

RESEARCH ARTICLE SUMMARY

SINGLE-CELL ANALYSIS

Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics

Mireya Plass,* Jordi Solana,* F. Alexander Wolf, Salah Ayoub, Aristotelis Misios, Petar Glazar, Benedikt Obermayer, Fabian J. Theis, Christine Kocks, Nikolaus Rajewsky†

INTRODUCTION: Understanding the differentiation of stem cells into the vast amount of cell types that form the human body is a central problem of basic and medical science. The recent advances in single-cell sequencing techniques now make it possible to capture the transcriptomes of thousands of cells in a fast and cost-effective manner, opening a new way to study the cell composition of organs, tissues, and developmental stages. Yet, single-cell transcriptomics per se just provides a snapshot of cellular dynamics and transient cell populations. Computational algorithms have emerged that infer a pseudotemporal ordering of cells based on comparison of their transcriptomic profiles, allowing new insights into stem cell biology and tissue differentiation. However, these algorithms were designed for relatively simple scenarios, such as the differentiation of cells belonging to a specific lin-

eage or the lineage relationships among cells from a particular tissue, and cannot evaluate all possible cellular differentiation trajectories in complex animals. To this end, we use single-cell transcriptomic approaches to improve our molecular understanding of how stem cells differentiate into the set of cell types that make an entire complex adult animal.

RATIONALE: Freshwater planarians such as *Schmidtea mediterranea* offer a unique opportunity to approach this question. These animals are immortal and constantly renew and regenerate all tissues owing to the presence of a large pool of pluripotent stem cells that continuously differentiate into all mature cell types. Therefore, we reasoned that an unbiased single-cell transcriptomic approach should allow us to capture not only terminally differentiated cell types but also intermediate cellular states, pos-

sibly enabling cell lineage reconstruction of the whole animal from transcriptomic data.

RESULTS: We performed massively parallel single-cell transcriptomics profiling of thousands of cells from adult planarians. At the molecular level, we identified and characterized dozens of cell types, including stem cells, progenitors, and terminally differentiated cells. We then applied a new computational algorithm, partition-based graph abstraction (PAGA), which can predict a lineage tree for the whole animal in an unbiased way. By combining the predictions from PAGA with several independent

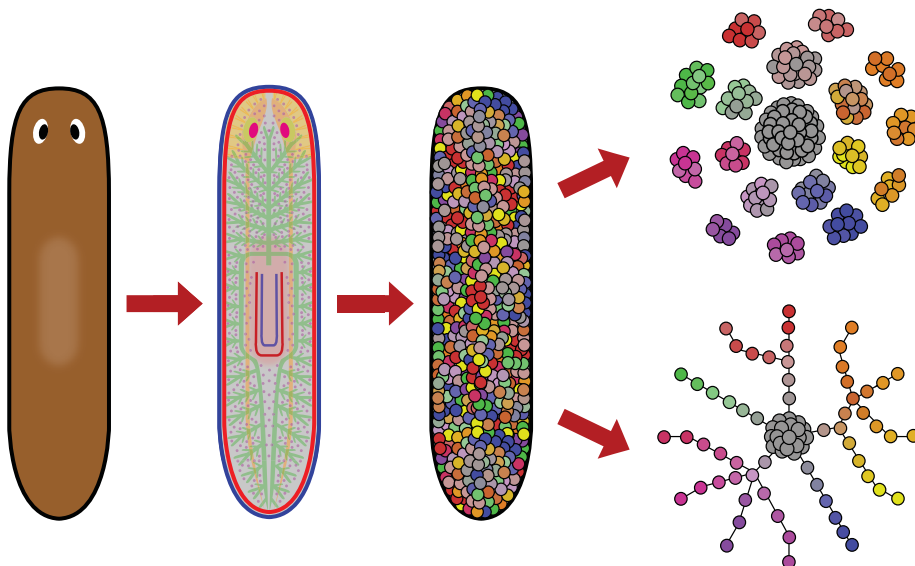
ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aag1723>

lines of evidence, including single-cell transcriptome data from purified stem cells and stem cell-depleted animals, analysis of gene expression dynam-

ics, and a method called velocity to that predicts future gene expression from mRNA metabolism, we produced a consolidated lineage tree that included all identified cell types rooted to a single stem cell group. We used this information to identify gene sets co-regulated during the differentiation of many specific cell types. To show the power of our approach, we applied single-cell transcriptomics to regenerating planarians and characterized how each cell type in the adult planarian body dynamically responds to regenerative body remodeling at the transcriptomic and cellular levels. Our results highlight that some cell types that had been previously overlooked in molecular studies quickly decrease their abundance, indicating that they may serve as an energy reservoir that fuels the regeneration process.

CONCLUSION: We have shown that it is possible to use single-cell transcriptomics to (i) build a cell atlas of an adult animal, (ii) reconstruct the lineage relationships of its cells in an unbiased way, and (iii) identify gene sets which likely contain genes that are involved in programming the lineage tree. Moreover, we demonstrated how single-cell transcriptomics can be used to study complex and dynamic cellular processes such as regeneration. Notably, our approach is applicable to other model and non-model organisms, assuming that their differentiation processes are sampled with sufficient time resolution. To foster future studies, we provide a detailed tutorial on the application of our approach, and we make our data available through an interactive web interface. This study opens the door to powerful approaches for understanding molecular mechanisms of development and regeneration in animals. ■



A lineage tree for complex animals from single-cell transcriptomics. Planarians are multicellular organisms. They contain adult pluripotent stem cells that continuously renew all tissues and differentiate into all adult cell types. Using single-cell transcriptomics, we characterized all major mature cell types and many intermediate cellular states. We then derived a lineage tree describing planarian stem cell differentiation into all mature cell types of the animal.

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: rajewsky@mdc-berlin.de
Cite this article as M. Plass et al., *Science* 360, eaag1723 (2018). DOI: 10.1126/science.aag1723

RESEARCH ARTICLE

SINGLE-CELL ANALYSIS

Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics

Mireya Plass,^{1*} Jordi Solana,^{1*†} F. Alexander Wolf,² Salah Ayoub,¹ Aristotelis Misiros,¹ Petar Glazar,¹ Benedikt Obermayer,^{1‡} Fabian J. Theis,^{2,3} Christine Kocks,¹ Nikolaus Rajewsky^{1§}

Flatworms of the species *Schmidtea mediterranea* are immortal—adult animals contain a large pool of pluripotent stem cells that continuously differentiate into all adult cell types. Therefore, single-cell transcriptome profiling of adult animals should reveal mature and progenitor cells. By combining perturbation experiments, gene expression analysis, a computational method that predicts future cell states from transcriptional changes, and a lineage reconstruction method, we placed all major cell types onto a single lineage tree that connects all cells to a single stem cell compartment. We characterized gene expression changes during differentiation and discovered cell types important for regeneration. Our results demonstrate the importance of single-cell transcriptome analysis for mapping and reconstructing fundamental processes of developmental and regenerative biology at high resolution.

