTCCCATGTTGTGAC	TGTGTCGTGGTTCGTTCCC	CAAACGGTAAATTTGCCAAGAGA
dd_Smed_v4_2679_0_1	ATATCGGTCAGGCTGGCG	TGCTGGGAGTTGTACTGTCT	CCTCATCTTCGTTATCGTCTTGA
dd_Smed_v4_2789_0_1	ACTCGAAGCGGAAGAAGTGG	CCAATCATAACTGCGTCATCACA	TCTCTGTACACACGCCGC
dd_Smed_v4_28398_0_1	CGTCAATCATCTCAGAAACACCA	CAATATGCTTTCACCAGACACCA	TGACATTCAACTTTGCAACACCA
dd_Smed_v4_2844_0_1	TCAGCAGCAGCAGCATGT	CCGCTGCTGATGCCACTA	AGCAAACGGCCGATGTAGA
dd_Smed_v4_28487_0_1	TGTTGGTGGTCTGTTTTGGTC	TGCCCATTTTGTGTTGCCT	ACAGTAATCGATTTGGAGTTTGG
dd_Smed_v4_30088_0_1	TCTGTCACGGTGATGTTTGT	ACAGCAGTTGATTATCAAGGCG	TCAGCCAATGGAAAATCAGTTGA
dd_Smed_v4_3012_0_1	CTCGTCTCGCAAGCGTCA	ACAAGCTCCATATGGAAGAGGC	CGGCGATGTCTGCTGTGA
dd_Smed_v4_3040_0_1	TGAAGGACGAATGTGACGGT	TTCTCGGTTTATTGTTGGACAAC	AGCTTGTATGGCGCTACACA
dd_Smed_v4_30891_0_1	TCGGGCAGCTTCCTTGTC	AGGTCCATGTGCAATGTGGT	TGTAAGCACATTGAGTTACAGGA
dd_Smed_v4_31236_0_1	TGCTTGGCCTTGTCGGTC	TGGTTAAGCATTTCTGTGGGTC	GGCGACATGACATCGTCCT
dd_Smed_v4_320_0_1	TGAACCAACAGCTGCTGC	CGCTGAACGCAATGTGTT	CAGCCCCTCCTGGTCCT
dd_Smed_v4_3257_0_1	GCGACGTCATTAAGAAGCTAGTG	GATGCAGTGTAGTGAAAATGTCA	ACTGTTTGCCACGCAGGA
dd_Smed_v4_3259_0_1	ACAACGCTTCCATCAACAACA	TCCTCACCTTCATCATCTTCGA	TGGCAGTGCTCAAAAGTACAC
dd_Smed_v4_3269_0_1	TGCAGTTTCTCAATGTCATGACT	AGAATCGCAAGGAGTTGGTGT	CCCATCAGTCCTAGATCGGC
dd_Smed_v4_32934_0_1	AAAGACGACGAAGGGCGC	TCCATCATGCAGAAAGTCGGT	AGGCTTCCAAATCTTTTTCTGTG
dd_Smed_v4_33456_0_1	TCCGACTCAGTTCATGACCA	CTCTTGAAACATCTTTGCCAGGA	GCTCGGAGCGAATGGAAA