Understanding differentiation from stem cells into the different cell types that make up the human body is a central problem of basic and medical science. Although numerous mechanisms of cellular differentiation have been identified and many cell types have been characterized, it will require a huge coordinated undertaking to systematically map all human cell types and cellular differentiation states (1). Owing to the advances in single-cell transcriptomics, it has already been possible to study the cell type composition of mammalian organs and tissues (2–6), as well as development stages (7, 8). However, single-cell transcriptomics provide just a snapshot of the dynamics of the cell populations unless cells can be traced or tagged experimentally (9–12). Thus, reconstructing cell lineages from stem cells to differentiated cells remains a challenge. Recently, algorithms to order developmental states and compute lineage trees based on comparing single-cell transcriptomes have made considerable progress (13–15) and have revealed insights into stem cell biology (16) and tissue differentiation (17–20). However, these

algorithms have been developed for the study of differentiation in specific cell lineages or tissues and are not suitable to reconstruct all the cell differentiation trajectories present in complex animals.

Given these problems, can single-cell transcriptomic approaches improve our molecular understanding of how stem cells differentiate into all the cell types of an entire complex adult animal? Freshwater planarians such as *Schmidtea mediterranea* offer a unique opportunity to answer this question. Planarians are immortal, as they contain as adults a large pool of pluripotent stem cells (neoblasts) that continuously differentiate to all mature cell types to turn over all tissues (21). Hence, all cell differentiation pathways are constantly active in adult individuals. We therefore reasoned that an unbiased single-cell transcriptomics approach should yield terminally differentiated cell types as well as many intermediate cellular states, making planarians an ideal model system to attempt the lineage reconstruction of a whole animal.

Here, we performed highly parallel droplet-based single-cell transcriptomics, Drop-seq (3), to characterize planarian cell types. We molecularly characterized dozens of cell types and uncovered many new ones. By applying a newly developed algorithm, partition-based graph abstraction (PAGA), which reconciles the principles of clustering and pseudotemporal ordering (22), and combining it with independent computational and experimental approaches, we derive a consolidated lineage tree that includes all identified cell types rooted to a single stem cell cluster. Along this tree, we identify 48 gene sets that are

co-regulated during the differentiation of specific cell types. Finally, we used single-cell transcriptomics to characterize cellular processes that happen during regeneration. Our results reveal a strong depletion of newly characterized cell types, suggesting that these cells are used as an energy source for regeneration.

A high-resolution cell type atlas for planaria

To comprehensively characterize different cell types and progenitor stages present in adult planarians, we performed genome-wide expression profiling in individual cells using nanoliter droplets (Drop-seq) (3) of cells isolated from whole adult animals. These cells were obtained, after dissociation, by fluorescence-activated cell sorting (FACS), which separated intact live cells from dead cells and enucleated cellular debris (Fig. 1A). From 11 independent experiments, we captured a total of 21,612 cells. We detected on average 494 genes and ~970 transcripts [identified by using “unique molecular identifiers” (UMIs)] per cell. The individual data sets correspond to five wild-type samples (10,866 cells), two RNA interference (RNAi) samples (3314 cells), a high-DNA content G₂/M population corresponding to cycling planarian stem cells (typically defined as x-ray-sensitive “X1 cells”; 981 cells) (23, 24), and three wild-type regeneration samples (6451 cells; table S1). Sequencing depth was comparable across samples (fig. S1A). Biological replicates showed highly correlated gene expression profiles (fig. S1B). In addition, all samples showed high correlation with published RNA-sequencing (RNA-seq) data from equivalent bulk cell populations (24–27) (fig. S1C). We pooled and analyzed all single-cell data sets together using Seurat (3). Eight of 11 samples were fixed with methanol (28) or frozen with dimethyl sulfoxide (DMSO) (29) to facilitate sample handling. To assess batch effects, we compared the overall quality across wild-type samples. Cells from each batch were distributed similarly on the t-distributed stochastic neighbor embedding (tSNE) (fig. S1D), which resulted in comparable proportions of cells per cluster (fig. S1E and table S2). Although we observed a mild bias in gene expression due to the preservation procedure of the samples, clustering was not affected (fig. S1F). However, we observed differences in the number of UMIs per cell across clusters (fig. S1G). Together, these analyses confirmed that sample preparation did not compromise data quality or introduce bias. Therefore, we clustered the expression profiles of the individual cells from all samples together using Seurat (3). In total, we identified 51 cell clusters (fig. 1B).

We elucidated the cell type identity of clusters by examining marker genes and comparing them to those in previous studies (fig. S2, A and B) (supplementary note 1). The largest cluster and 14 smaller clusters located in the center of the tSNE plot express combinations of well-known stem cell markers (fig. S3A), such as *Smedwi-1*, *Smedtud-1*, and *bruli* (fig. S3B). The remaining clusters corresponded to the previously described neural, epidermal, secretory, muscle, gut, and

¹Laboratory for Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology, Max-Delbrück Center for Molecular Medicine, Berlin, Germany. ²Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany. ³Department of Mathematics, Technische Universität München, München, Germany.

*These authors contributed equally to this work. †Present address: Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK. ‡Present address: Core Unit Bioinformatics, Berlin Institute of Health, Berlin, Germany.