dd_Smed_v4_3362_0_1	AATGTGTGGTCATTGGGGATG	TGCAGTTGGGAAAACATGCA	CAAAACTTTGTGCGTTTCCGT
dd_Smed_v4_345_0_1	CGGCGAGTATAAATCGGGGG	ACAAAATGCAATTCAACATGCAA	TCCGTTCTTTTGGATCATGGT
dd_Smed_v4_3541_0_1	AAATGACGGATTTCGCGCC	GCTCAGCTCACATATTGCAGG	ATTCAATGTGGGAAATTTGCACA
dd_Smed_v4_3603_0_1	GCCGCACTAGAGTTGGCA	AGCGAGCGATGTTTATGAAAAGG	AGTGCCACTTCGTGAGCC
dd_Smed_v4_3606_0_1	ACTCTTAATTGTCGCGTTTGTT	AGAATTGACTGAACTCGGAAAGA	AGGTTCATCATAGCATTGGCCA
dd_Smed_v4_3632_0_1	CGTGCTGCGTTTCTTCGG	GCGAAACTTCTGGTGATTGCA	GCACATTTTTGTTGCACAGCA
dd_Smed_v4_3638_0_1	TCCCAAACACTTTGCCAACA	ACAGAAGAAACTTTTTCCCTGCA	TCAGGAAACCGAGGATAAAACT
dd_Smed_v4_3667_0_1	CGTTCTCCGAGTGGCTGG	GGCGAGACACTGAGCTCG	GACCACACGTCGGCCTTT
dd_Smed_v4_3674_0_1	TCTCACAGCCCTCTTCGGA	ACTCAATTTCATAAGAAACGCGT	CCCTCTCGCTCCCTCCTT
dd_Smed_v4_36829_0_1	ATCGACGAAAACCAAATGTTGA	ACAGCAGTCAAGAATACGATGC	GACAGAGATAAATCAGACGGAGC
dd_Smed_v4_3703_0_1	CCAGCAGGGTGCCAGAAT	ACCATGTCTGCTATCAGCTCA	ATTTGGAAATGGCTCGAAGTG
dd_Smed_v4_39545_0_1	ACTAATTCATCGCCACCAACAC	TCGATACAATGAAAGACGACTGG	TCGGAAGTACTTGGAAATTCCCA
dd_Smed_v4_4012_0_1	CAGTGTCACGAAGAAATTGGTCA	AGGGGGTTTCGGAACAGT	ATGAGATCTGACGTTGTCTGAA
dd_Smed_v4_4154_0_1	GGGCTGCTCATGACGTGT	CTCAAAAGCTGATGCCATCGA	GATCGGTCCGCGGGAATC
dd_Smed_v4_4279_0_1	TGGGTTAATTTTATGTTGCACGT	CGTTTTCTGCTTTAACGTTTGCT	ATTTGACAGACAACTGAGTCCA
dd_Smed_v4_4299_0_1	GCAAAGGACCCCATGGCA	ACCCCAAAATGAAACAGTTGCC	ACCGACAAGACACAAAGGACA
dd_Smed_v4_4381_0_1	TGCGTCGACAATGAAATGGA	TGAAATTGGAAAACGGCATGAA	CCCGCGTAGAAATCGGCA
dd_Smed_v4_4392_0_1	TCCGAAATTCTCAGAGCAGATCA	TGGAATCGACAATTGTCTCTGA	TGTTTGGACTTTGATTGCGAGT
dd_Smed_v4_4435_0_1	TGCAATTTATGGGAAAATCGGTG	TCTCAAATGGAAAATCTGTCGCT	TCAAATCTCGACATTCTGCTGA
dd_Smed_v4_4486_0_1	GCCGCCCTCCGTTATGTG	CCCCTCCCAAACTGAAATCCC	TGCCAATTCCACTGCGGT
dd_Smed_v4_4619_0_1	AATCATCCACTTCGATGCCAAC	ATCTATTATCCGAAGAGCCGTCC	TGAAGTTCCCGTAAACAATGTCG
dd_Smed_v4_4793_0_1	GTGGGACTCTGTGCTGTTCA	TCAATTCAAAGTTGTGCACGGA	TGGGAGGTCAGTTGCACTC
dd_Smed_v4_4808_0_1	ACCATCGAAACTCGTGTGCA	ATGGCTCCTAAAGGTAAAGTAGC	ACAACCATCATTGTGGTCCT
dd_Smed_v4_4902_0_1	GGGATTATTCTGGTTTCCGGGT	TGGGCGTCGGTGGAGTAT	ACATCGTATCCAAAACCGCA
dd_Smed_v4_4944_0_1	AGCAGACAAGTGTTTCGTCA	CTCTAATGTGAAATACGGTCAGC	TCTGAGAACAAGAAATCATGCGA
dd_Smed_v4_5081_0_1	TTTCTGTTGTCGCCCCCG	ACAAAGGTGAACTAGGAGTCTTT	ACGGTTCGGCGTACACAG
dd_Smed_v4_5102_0_1	GGTCAGCCAAAGTCCCCC	ACAGTATTTCTTAACACGGGTCA	TGAACCATACGGAGCGGT
dd_Smed_v4_511_0_1	GGCGATACTCACTTGGGAGG	AGGAAAGGATATCACCGATGACA	GCTTGATCTGAGAAAGTGAAAGT
dd_Smed_v4_5120_0_1	TTACAGATCGGCAGGAAGC	TGCACATCGAATGAAAACAGATC	TGAAGTTCTAGAAAATCCAGCCA
dd_Smed_v4_526_0_1	GCTGAATGGGGAAGGAAGACA	TGCAGAAAATGAAATGCCTGGT	AGCCGCTCTAAATGAACCACA
dd_Smed_v4_5390_0_1	GTTGTGGACGTTCCTCTCGA	GGTTTGGCATTGGCATTCAGT	TGCAATCTTGTCAACCATTTCGA
dd_Smed_v4_5406_0_1	TTTCGCTATTTAGATGAGCCGA	TGGCCAGAAATTACTCATGTTGG	TGTTGGTAGCTTCAATTGGGA
dd_Smed_v4_5469_0_1	ACTCTAATGGATCCGAAACTGGA	GGAACTGAAGGATCTGAACCT	ACAACAAGAAATCTCGGTCAGT
dd_Smed_v4_5525_0_1	CTGGTGCTCATAATTTGGAAGCA	GATGTGTTTATGAAACGTCCTGA	GAAGGCTGAGAAATTCGATCGG
dd_Smed_v4_5531_0_1	CGAATCACCCCAGTTCCAGA	GTCTGTCATGACAACCAAACTCT	ACTTGGGGAGTTATCAATTCCGA
dd_Smed_v4_5630_0_1	GTCCTACCGGCGGAAGTG	TGCCGCAAAACTTGTGACTG	GGCACTGACCCTTGCAGT
dd_Smed_v4_5638_0_1	CCTTCTGAAAGGCCTCCATT	ACAACTCTAGGTGTTATTGTCGT	CGCCAACAGTAACATTAGCTGT
dd_Smed_v4_5700_0_1	ACTCAAAATGTCTGGTCGAGGA	GGCTTAGGTAAAGGTGGAGCT	ACTACAATCAACATTTGTGGCCC
dd_Smed_v4_574_0_1	TCGCTGCAGTGCTGATTCA	TGTGCCAGTGTCAAGGCC	AGTGGATCTAAAAAGGCTGTCCA
dd_Smed_v4_5749_0_1	TGCTGTCTTCTAAGCATAAACCA	AGACTACAATAACACCACAGGTC	TCGAGTCTGCTTCATGAATGACA
dd_Smed_v4_5781_0_1	CAGTTGACGCGATCGGGA	ACGCAATTTGACCAGATTCAACA	GAGTGCTGTCCGCTCCAG
dd_Smed_v4_5786_0_1	TGCAAATTTCAGCCGAAAATTTG	TCGATGTTGCAAGGGACAA	TGCTGAGCAATTAAACTTCATCA
dd_Smed_v4_5818_0_1	AAACCATTTTCCCTTGCCAAA	GAGCACCGACACTAGTGGT	AAACCAACTAAAACCGATTTCGA
dd_Smed_v4_5853_0_1	ACTGTTTCAAACATTTCTCCGCA	CAGTCTTCGAATGCAATTAACGA	AGCGAGGTAGTGAATCCTACA

dd_Smed_v4_5862_0_1	ACGATATTTATGCCGCCTATCA	TCCTCAACAATTCGGTACTGAA	GGCTATTTGAATGGATTCTCGCC
dd_Smed_v4_5924_0_1	GGTTCCGGTGCACAAGGA	TGTAGCTGCACTTGATTCGGT	ACCGCTATGTCAAAATCAACCA
dd_Smed_v4_5999_0_1	TTTTTCTGCTCACGGGAAATCC	TCAAATCTCAGTAGGCTAAGGGA	ATGAAAGAATTGATTGCCAACGG
dd_Smed_v4_6047_0_1	GCCCCGAAAACAACAACACA	ACCTGCAAGATCCTCGAGA	TGCCGGATGTTGGTCCAG
dd_Smed_v4_6053_0_1	TGGTGAGGAAATTATGCCTACTG	CCGATCGAATAAGATTTCCAAGC	TGGGCGACGTAGATGTCT
dd_Smed_v4_6075_0_1	TTTTTCAATCTTTCAGCTCTGGC	CTAGAGCGTGTTTTTCTTTACCG	TCAATGATATTGATGATGCAGCC
dd_Smed_v4_6278_0_1	ACATGCCACCGAAGAAACT	TGACTGCATTGAAAAAGGAATCA	TCCTCCTCCTCGTCGAGA
dd_Smed_v4_6349_0_1	AAACCAGTAATTAAGCGACCCT	TGGCTTTTCTTTTATCAGCTGCC	ACCATTGATAAACGTGATGAACC
dd_Smed_v4_6420_0_1	TTTGGAAATTATTGGCGAAGGAG	AGAAAAGCTATTCGTCGATCCGA	TGTTTGTCTTTTTGGAAGAGGTG
dd_Smed_v4_6444_0_1	AAATCCACAAAGACAACAACAGC	GGGTGACCGCTCCTGTG	AAGGTTCTAACTAGCAAATGGGC
dd_Smed_v4_6463_0_1	GACGTTTAACAATCGGCGCT	TGAGTTTTTGTGGGTTCCTGA	ATAGAGAAAGGCCGCAGC
dd_Smed_v4_6562_0_1	ACCGATGCTTGGGGAATGA	GTCGAAGTGAAAGATGTTCGGT	ACAGTGCAGTCGGAACTTGT
dd_Smed_v4_668_0_1	TGATCTTTGCCAAATCAAGCCA	GCTACTTAGCATGGGAGCTACT	TCCAAGTCTAGTCCAAATCGTCA
dd_Smed_v4_678_0_1	CTTTTCAAGCTGAAATCGCACA	GCGATGCACTAGACAAAATTCGA	TAGCGCAGGAAGTCAGCC
dd_Smed_v4_6794_0_1	CACCTTATTTACCCGGGGCA	TGATGCTTACTTTACGAGATGGT	TGGCCAAATTGTAAAACGAGACT
dd_Smed_v4_681_0_1	ACAATTTGCCACTGTGACGTG	CCCCCGATCAGAAAAAGGCT	ACTGCATAGTCCATCATTGCA
dd_Smed_v4_6813_0_1	TGCCTATTTATCCCTTGTCTTGG	CTTCCACAAAATCTCCAATCTGG	ATCATCTGCTGTTGTGGTTTTGC
dd_Smed_v4_6882_0_1	CCTGTTGAAGGGGTCGATT	TGCGGAGAATGTGAATTACCT	GCAATTAACGCTTTGCATCTCC
dd_Smed_v4_6884_0_1	CCGGAGGTTCTTGGCACAA	TGTTGGATATTTGTCGGTGGACT	CGTTTGTGAGTACTTCTTGATCG
dd_Smed_v4_6929_0_1	CCTTGTCACGGTAGCGCA	TCCGTTGTCAATTGTATCTGTCC	CACAGATCCAGCACTCGGT
dd_Smed_v4_6948_0_1	AGCCGGTGTCATTCCTCA	ATTATCTCTGCGAGAACTGGATC	GTGACCGTTTGCGTTTGCT
dd_Smed_v4_7038_0_1	TTCAGCGTGGTCGG	ACAATGCGACAAATGTGCCA	GTTTCTCACCGCTGTGGA
dd_Smed_v4_7063_0_1	TCCTTGCTCATTGCTGCCA	GCTCGGATTAATGGCAGCG	ACGGACGGCTCTTTTCGA
dd_Smed_v4_7128_0_1	TGTGTTCACGAGTTTTTGATTCA	TGATTTAGCTACATCCGAGGAAA	AACGGTGACCAGGCATCG
dd_Smed_v4_7166_0_1	GAAAGTAACCTTTGCCGACGA	ACCACTTGCATTTCAAAAATGGA	TGCCAATTGTGTCATAAACCACT
dd_Smed_v4_7168_1_1	ACAGATGCATGAGTTTGTGAAAT	ACACATCAACAATAGCTCTGACG	AGTTGCAAGGTCAGCGTGA
dd_Smed_v4_7262_0_1	CAACACGCGCAGACACAC	TCCGTTTCTATTTGATCGCCA	TCAGCAATCTGACGAACCTGA
dd_Smed_v4_7295_0_1	GGACTTCGATAAAACACTTGTCA	CACAATTGACATTGGTGTTTCGA	TGTCAACCAGCAAACCGA
dd_Smed_v4_7326_0_1	TGCATATCTGGACGTGGATTAGT	TTCACAAAATTGGAACACGTCA	ACTTTTCTCTTGCAAGTTTCACA
dd_Smed_v4_7413_0_1	TCCATTGAACCAGAAATTCGGC	AGTCGGATGGCAAATGCTGA	AAATTGGGCGCTGAAGCAC
dd_Smed_v4_7444_0_1	ACCAAGACGCAGAGTTGATGA	ATTGCTCCATTTTGGTTTCCAA	CAGCAGCATTAGCATCAGCA
dd_Smed_v4_758_1_1	CCCTGACAGACAGCACCG	CATATTGTCGATACAGGTGTGGG	TTCCCGCTGCTCTTTGGC
dd_Smed_v4_7607_0_1	CATCATGAAGCGAAACACAATGT	TCAAATTGAGACAACTCCGAACA	TGTTACAATGTAGCAGTTGCCA
dd_Smed_v4_766_0_1	TCGTGGCAAAAGGCGTCA	ACAGTTAAAGCGGGAGGC	TGCAACACAGCATAGCACT
dd_Smed_v4_7701_0_1	CCGCTCCAGTACGAACGG	TCAGTGCGATCAAAGAAAAGCA	AAAGCCGACGCCATGTGA
dd_Smed_v4_7731_0_1	AACTCTACCAGTGAAAATCGACA	ACTCCGTTGCCAGGAATTCA	TGCCTGAGCCTTTCATCAGA
dd_Smed_v4_7788_0_1	AGGTACAGGGTTTAAAGCAGCA	ACACCAAGGCGCCAAAGT	TGATTGTCGTTTGTAAATGCCTT
dd_Smed_v4_7921_0_1	ATGGTGCCATTGTCCCGG	TGGATGACGGAAATCAAGGTCA	AACCGAGAGTTGCCGGTG
dd_Smed_v4_8252_0_1	TGACAGTGCCAATTTGCTACA	GGATCCGTGATCATTCTTGGC	ATTTGTGAAGGGCCCCCA
dd_Smed_v4_8302_0_1	TGAAGCTGACAACGGGCA	AGCTGTATCGGTTGAGGCAC	GGTTGACGGTTGAGGGGT
dd_Smed_v4_8340_0_1	GATGCAGTCTGACCCGCA	TGCAACAGGAAGGAAAGTTACTG	ACCCTGTTGAACCAAGCCA
dd_Smed_v4_8356_0_1	GGAACCGTCTATGAATGCGC	CGTAAAGGAAGAATGCCCCCT	CCCGAATATCCCGCTGGG
dd_Smed_v4_8439_0_1	TTGTTACAAATGCACGGTAGTTT	TGAGTTTCGGTGCTATACGGG	TGCAAATGTCCAATTGCAAGACT
dd_Smed_v4_8569_0_1	AGGCTTTGAAACCCAACAGGA	TGCATTGAAAGATCTTATTCCCG	GCCAGCGACAACTTTCGG