§Corresponding author. Email: rajewsky@mdc-berlin.de

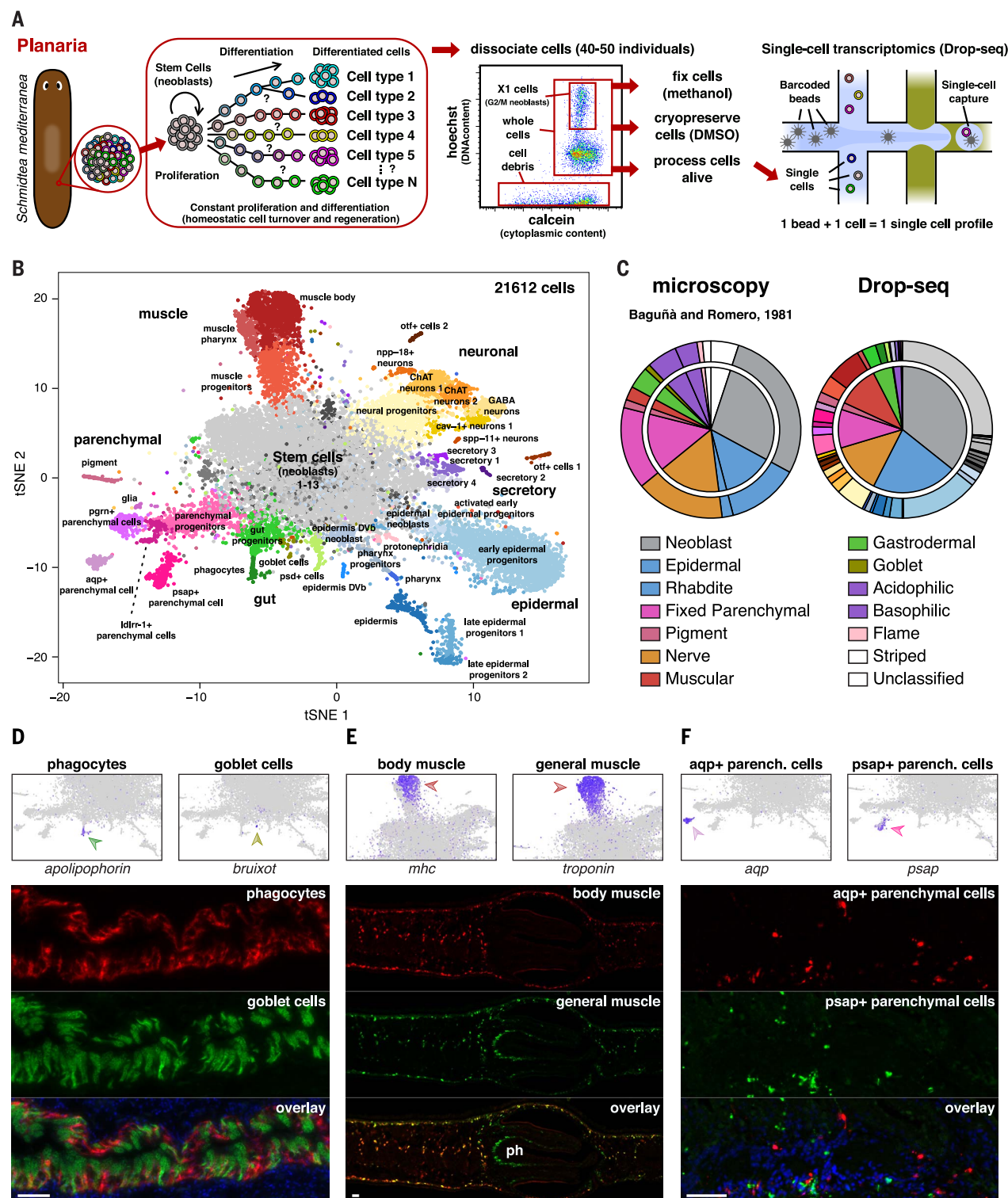


Fig. 1. Cell type atlas by single-cell transcriptomics. (A) Experimental workflow. (B) tSNE representation of the single-cell transcriptomics data with clusters colored according to the expression of previously published marker genes as follows: gray, neoblasts; orange, neuronal lineage; red, muscle; purple, secretory; blue, epidermal lineages; pink, protonephridia; green, gut; magenta, parenchymal lineages. (C) Proportions of cell types identified by Baguña and Romero by microscopy (left) and as identified by tallying up our annotated Drop-seq clusters (right). The outer ring shows the proportion of each individual cluster, which includes neoblasts, epidermal (epidermal and rhab-

dite), parenchymal (fixed parenchymal), pigment, neuronal (nerve), muscular, gut (gastrodermal and goblet), secretory (acidophilic and basophilic), and protonephridia (flame) cells. We did not find “striped” cells in our data set. Overall, we find many subtypes for each of the original cell types. (D to F) tSNE plots (upper panels) showing the expression of marker genes and their expression patterns in adult animals with double in situ hybridizations on tissue sections (lower panels). Nuclei in (D) and (F) were stained with Hoechst and are shown in blue in the overlay. Scale bars: 100 μ m. The color scale for tSNE plots ranges from light gray (no expression) to blue (high expression).

protonephridia cell types (fig. S2, A and B). However, in each of these categories, we found several distinct clusters (Fig. 1B) that express different combinations of marker genes (table S3 and fig. S4). This result suggests that our approach can distinguish more cell types than previous studies.

Single-cell transcriptomics unveils previously uncharacterized cell types

In the 1980s, Bagnà and Romero used microscopy to morphologically characterize and count all major cell types in *S. mediterranea* (30). We used this resource as a reference to validate the cell types identified by our Drop-seq data and cluster annotation. Even though the microscopy data are of a qualitative nature, we observed a strong correlation between it and our molecular, unbiased Drop-seq annotation (Fig. 1C), suggesting that FACS sorting, cryopreservation or fixation, and cell capture in nanodroplets did not influence cell type proportions. We validated the identity of several clusters by designing RNA probes targeting marker genes and performing in situ hybridizations, both whole mount and in histological sections (fig. S5). We could confirm major known cell types such as different types of neurons, muscle, protonephridia, epidermis, and secretory cells. We identified the two main cell types of the planarian gut: phagocytes (Fig. 1D, red) and goblet cells (Fig. 1D, green), and discovered markers of planarian goblet cells, including a gene without apparent homologs in other phyla. We named this gene *bruixot*. We also distinguished body and pharynx muscle (Fig. 1E). General muscle markers colocalized with body

muscle markers throughout the entire body except in the pharynx (Fig. 1E). Pharynx muscle was characterized by the expression of *laminin* (31) (fig. S5). The protonephridia cluster (0.3% of our wild-type cells) contained the two main cell types of these organs, flame and tubular cells (32) (figs. S4 and S5). In some cases, cell clusters contain several similar subtypes that we cannot distinguish at this resolution. For instance, previously described markers of eye pigment cup cells and photoreceptor neurons (33) are expressed in pigment and ChAT neurons 2 clusters, respectively, indicating that the former are subtypes of the latter (fig. S6).

We also validated a recently discovered epidermis cell type, which marks the boundary between the dorsal and ventral parts of planarians (fig. S5) (34). Additionally, we identified an epidermal-related pharynx cell type (fig. S5) and several parenchymal cell types previously undescribed molecularly (Fig. 1B and fig. S5). Among parenchymal clusters, we found a diversity of nonoverlapping cell types, including *aqp+* and the *psa+* parenchymal cells (Fig. 1F), which probably collectively correspond to the previously described fixed parenchymal cells (30, 35), pigment cells (cluster 44) (36, 37), and glial cells (38, 39) (cluster 47) (figs. S4 and S5). Altogether, these results show that we can identify known as well as unknown cell types using single-cell transcriptomics and measure their abundances in a reproducible way.

To investigate the function of newly identified cell types, we used pathway and gene set over-dispersion analysis (PAGODA) (40) to identify variable gene sets with particular Gene Ontology

(GO) terms annotated (fig. S7 and supplementary note 2). The clustering that emerges with these gene sets roughly recapitulates the one obtained with Seurat, showing the robustness of our clustering approach (fig. S7A). This analysis revealed that neoblasts and progenitors are functionally similar, both expressing gene signatures enriched for GO terms related to RNA processing. Additionally, parenchymal clusters showed enrichment for GO terms related to “lysosome,” “extracellular region,” and hydrolytic enzymes and appear to share metabolic functions with gut cells (fig. S7B).