dd_Smed_v4_8580_0_1	TCTCGTGCAGTAATTTCTACCGA	AGGAGAAAATGGGATTGCGGA	TCCCCAGTTGCAGTTCGAG
dd_Smed_v4_8829_0_1	TGGGGCAGAATCTTGTGCT	TGGCTGGTCAAGGATTTGGG	TATGTTGACGCAGCGGCC
dd_Smed_v4_8835_0_1	TTCGCCAACCTCCAGCAA	AGGGTGAAGAAGTGTCTCAAGA	GCGCTTTCACACACAAGCA
dd_Smed_v4_8839_0_1	GGATGACGGATTCTCTACGGT	TCAAAATCTTCTGCAAACGTTGA	TCGCCGAAAAATATTTTCCAACA
dd_Smed_v4_8858_0_1	GCGGTTCTTGTCCAGTGGA	TGAGTTGGCCGATATTAACAGT	CGTTCTCCGGTGTGGGTT
dd_Smed_v4_8901_0_1	TGAAGGTTACACTCGGGGG	CTTTGACTGTCAAGCTGGGC	ACTTGCAAGGACACAATTCGAG
dd_Smed_v4_8918_0_1	TCACAGCCTGGGAAAACTCC	CGATAGCATGAACATCATCACAA	TGCGACTGGTAAGCCGTT
dd_Smed_v4_8994_0_1	TAAATGTCGCGGGGCAGT	TGCCAGTATTGGGTGCACA	CCGGCTCCAGAACTGCTC
dd_Smed_v4_9050_0_1	TCCAGTCCGTTGGAAAGGA	TGGTTATGAGGAGAAACTTCGT	TGAATTGTCTGACAAGGCAGGA
dd_Smed_v4_9165_0_1	CGACAGCAAACAGGTAGCC	TGGGGTCAGTACAAAGAAAGAAG	ACGCACAAACCAAACTGACA
dd_Smed_v4_9202_0_1	AGAAATATAACACGGTGTTTGCA	TGACTTGTGCGAATTGTTGCA	ACCTTAAGTGGCGGATGTTGT
dd_Smed_v4_9204_0_1	ACAACTCGATCATTCCTTCTCGT	CGTTCTCGTTTCACCGTCA	TCGAACGCATTATGAGCGA
dd_Smed_v4_9273_0_1	AGATGGCAGTGAACTGGACA	TGGATTAACGCCTCCGCA	TGAGAACTGAACTTTTGGTAGCA
dd_Smed_v4_9402_0_1	GGAGGCTGGGGATGGGTA	TGGTGCATGTATTAGCAGATGGT	TCCATCCTGCCAAGGGGG
dd_Smed_v4_9410_0_1	CTCCTACTGGGAAATTTGGTACA	GACACACCACAACCTTTAGAAGA	TCAAAATTCAGTTTATTGCGGGT
dd_Smed_v4_9416_0_1	TCGAACAACGAGCAACGG	CGTGCCTTCATCATTTTTGGC	AATCGTCCACCCCTCGGA
dd_Smed_v4_9472_0_1	CAATTGTGCGTATTTTGTGGTGT	CGTAATTGGAGCCGGCCA	GATCAAACTAATCGCACCAGCA
dd_Smed_v4_9490_0_1	AGATGACAACCAAAGCCGGA	TGTATCGACAATTTACCGATCGA	AGGGGCCGGTTCAGACT
dd_Smed_v4_9491_0_1	TGAGCCAAAAAGAAGAAAGTGCA	GCATGGAAGATACTCAGGACGT	CGGATCAGATAAGCTCCATTTCG
dd_Smed_v4_9519_0_1	TGCAAAGCTAACGCAGAAGA	CTCTACGGTATTCGACTTTACCA	TCCCATGGAAGCCACGTTC
dd_Smed_v4_953_0_1	AGGACCACCTGGCAGCTA	CCGCAACGGCTGAAACTG	GCTGATCATCCTGCTCACA
dd_Smed_v4_9530_0_1	ACAGCCAGTCTTCGCCAA	TCCCTCGCAGCATTGTGT	ACGCTTCAACCTTTGATCGG
dd_Smed_v4_9546_0_1	CGTTGTTTTCAATGGGTAGCTGT	TTTGGTGAATATTCGCATTCCAT	TCTATCGCGCATGATAGCAA
dd_Smed_v4_961_0_1	GCTTATGCTATGCTCAATGTGGA	TTGGAGACATGGTTCTTAGCCC	AGGCACATCCATAATAGTCTCGT
dd_Smed_v4_9610_0_1	TCAATCTCATTTCTGGACAGTGT	TCCCTCAAATGTCTACGTAGTGG	TAAAATTGCGCTCATTCTGTTGA
dd_Smed_v4_9642_0_1	TGCCACAGACAATCTTGCT	TGGTTCTGCCAATGAGTTAGAT	CTGTCCAACAGCGGCAAC
dd_Smed_v4_9677_0_1	CCGGGGGCCTCAAATTGT	TTCTGCTGACAAAACCTCTCGA	AGCTCATGACGCCCGAAG
dd_Smed_v4_9905_0_1	ACAACAACCGAAAATTGTGCCA	CGTCCTAATTCTCACAATCGCAC	ATTGTCGGTGGGCAGTGG
dd_Smed_v4_996_0_1	ACGGTGTGAATGGATCTTCAGA	AACATGGGAAATGGGTATTGTGA	CCGTTTTGTTCACCGCGG

Additional sequences used for mapping

>rRNA_5s

TCACCCGATCTCGTTCGATCTCGGAAGTTAAGCAGGTTAAGGCCTCGTTAGTACTTGGATGGGTGACCGCCTGGGA
ATACGAGGTGTTGTGGACTTTATACTGTTTGTCCACGACCATACTAATCTGGGTTCACCCGATCTCGTTCGATCTC
GG

>mtRNA_2

TCTAAGCAGTGGTATCAACGCAGAGTACGGGGGGTATTGCACTGTTTAGTTGTGATATTTTCCTTTTTGTATTACG
GTTTGTAGGTATTTTTATGTTTTTATTCTCTCGAATGGAATATGATATATCTTTTGTGTGTTTTGTTTTATTTTCT
TTGGTTAACATTAGAA

>mtRNA_1

AGTTGGTGTTGTTGTTTTGTGCAGGTAAGTTAATTAAAACTAGCAGATTCATGTTCTGTCTATGAGTCCTTTCTCT
GTATATGTGGTTAAGATAGTTTATTCAGAATGTTAATTTGTGGAGTTAATGGTAAAAGACTTGTTTTTCTTAATAT
TTGTTTTAATAGCTTA

>unidentified

CGGCCGGACGTAGCGGTGTGCGTCTGTAATCCAACTACTAGGAGGCCGGGTATATGAATGGTTTGAGATGAGGAGT
TCTGTGAGCATTGCGCCTATGTAGATCGGATGTCCACAGTAAGCTTGGCGTCAACATGGTTATATTGTCGGAGGAT
AGAATACCCAGGTTGT

>SMED_11901_V2

ATGAATGAAATTTTGGAAAAGGATATGAAAGCGATTGAATCCATTAAAGTAAAAGAAAAAAAGGCTGTTGATGGTT
TTATGGGTACCTCATCGTTTCATGGAGTGATTCAAGCATATCATAAACGAAATAAAATTGATAAAGGGAGCTGGTT
CATCAGTTTAGTTATTTGTATGTTTGGCTTAATTGGGCATCTCTACCTAATAATCAGTAGATATATAAGTTTGCCC
ACAACTATTGACATGGTCTCTTCAGTGAATTTTGATCCTTTTCCTGCTGTCGCAATATGTCCGGTTACCTTTATTA
GCAGGGATAAATTCACCAAGTATTACAATACAACTCAAGTTTCCCTTAATAAAAAGCTAGTTGGGGATATTTTCTA
CGTCGATGTAAGTGCCTTGAATTTCTGGAGGTCCCTAAGTAAACAACAAGGCAAAGACATAAACAGTAGTTCAGTT
CTTGGAAAGTATTGGGATGAAGCTGAAACCACTTTCTATAGATTCCAGAAAATGATGAATGTTTCAATAGGTCATC
GAAATTATGAAATGATTTTCTTTTGTGAAATTAACAATAAACCTTGCTCATGGGAACATTTCCTTGAATTCGATCA
TCCGATTTATAAGCGATGTTTTAAATTCTCCTATCCGGTAACTGATGAAGATGAAATTCCAGATAAATTGATATTG
GGGCTTTATGTTGATGATGACTATCAAAGAGACACTGATGATATTAAAACGATAATAACCTCTCATGGAGGAAAGG
TTACTATAAATGAAGCAAGTATTTACCCTGGAACTGAAAGTTCATTTGAACATTTTCCGTCAGGATTCCAAACGAT
GTTTCGATTGAAACAAGAAGGTAGCAGTCAAATCAATAAACCAAGGTCTCCATGCCAAGTTAATACTGATTCAGTG
ATCAACGTTTTCAACGATTATGAATATGATGGCTCAACAAATATCACAATACCATATAAATACAATGTGATACTTT
GCAGACAATACCATCAACAAATAGAATGCGTTAAAAGATGCAAGTGTTTAAATCCGAACATTCCAGTATTTGTTGA
TGCTATTAAGAATTCTGAAAATAAATCATTCTTTTGCGATGAGATTCAGCTTAATTCTTCCTTTTCAAGCATTATT
AATCAGCTTGATTGTCTTTATAATTTAGATTATGATCAGTATTTTAATGAGAATGTTATATCATTATGTTCGGGAT
TGTGTAATCAGGTAGAATATTCAATGTATTCTTATACTATGCCTTGGTTCGGTAAAACAATGATCAAAGAAATGGA
GTTTGTCNAATGAGAAATTCATGGCCCATTACAACAGTCTAATTAATTCCATCCAATTGTGAAGGATTATGGAACC
ATTGAATAGAGCACGCAATTGCGTAATCAAATCCATGAAAGATAATGATCAAGCCAGCTTGTGTTTCGCAATGATT
AATATTCAATTTGAATCTCCCAGAAAAGAAATTATTCGAGAATATGAGGCATATTTATTGGGGAATTTACTCAGTG
ATTTCGGCGGGATTTTAGGACTGTGGATTGGAATGTCTCTGATAACAATTATTGAAATCATATACTTAGCATGCTC
GTTGAGTAAACACAAAACTGAACGCGCTGCTTCAGTTTTCAAAAAGTCAATCCACAAGAGAAGTCTGAAAAGGAAT
TCCGATAAAAACAAAATTATCAGAATCGGAATAGAAAATGAGGCGTATGAAAATTAG