Single-cell transcriptomics of purified stem cells and stem cell-depleted animals reveals stem, progenitor, and differentiated cell populations

The great diversity of cell types identified, which included stem cells, differentiated cells, and presumably many progenitor cells, offered a unique opportunity for exploring stem cell differentiation and lineage relationships between all cell clusters. We focused on the X1 cell sample, which is enriched in G_2/M neoblasts (23, 24), and the *histone 2b* (*h2b*) RNAi-treated whole-planaria sample, in which stem and progenitor cell populations are depleted (41). Cells from these data sets showed a clear distribution pattern: X1 cells were located in the middle of the tSNE plot (Fig. 2A, red dots), whereas *h2b*(RNAi)-resistant cells were clearly enriched in the periphery (Fig. 2A, blue dots). This distribution was specific and not the result of batch effects, as evident from the respective control samples [Fig. 2A, X1 control and *gfp*(RNAi) samples]. Given that each data set

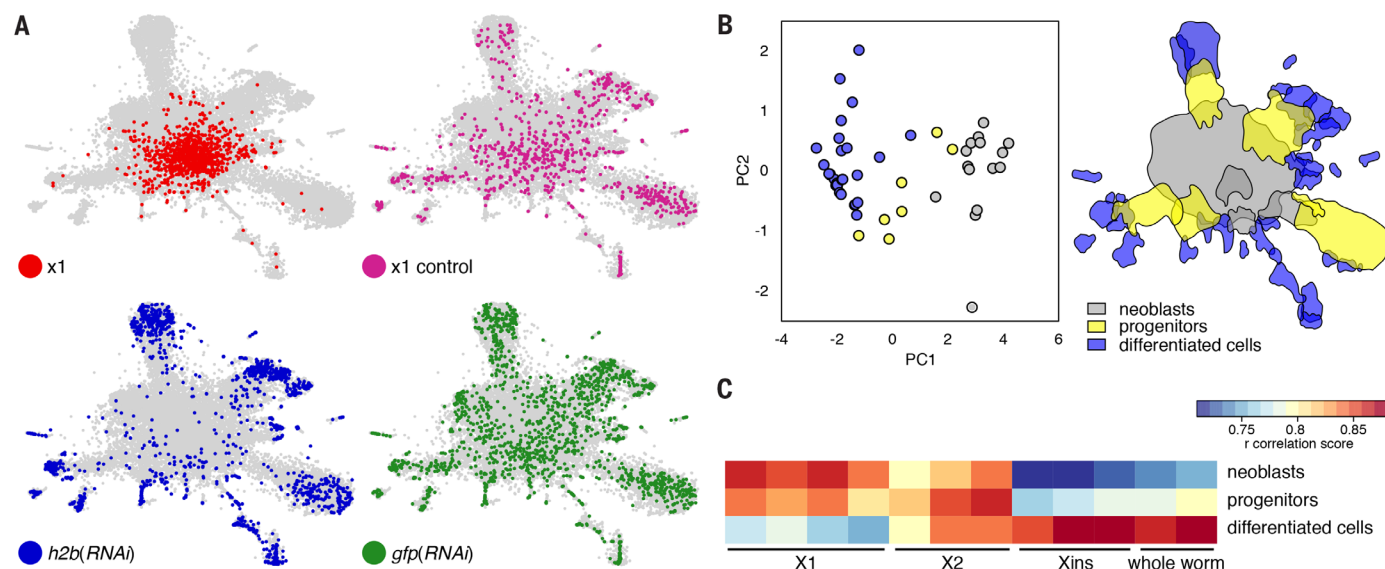


Fig. 2. Neoblast ablation and enrichment experiments show stem and progenitor clusters. (A) tSNE plots showing the distribution of the cells of an X1 FACS-sorted sample (red) and its whole-cell population control (X1 control, magenta), and a *h2b*(RNAi) sample with its negative control [*gfp*(RNAi), green]. X1 cells are enriched in the center of the plot, whereas *h2b*(RNAi) cells are enriched in the periphery. (B) PCA analysis considering the expression level of neoblast marker genes and the log odds ratio of the amount of cells per cluster from *h2b*(RNAi) and X1 experiments

compared to wild-type and control samples separates neoblasts (gray), progenitor clusters (yellow), and differentiated cell clusters (blue). The location of these clusters is shown on the tSNE plot on the right. (C) Gene expression correlation between bulk RNA-seq data from FACS-sorted X1, X2, Xins populations, and whole worms and the pooled clusters as defined in (B). Neoblasts show a stronger correlation with X1, progenitors with X2, and differentiated cells with Xins and whole worms.

is enriched in particular cell populations, we reasoned that they could be used to distinguish cells in varying differentiation states. We quantified the fraction of cells per cluster from the X1 and *h2b(RNAi)* samples and compared them to wild-type and control samples. We performed a principal component analysis (PCA) using these cellular proportions as well as the mean expression of the three top neoblast markers (*Smedwi-1*, *tub-a1*, and *h2b*) in each cluster. The first two principal components resulting from this analysis separated clusters according to their gene expression profiles as neoblasts, progenitors, and differentiated cell clusters (Fig. 2B). Mapping onto the tSNE revealed that progenitor cell clusters were located between differentiated cells and neoblast clusters (Fig. 2B). To corroborate the differentiation state of the cells in the different clusters, we pooled the cells in each group and correlated their gene expression profiles to previously described FACS populations. Neoblast clusters best correlated with X1 populations, corresponding to high-content DNA G₂/M neoblasts; progenitor clusters correlated with X2 populations, a mixture of G₁/S neoblasts and early progenitors; and differentiated cell clusters correlated with Xins samples, a pool of all differentiated cells (Fig. 2C). Altogether, our functional experiments reveal the stem, progenitor, or differentiated status of each cell cluster.

Computational lineage reconstruction predicts a single tree for all major planarian cell differentiation trajectories

Existing methods to investigate cell differentiation using single-cell transcriptomics data were designed to study individual lineages or organs, allow few branching trajectories (13, 15, 18), and often require high sequencing depth and associated costs (16). To overcome these limitations, we developed the general framework of PAGA, which reconciles clustering and pseudotemporal ordering algorithms and allows the inference of complex cell trajectories and differentiation trees (22). Starting from the neighborhood graph of single cells, in which cells are represented as nodes, the algorithm quantifies the connectivity of cell clusters and generates a much simpler abstracted graph in which nodes correspond to the clusters identified by Seurat and edges represent putative transitions between clusters. The differentiation tree is then computed as the treelike subgraph in the abstracted graph that best explains all continuous progressions along the original single-cell graph (supplementary note 3).

When running this algorithm, without any assumptions about the tree structure, we obtained an abstracted graph that shows high confidence of the branching events (Fig. 3A) from which we can derive a single differentiation tree that included all the cell types and linked them to a single root, the neoblast 1 cluster. This tree defines independent differentiation branches for all the major tissues such as neurons, muscle, parenchyma, and gut (Fig. 3A). Additionally, the tree reflects the relation between

different groups of cells. For example, it predicts the existence of independent progenitor cells for the epidermis dorsoventral boundary and the pharynx cell type lineages, although both lineages are related to the epidermal lineage. By contrast, it shows the presence of a shared progenitor for all parenchymal lineages despite containing cell types as different as glia and pigment cells. The connections in the tree are highly consistent with the continuity of gene expression patterns along the various lineages (fig. S8A) except for two cases: The epidermis cluster itself is disconnected from epidermal lineage, and muscle pharynx is connected to muscle body instead of muscle progenitors (fig. S8A). Together, from 51 clusters (with 1275 possible transitions between them), PAGA predicts 53 transitions that are mainly consistent with our marker-based analysis.

Furthermore, PAGA yields a pseudotemporal ordering of individual cells within each cluster consistent with our stem cell ablation and purification experiments and therefore confidently predicts their differentiation status, even for cell types for which separate progenitor clusters could not be identified (fig. S8B). For instance, when we sorted the goblet cells by pseudotime, we observed a higher percentage of X1 cells in early pseudotime and *h2b(RNAi)* cells in the late pseudotime (fig. S8B). To validate this observation, we performed double FISH (fluorescence in situ hybridization) of *bruiixot*, our newly identified goblet cell marker, and *adb* (*aprenent de bruiixot*), a gene expressed earlier in the goblet cluster pseudotime (fig. S8B). Consistently, *adb* was expressed in the gut (fig. S8C) overlapping with *bruiixot*, but staining more cells located in the periphery of the gut that clearly lacked goblet cell morphology (fig. S8D). This indicates that *adb* is a marker of immature goblet cells and that computationally estimated pseudotime correctly orders cells according to their differentiation status.

Although the tree predicts the connectivity of cell clusters, it does not give any information about the direction of the trajectories. Thus, we used the tree topology to estimate the developmental potency of each cluster, i.e., their ability to give rise to other cells. We developed a potency score that is conceptually similar to the stemID score previously proposed to identify stem cells (16) but additionally estimates pluripotency versus multi- or unipotency of cell populations. It is computed as the normalized degree of each cluster in the abstracted graph (supplementary note 4). This analysis showed that neoblast 1, the largest stem cell cluster, had a score of 1 (Fig. 3B), correctly assigning pluripotency to neoblasts as expected from earlier studies (21). We note that the potency score is independent of prior information and therefore can be used to identify stem cells from single-cell transcriptomics data alone, a feature that is particularly useful in less well-studied non-model organisms. Progenitor clusters showed lower potency than neoblasts and higher potency scores than differentiated cells (Fig. 3B), in agreement with a gradual potency loss. To assess the stem cell and progenitor

status of the clusters connected in the center of the PAGA topology, we mapped X1 and *h2b(RNAi)* data onto the tree. Most X1 cells were located in the neoblast 1 cluster (Fig. 3C), whereas *h2b(RNAi)*-resistant cells were more enriched in the leaves of the tree (Fig. 3D). Thus, both PAGA and stem cell ablation and purification independently support the stem and progenitor status of these clusters.

The remaining neoblast clusters had lower potency scores than the neoblast cluster 1 and were connected to it. These clusters share the majority of marker genes with the neoblast 1 cluster (table S3) and do not correspond to previously identified specialized neoblasts of the sigma, gamma, and zeta class (26, 42, 43) (fig. S9). Although some of these neoblast clusters are connected to differentiated cell types (Fig. 3A), most do not give rise to differentiated cell types, raising the possibility that they represent neoblasts in different metabolic, cell cycle, or activation states (supplementary note 5).

We detect expression of specialized neoblast markers among both neoblast and progenitor clusters (figs. S10 and S11). Although present in neoblasts, sigma markers were most highly expressed in neural and muscle progenitors, gamma markers in gut, and parenchymal progenitors and zeta markers in epidermal progenitors (fig. S9D). These clusters are mostly devoid of X1 cells (Fig. 3C) and therefore correspond mainly to postmitotic progenitors.

RNA velocity confirms lineage relationships predicted by PAGA

To independently validate the differentiation trajectories predicted by PAGA, we used velocity (44). This method computes RNA velocity, defined as the rate of change of mRNA levels for a gene in time, in every single cell. In differentiating cells in which changes in gene expression are dominated by changes in transcription rates, the ratio of unspliced to spliced reads for a given gene within a cell will be proportional to the temporal change of the logarithm of spliced reads (or mature mRNAs) (44). Thus, one can estimate the future mRNA level of a gene by computing its velocity and a linear fit. By aggregating over many genes in a cell, one can estimate the cellular expression state to which the cell is apparently moving in time. We estimated mRNA velocities for each cell and projected the estimated future cell states onto the tSNE, which describe the paths predicted by the mRNA velocity model (Fig. 3E and fig. S12A). These paths show a highly homogeneous stem cell population that moves slowly to progenitors, which will differentiate to mature cell types. The long arrows at the edges of the clusters likely are due to the averaging on the force field, as they do not appear when individual arrows are plotted (fig. S12A). These paths largely agree with the trajectories predicted by PAGA and also confirmed the connection between muscle progenitors and pharynx muscle predicted from gene expression changes (fig. S12A). Additionally, velocity can also model longer cell trajectories to identify

their root (Fig. 3F) and terminal end points (Fig. 3G), which corresponded to the tSNE regions containing stem cells and terminally differentiated cells, respectively. Velocityto cannot provide information from disconnected clusters. As a result, all disconnected clusters contain differenti-

ation trajectories with independent start and end points (Fig. 3, F and G).

The estimates of RNA dynamics obtained with velocityto also identified regions where genes are mainly induced or repressed compared to the steady-state level. This information can be help-

ful to investigate relations between clusters that appear disconnected on the tSNE. We used these estimates to study the expression of marker genes from the epidermis cluster. These genes are clearly induced in epidermal progenitors and become repressed in mature epidermis, where