Contig Ids corresponding to genes shown in figures

Fig.	Panel	Gene annotation in figure	Contig	Best-blast hit description	ID	E-value	Organism
1	C	*smedwi-1*	dd_Smed_v4_659_0_1	smedwi-1	DQ186985.1	0	Smed
1	C	*tropomyosin*	dd_Smed_v4_436_0_1	tropomyosin 1 (alpha) (TPM1)	uc002alp.3	1.00E-54	Human
1	C	*vim-1*	dd_Smed_v4_364_0_1	vimentin (VIM)	uc001iou.2	3.00E-29	Human
1	C	*synapsin*	dd_Smed_v4_3135_0_1	synapsin II (SYN2)IIb	uc003bwl.1	1.00E-121	Human
1	E	*znf91*	dd_Smed_v4_7664_0_1	zinc finger protein 91 (ZNF91)	uc002nre.3	3.00E-09	Human
1	E	*zfp-1*	dd_Smed_v4_8720_0_1	ZFP-1	JQ425154.1	0	Smed
1	E	*hnf4*	dd_Smed_v4_1694_0_1	HNF4 (hnf4)	JF802199.1	0	Smed
2	C	*egr-2*	dd_Smed_v4_9273_0_1	clone SMED_20251_V2 early growth response-2	JX010482.1	0	Smed
2	C	*RPSAP58*	dd_Smed_v4_8634_0_1	ribosomal protein SA pseudogene 58 (RPSAP58)	uc002nrn.3	3.00E-22	Human
2	C	*wntless*	dd_Smed_v4_9546_0_1	Evi/Wls	FJ463748.1	0	Smed
2	C	*svopl*	dd_Smed_v4_12695_0_1	SVOP-like (SVOPL)	uc011kqh.2	1.00E-42	Human
2	C	*dd_9490*	dd_Smed_v4_9490_0_1	Smed06730_V2 hypothetical protein	JX010552.1	0	Smed
3	A	*Tob2*	dd_Smed_v4_7444_0_1	transducer of ERBB2, 2 (TOB2)	uc021wqf.1	8.00E-28	Human
3	A	*dd_9519*	dd_Smed_v4_9519_0_1	NA	NA	NA	NA
3	A	*svopl*	dd_Smed_v4_12695_0_1	SVOP-like (SVOPL)	uc011kqh.2	1.00E-42	Human
3	B	*egr-2*	dd_Smed_v4_9273_0_1	clone SMED_20251_V2 early growth response-2	JX010482.1	0	Smed
4	B	*notum*	dd_Smed_v4_24180_0_1	notum	JF725701.1	0	Smed
4	B	*sulfotransferase*	dd_Smed_v4_15647_0_1	sulfotransferase family, cytosolic, 1C, member 3 (SULT1C3)	uc010ywo.2	6.00E-48	Human
4	B	*klf*	dd_Smed_v4_3638_0_1	Kruppel-like factor 13 (KLF13)	uc001zfo.3	9.00E-34	Human
4	B	*TRAF-1*	dd_Smed_v4_4392_0_1	Smed19658_V2 TNF receptor associated factor-1	JX010627.1	8.00E-136	Smed
4	B	*H2B*	dd_Smed_v4_4808_0_1	Smed15708_V2 histone h2b-2	JX010617.1	1.00E-104	Smed
4	B	*dd_6806*	dd_Smed_v4_6808_0_1	NA	NA	NA	NA
4	B	*rhomboid*	dd_Smed_v4_13835_0_1	5B07 rhomboid-like protein	KJ573355.1	0	Smed
5	B	*mex-3*	dd_Smed_v4_6053_0_1	mex-3 homolog A (C. elegans) (MEX3A)	uc001fnd.4	2.00E-08	Human
5	B	*hsp70*	dd_Smed_v4_320_0_1	heat shock 70kDa protein 8 (HSPA8)	uc001pyo.3	0	Human
5	B	*traf2*	dd_Smed_v4_10569_0_1	Smed07121_V2 TNF receptor associated factor-2	JX010549.1	0	Smed
5	B	*sytl2*	dd_Smed_v4_21069_0_1	synaptotagmin-like 2 (SYTL2)h	uc001paw.3	1.00E-35	Human
5	B	*dd_14011*	dd_Smed_v4_14011_0_1	NA	NA	NA	NA
5	B	*CALCR*	dd_Smed_v4_15499_0_1	calcitonin receptor (CALCR)	uc003umw.2	8.00E-22	Human
5	B	*wntless*	dd_Smed_v4_11629_0_1	Evi/Wls	FJ463748.1	0	Smed