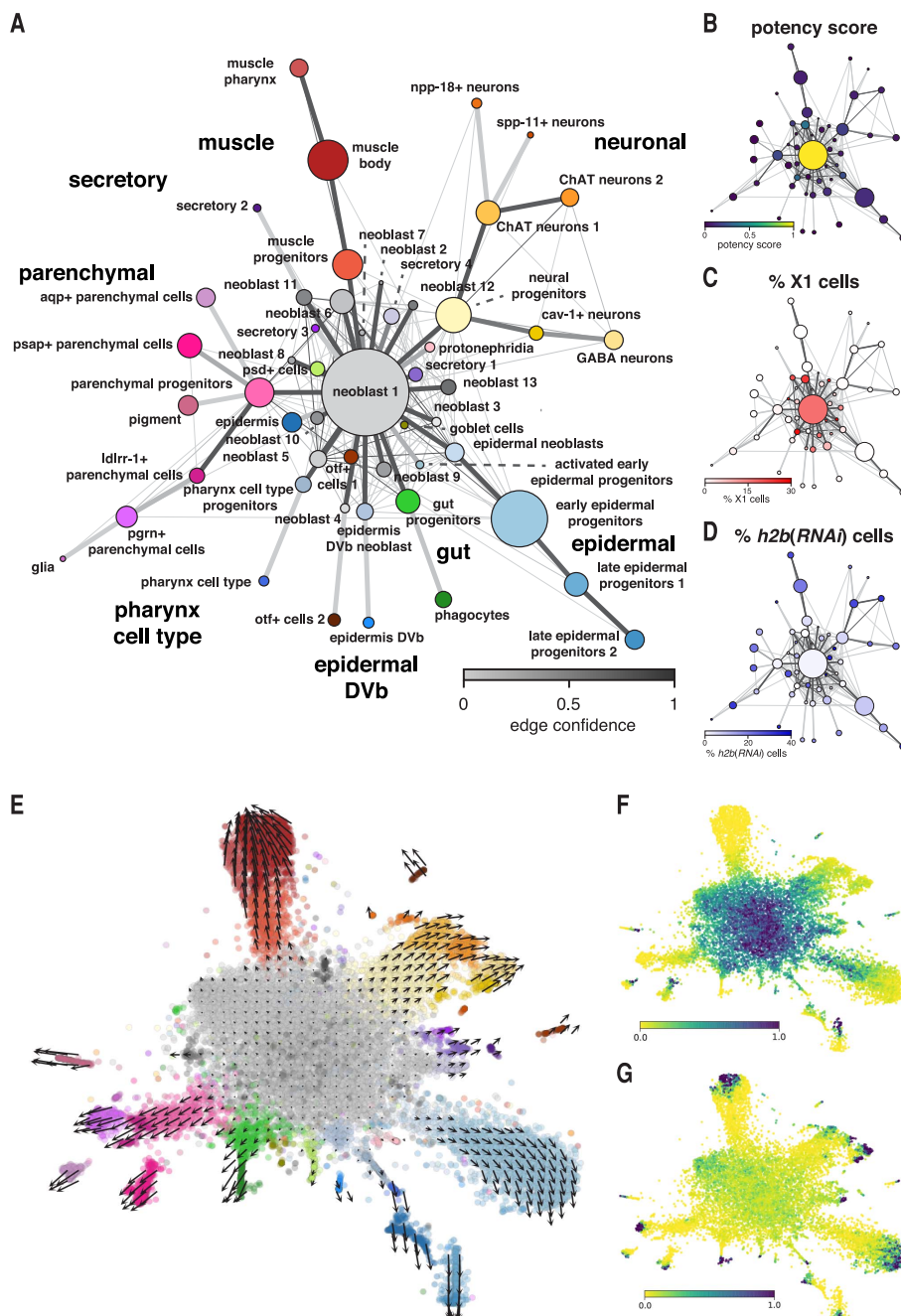


Fig. 3. Lineage tree reconstruction by PAGA and velocityto. (A) Abstracted graph showing all the possible edges with a probability higher than 10^{-6} connecting two clusters and their confidence. Each node corresponds to each of the clusters identified with Seurat. The size of nodes is proportional to the amount of cells in the cluster. The most probable path connecting the clusters is plotted on top, with thicker edges. (B) Lineage tree colored according to potency score, which ranges from blue (0) to yellow (1). (C and D) Lineage trees colored according to the percentage of X1

(C) or *h2b(RNAi)*-resistant (D) cells in each cluster. X1 cells are most abundant in the neoblast 1 cluster, whereas *h2b(RNAi)*-resistant cells are mostly located in the leaves of the tree. (E) Velocityto force field showing the average differentiation trajectories (velocity) for cells located in different parts of the tSNE plot. (F and G) Root (F) and terminal end points (G) obtained after modeling the transition probabilities derived from the RNA velocity by using a Markov process. The color scale represents the density of the end points of the Markov process and ranges from yellow (low) to blue (high).

they are highly expressed (fig. S12B). Thus, mRNA metabolism patterns provide additional support to the differentiation trajectory connecting late epidermal progenitors to epidermis that we predicted on the basis of gene expression changes (fig. S8A).

A consolidated lineage tree of planarian stem cell differentiation into all major cell types

Taken together, our results show that both computational and experimental methods agree in the identification of stem cells, progenitors, and differentiated cells. By combining all four independent lines of evidence (PAGA, gene expression changes, stem cell ablation and enrichment experiments, and velocyto), we provide a single consolidated tree that models stem cell differentiation trajectories into all identified cell types of adult planarians (Fig. 4A). The resulting cell lineage tree correctly recapitulates the known expression changes described during epidermal differentiation (26, 34, 45) (Fig. 4B). We observed a continuous decrease of the expression of *Smedwi-1*, a well-characterized neoblast marker (fig. S3), with pseudotime progression, whereas early (*prog-1*) and late (*agat-1*) epidermal progenitor, as well as mature epidermis (*vim-1*) markers, increased their expression at consecutive time points (Fig. 4B).

According to the consolidated lineage tree, neoblasts (35% of our wild-type cells) differentiate into at least 23 independent cell lineages. There are six major differentiation fates (57% of cells) (table S2), each representing more than 1% of total cells: epidermal, parenchymal, neural, muscle, gut, and a pharynx cell type. For these major fates, we identified progenitor and differentiated states. Additionally, we identified 10 minor lineages (6% of cells; each less abundant than 1% of total cells) that differentiate from the neoblasts, but for which we were unable to identify progenitors.

Self-organizing maps identify gene programs underlying cell differentiation

We used our data to identify gene sets that coordinately change their expression during differentiation. For this analysis, we discarded all cells from neoblast clusters that did not give rise to differentiated cell types in our consolidated cell lineage tree. The remaining cells were ordered following the tree for each lineage and sorted within each cluster according to their pseudotemporal ordering (Fig. 5A). Subsequently, we used self-organizing maps (SOMs) (46) to identify 48 sets of highly variable genes that coordinately change their expression during differentiation (47) (Fig. 5B, fig. S13, and table S4). Many of these sets contain some genes previously known to be expressed in the respective lineages and in some cases involved in their differentiation (table S5). For instance, gene sets 10 and 11 contain genes that are highly expressed in neoblast and progenitor clusters, such as *Smedwi-1* and *tub-α1*, whose expression drops during differentiation (Fig. 5B and fig. S13). Similarly, we found gene