5	B	*dd_8302*	dd_Smed_v4_8302_0_1	NA	NA	NA	NA
5	B	*slc16a14*	dd_Smed_v4_9402_0_1	solute carrier family 16, member 14 (monocarboxylic acid transporter 14) (SLC16A14)	uc002vqf.3	8.00E-39	Human
5	B	*mpped1*	dd_Smed_v4_9610_0_1	metallophosphoesterase domain containing 1 (MPPED1)	uc011apy.2	2.00E-16	Human
5	B	*rrm2b*	dd_Smed_v4_5862_0_1	Smed05893_V2 ribonucleoside diphosphate reductase subunit M2	JX010583.1	0	Smed
5	B	*dd_8901*	dd_Smed_v4_8901_0_1	NA	NA	NA	NA
5	B	*dd_9519*	dd_Smed_v4_9519_0_1	NA	NA	NA	NA
5	B	*notch*	dd_Smed_v4_10716_0_1	notch 1 (NOTCH1)	uc004chz.3	8.00E-57	Human
5	B	*jun-1**	dd_Smed_v4_5749_0_1	Smed03061_V2 1-Jun	JX010576.1	0	Smed
5	B	*nlg-1**	dd_Smed_v4_14068_0_1	noggin-like protein 1	EF633691.1	0	Smed
5	B	*inhibin-1**	dd_Smed_v4_7607_0_1	clone SMED_01282_V2 inhibin-1	JX010479.1	0	Smed
5	B	*glypican-1**	dd_Smed_v4_4154_0_1	clone SMED_05117_V2 glypican-1	JX010469.1	0	Smed
5	B	*dd_20048*	dd_Smed_v4_20048_0_1	NA	NA	NA	NA
5	B	*inx-13*	dd_Smed_v4_11501_0_1	INX-13	JQ425145.1	0	Smed
5	B	*cyp2j2*	dd_Smed_v4_2394_0_1	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)	uc001czq.3	3.00E-36	Human
5	B	*sbspon*	dd_Smed_v4_5786_0_1	somatomedin B and thrombospondin, type 1 domain containing (SBSPON)	uc003xzf.3	9.00E-15	Human
5	B	*pif1*	dd_Smed_v4_16842_0_1	PIF1 5'-to-3' DNA helicase (PIF1)	uc010uiq.1	1.00E-119	Human
5	B	*dd_13860*	dd_Smed_v4_13860_0_1	NA	NA	NA	NA
5	B	*pxdn*	dd_Smed_v4_3603_0_1	peroxidasin homolog (Drosophila) (PXDN)	uc002qxa.3	0	Human
5	B	*sfrp1*	dd_Smed_v4_13985_0_1	secreted frizzled-related protein 1 (sfrp1)	EU296635.1	0	Smed
5	B	*med12l*	dd_Smed_v4_11943_0_1	mediator complex subunit 12-like (MED12L)	uc003eyp.3	8.00E-64	Human
5	B	*plasminogen-1*	dd_Smed_v4_23420_0_1	Smed27240_V2 plasminogen-1	JX010625.1	0	Smed
5	D-E	*egr-l1*	dd_Smed_v4_7731_0_1	EGR-like protein 1	JF914965.1	0	Smed
5	D-E	*runt-1*	dd_Smed_v4_16222_0_1	runt-like 1 protein	JF720854.1	0	Smed
5	D-E	*Inhibin-1*	dd_Smed_v4_7607_0_1	clone SMED_01282_V2 inhibin-1	JX010479.1	0	Smed
S2	A	*Rab-11B*	dd_Smed_v4_7604_0_1	EF-hand calcium binding domain 4B (EFCAB4B)	uc010sen.1	3.00E-45	Human
S2	A	*anoctamin 7*	dd_Smed_v4_4761_0_1	anoctamin 7 (ANO7)NGEP-L	uc002wax.2	0	Human
S2	A	*ESRP-1*	dd_Smed_v4_5053_0_1	epithelial splicing regulatory protein 1 (ESRP1)	uc003ygt.4	4.00E-117	Human
S2	A	*myoferlin*	dd_Smed_v4_6816_0_1	myoferlin (MYOF)	uc001kio.3	0	Human
S3	C	*plasminogen-1*	dd_Smed_v4_23420_0_1	Smed27240_V2 plasminogen-1	JX010625.1	0	Smed
S3	C	*fos-1*	dd_Smed_v4_2789_0_1	clone SMED_00055_V2 fos-1	JX010471.1	0	Smed
S3	C	*hadrian*	dd_Smed_v4_3606_0_1	clone SMED_02793_V2 hadrian	JX010472.1	0	Smed
S3	C	*glypican-1*	dd_Smed_v4_4154_0_1	clone SMED_05117_V2 glypican-1	JX010469.1	0	Smed
S3	C	*dusp10*	dd_Smed_v4_4619_0_1	dual specificity phosphatase 10 (DUSP10)	uc001hmy.2	7.00E-39	Human

S3	C	*wntless*	dd_Smed_v4_11629_0_1	Evi/Wls	FJ463748.1	0	Smed
S3	C	*egr-l 1*	dd_Smed_v4_7731_0_1	EGR-like protein 1	JF914965.1	0	Smed
S3	C	*egr-4*	dd_Smed_v4_9410_0_1	clone SMED_09938_V2 early growth response-4	JX010483.1	0	Smed
S3	C	*HSP20**	dd_Smed_v4_5406_0_1	NA	NA	NA	NA
S3	C	*innexin-1*	dd_Smed_v4_11254_0_1	Smed09630_V2 innexin-1	JX010623.1	0	Smed
S3	C	*dd_4944*	dd_Smed_v4_4944_0_1	NA	NA	NA	NA
S3	C	*traf2*	dd_Smed_v4_10569_0_1	Smed07121_V2 TNF receptor associated factor-2	JX010549.1	0	Smed
S3	C	*mex-3*	dd_Smed_v4_6053_0_1	mex-3 homolog A (C. elegans) (MEX3A)	uc001fnd.4	2.00E-08	Human
S3	C	*Jun-1*	dd_Smed_v4_5749_0_1	Smed03061_V2 1-Jun	JX010576.1	0	Smed
S3	C	*sfrp1*	dd_Smed_v4_13985_0_1	sFRP1	EU296635.1	0	Smed
S3	C	*MPPED1*	dd_Smed_v4_9610_0_1	metallophosphoesterase domain containing 1 (MPPED1)	uc011apy.2	2.00E-16	Human
S3	C	*7tm**	dd_Smed_v4_20048_0_1	NA	NA	NA	NA
S3	C	*inx-13*	dd_Smed_v4_11501_0_1	INX-13	JQ425145.1	0	Smed
S3	C	*CPO*	dd_Smed_v4_5999_0_1	carboxypeptidase O (CPO)	uc002vby.2	8.00E-63	Human
S3	C	*ldlr-1*	dd_Smed_v4_1581_0_1	Smed05022_V2 low density lipoprotein receptor-1	JX010530.1	0	Smed
S3	C	*SLC16A14*	dd_Smed_v4_9402_0_1	solute carrier family 16, member 14 (monocarboxylic acid transporter 14) (SLC16A14)	uc002vqf.3	8.00E-39	Human
S3	C	*CYP2J2*	dd_Smed_v4_2394_0_1	cytochrome P450, family 2, subfamily J, polypeptide 2 (CYP2J2)	uc001czq.3	3.00E-36	Human
S3	C	*dd_5390*	dd_Smed_v4_5390_0_1	NA	NA	NA	NA
S3	C	*rrm2b*	dd_Smed_v4_5862_0_1	Smed05893_V2 ribonucleoside diphosphate reductase subunit M2	JX010583.1	0	Smed
S3	C	*dd_9642*	dd_Smed_v4_9642_0_1	NA	NA	NA	NA
S3	C	*sbspon*	dd_Smed_v4_5786_0_1	somatomedin B and thrombospondin, type 1 domain containing (SBSPON)	uc003xzf.3	9.00E-15	Human
S3	C	*HYOU1*	dd_Smed_v4_2324_0_1	hypoxia up-regulated 1 (HYOU1)	uc010ryu.1	0	Human
S3	C	*HSP90**	dd_Smed_v4_758_1_1	NA	uc001tkb.1	0	Human
S3	C	*dd_9519*	dd_Smed_v4_9519_0_1	NA	NA	NA	NA
S3	C	*HSP70**	dd_Smed_v4_1087_0_1	heat shock 70kDa protein 4-like (HSPA4L)	uc003ifm.3	0	Human

Supplementary References

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Chechik, G., and Koller, D. (2009). Timing of gene expression responses to environmental changes. Journal of computational biology : a journal of computational molecular cell biology *16*, 279-290.

Chechik, G., Oh, E., Rando, O., Weissman, J., Regev, A., and Koller, D. (2008). Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. Nature biotechnology *26*, 1251-1259.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics *31*, 545-554.

Eisenhoffer, G.T., Kang, H., and Sánchez Alvarado, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian Schmidtea mediterranea. Cell stem cell *3*, 327-339.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell *159*, 188-199.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150-3152.

Gurley, K.A., Elliott, S.A., Simakov, O., Schmidt, H.A., Holstein, T.W., and Sánchez Alvarado, A. (2010). Expression of secreted *Wnt* pathway components reveals unexpected complexity of the planarian amputation response. Developmental biology *347*, 24-39.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M*., et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols *8*, 1494-1512.

Hayashi, T., Asami, M., Higuchi, S., Shibata, N., and Agata, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. Development, growth & differentiation *48*, 371-380.

Kao, D., Felix, D., and Aboobaker, A. (2013). The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. BMC genomics *14*, 797.

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics *12*, 124.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, S.Y., Selck, C., Friedrich, B., Lutz, R., Vila-Farre, M., Dahl, A., Brandl, H., Lakshmanaperumal, N., Henry, I., and Rink, J.C. (2013). Reactivating head regrowth in a regeneration-deficient planarian species. Nature *500*, 81-84.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461-467.

Petersen, C.P., and Reddien, P.W. (2008). *Smed-betacatenin-1* is required for anteroposterior blastema polarity in planarian regeneration. Science *319*, 327-330.

Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods *10*, 1096-1098.

Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature protocols *9*, 171-181.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
Reddien, P.W. (2011). Constitutive gene expression and the specification of tissue identity in adult planarian biology. Trends in genetics : TIG *27*, 277-285.
Reddien, P.W., Oviedo, N.J., Jennings, J.R., Jenkin, J.C., and Sánchez Alvarado, A. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science *310*, 1327-1330.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
Sandmann, T., Vogg, M.C., Owlarn, S., Boutros, M., and Bartscherer, K. (2011). The head-regeneration transcriptome of the planarian Schmidtea mediterranea. Genome biology *12*, R76.
Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature biotechnology.
Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S.*, et al.* (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148-162.
Scimone, M.L., Kravarik, K.M., Lapan, S.W., and Reddien, P.W. (2014). Neoblast specialization in regeneration of the planarian Schmidtea mediterranea. Stem cell reports *3*, 339-352.
Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N.*, et al.* (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363-369.
Sivriver, J., Habib, N., and Friedman, N. (2011). An integrative clustering and modeling algorithm for dynamical gene expression data. Bioinformatics *27*, i392-400.
van Wolfswinkel, J.C., Wagner, D.E., and Reddien, P.W. (2014). Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. Cell stem cell *15*, 326-339.
Vogg, M.C., Owlarn, S., Perez Rico, Y.A., Xie, J., Suzuki, Y., Gentile, L., Wu, W., and Bartscherer, K. (2014). Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. Developmental biology *390*, 136-148.
Wagner, D.E., Ho, J.J., and Reddien, P.W. (2012). Genetic regulators of a pluripotent adult stem cell system in planarians identified by RNAi and clonal analysis. Cell stem cell *10*, 299-311.
Wenemoser, D., Lapan, S.W., Wilkinson, A.W., Bell, G.W., and Reddien, P.W. (2012). A molecular wound response program associated with regeneration initiation in planarians. Genes & development *26*, 988-1002.