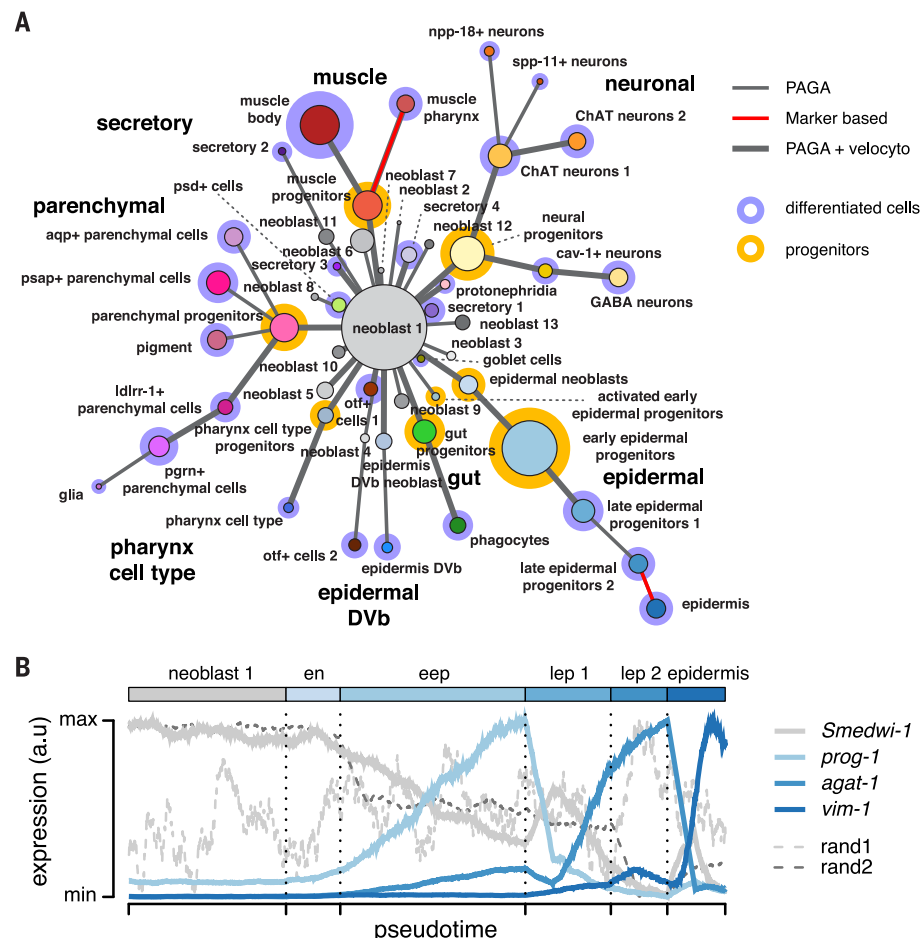


Fig. 4. Consolidated lineage tree of planarian stem cell differentiation into all major types.

(A) Consolidated lineage tree including four independent sources of evidence. The topology of the tree is shown according to PAGA, and marker-based connections are shown with red edges. Velocyto-supported connections are shown with thick edges. Progenitor and differentiated cell clusters according to neoblast ablation and enrichment experiments are shown with yellow and blue halos, respectively. (B) Gene expression changes of marker genes for the individual stages during epidermal differentiation (in pseudotime). Relative expression of marker genes from neoblast (*Smedwi-1*), early (*prog-1*) and late (*agat-1*) progenitors, as well as from the epidermis (*vim-1*). A maximum of 1000 cells from neoblast 1, epidermal neoblasts (en), early epidermal progenitors (eep), late epidermal progenitors 1 (lep 1) and 2 (lep 2), and epidermis were sampled. Gray thin dashed lines show the expression of *Smedwi-1* after randomly permuting cells (rand1) or after randomly sorting cells within each cluster (rand2).

sets that are regulated along muscle, neuronal, parenchymal, gut, and epidermal differentiation (Fig. 5B, top row). They contain genes expressed in these lineages, such as *mhc* for the muscle and *chat* for the neuronal lineage, but also include well-known regulators of their differentiation, such as *myoD* (48) and *cof* (49). As a consequence of analyzing all detected planarian cell lineages simultaneously, we not only identified gene sets involved in lineage specific programs but also gene sets co-regulated during the differentiation of several fates (Fig. 5B, middle and bottom row). Taken together, these results show that single-cell transcriptomics of a whole organism allows the reconstruction of specific differentiation events for many differentiation fates in parallel, enabling the identi-

fication of previously undetected combinations of co-regulated genes.

Molecular profiling of planarian regeneration by single-cell transcriptomics

Freshwater planarians are well known for their remarkable regenerative capacities. Planarians can be cut into small pieces, and each piece (except for the pharynx and the most anterior tip of the head, which are devoid of neoblasts) can regenerate a complete, albeit much smaller, organism in a matter of days. This process is dynamically complex and involves the orchestration of all cellular differentiation pathways. The animal does not grow (as it cannot eat) during the process. Thus, the truncated body fragments need to

reshape their body proportions to adjust to their new size by a process termed morphallaxis (50). It is still largely unknown how each individual cell type behaves in this process.

Given the detected cell type abundances and the cell differentiation tree of steady-state adult animals, we asked if we could use Drop-seq to profile the cellular and transcriptomic changes that occur during regeneration. We cut planarians into five to seven pieces, discarded the head piece, and prepared the remaining body pieces for single-cell transcriptomics immediately after cutting (day 0), and 2 and 4 days after cutting (Fig. 6A and fig. S14). We compared regenerating samples to those at day 0 using Seurat and detected hundreds of differentially expressed genes in both samples (tables S6 and S7). By pooling all cells, we could detect up-regulation after 2 days of regeneration of 16 of the 128 wound-induced genes described in a previous study (42, 51) (table S8 and fig. S15A). The shallowness of Drop-seq data makes it difficult to assess differences in lowly expressed genes. However, Drop-seq allows

the cell types that undergo these changes to be distinguished, showing that *run1-1* and *egr-2* are up-regulated in the neoblast 1 cluster (fig. S15B) and *jun-1* in the muscle body cluster (fig. S15C and tables S6 and S7).

All cells from the regenerating samples fall into clusters that are present in wild-type samples, indicating an absence of regeneration-specific types or trajectories (table S2). However, our analysis revealed notable changes in cell composition during regeneration (Fig. 6B and fig. S16). On one hand, we observed a large increase in the number of neoblasts, consistent with an increase in mitotic activity, and of neural progenitors, reflecting active neurogenesis to replace missing brain structures after head removal. On the other hand, we detected that both parenchymal progenitor cells and differentiated parenchymal cell types were depleted (Fig. 6B), indicating that these cells are cleared in the process of reshaping the planarian tissue. The cell proportion changes at days 2 and 4 were clearly correlated (Fig. 6C), indicating that

aqp-positive parenchymal cells are the most depleted cell type. We experimentally confirmed this observation by in situ hybridization on planarian tissue sections (Fig. 6D) and counting *aqp*-positive parenchymal cells (Fig. 6E) (Mann Whitney U test p -value $< 10^{-7}$). Our results indicate that parenchymal cells are highly depleted upon regeneration, implying that they may be used to metabolically fuel the regeneration process (52).

Discussion

In this study, we made use of the stem cell population and the extreme regeneration capabilities of adult flatworms to generate an atlas of cell types at high resolution. We identified, quantified, and molecularly characterized 37 cell types, including 23 terminally differentiated cell types, and numerous progenitor and stem cell clusters. Although our sequencing data are relatively shallow, molecular characterization of cell types with computational methods was robust, agreed well with previously published microscopy data, and revealed progenitor and differentiated cells. This implies that the grouping of incomplete transcriptomes of thousands of cells into clusters did not suffer appreciably from capture rates or other confounding factors.

The resolution of our data depends on both the number of cells sequenced and the number of genes detected per cell. Considering only wild-type and control samples (~11,000 cells), we can identify differentiated cell clusters containing about 10 cells. Therefore, we estimate that cells present at a frequency of $<1/1000$, such as *cintillo*+ cells (53) and photoreceptor neurons (33), will be missed by our approach. Furthermore, we failed to identify certain neoblast subpopulations previously described in the literature (26, 42). This result could be due to the low sensitivity of Drop-seq, which captures only a fraction of mRNAs in a cell. However, we do detect the expression of the proposed marker genes of these subpopulations. They appear to be spread among neoblasts and progenitor clusters (fig. S9), which still express neoblast markers such as *Smedwi-1* at low levels (figs. S10 and S11). This result indicates that the boundary between stem cells and lineage-committed progenitors is probably not sharp. Further studies will help in describing and delimiting these boundaries.

Projecting high-dimensional gene expression data of thousands of transcriptomes onto a two-dimensional plot [for example, by the widely used tSNE method (54)] visually reveals clusters. However, it is impossible to infer the relationships among them, as the distances between clusters cannot be interpreted as differentiation trajectories. To solve this problem, we used computational and experimental methods to reconstruct a lineage tree. PAGA and velocity provide two complementary approaches to study cell differentiation by using single-cell transcriptomics. Whereas velocity allows the differentiation trajectories (including their directionality) of individual cells to be located within a cell continuum on the basis of mRNA metabolism, PAGA allows the average differentiation paths of a group of

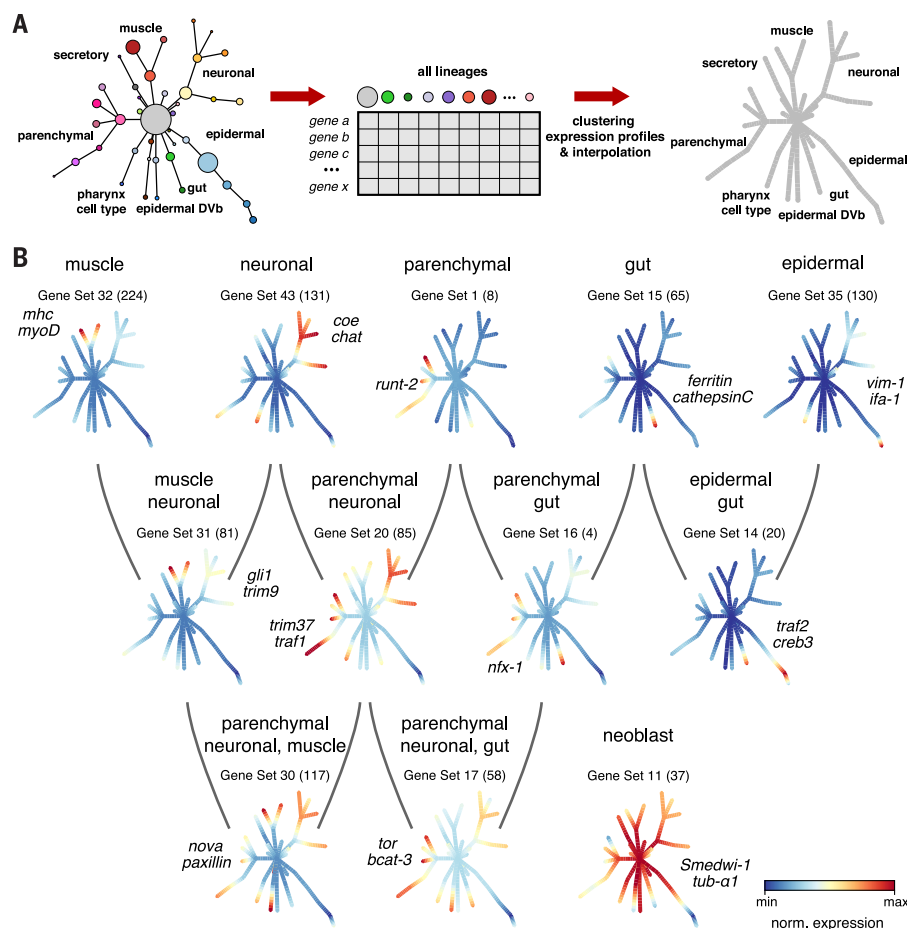


Fig. 5. Identification of gene sets regulated and co-regulated in cell differentiation. (A) Schematic workflow of the analysis performed to identify gene sets involved in the differentiation processes. Pseudotemporal ordering of the cells from all lineages and clustering of variable genes using SOMs allowed the identification of 48 gene sets. (B) Graphical representation of gene expression changes during cell differentiation of 12 gene sets. For each gene set, the normalized expression of the genes is shown on the edges of the tree and ranges from blue (low expression) to red (high expression). Next to each tree, representative genes from each gene set are highlighted.

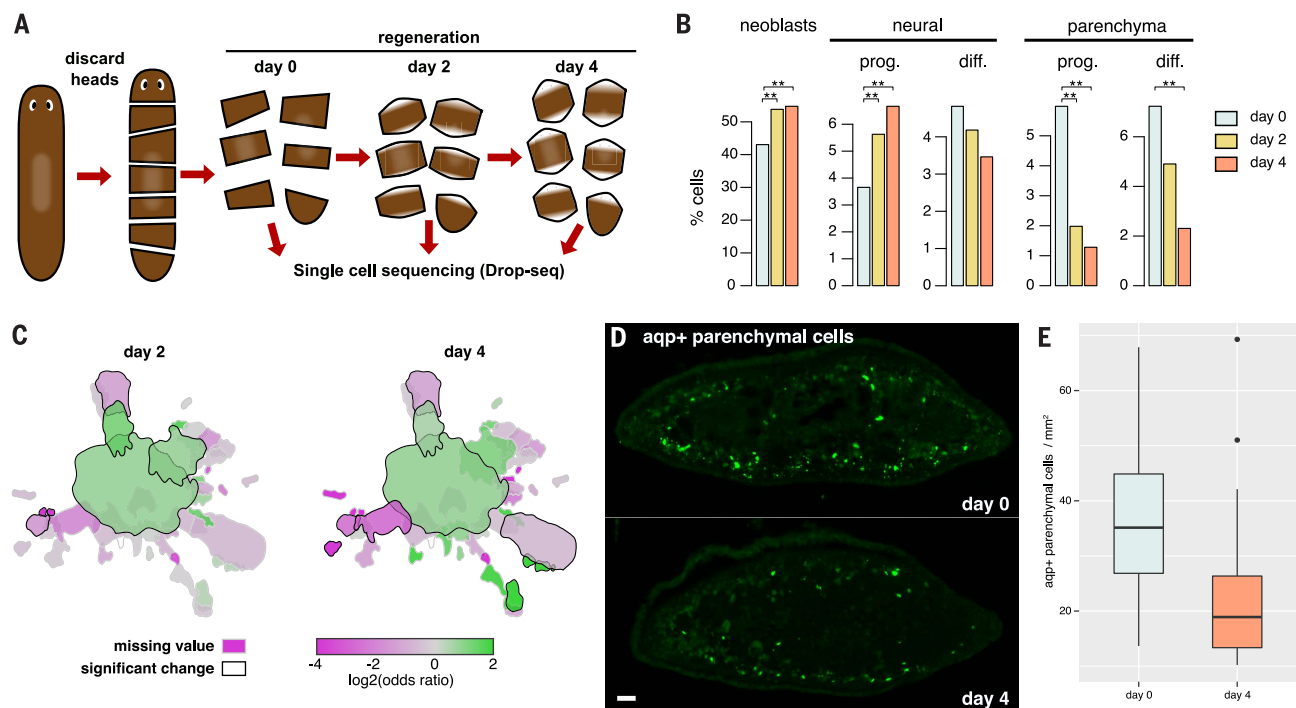


Fig. 6. Molecular profiling of regeneration by single-cell transcriptomics.

(A) Experimental workflow: Planarians were cut into small pieces; head pieces were discarded and the remaining pieces were processed for single-cell RNA sequencing 0, 2, and 4 days after cutting. (B) Quantification of neoblasts, neural progenitors, and differentiated clusters and parenchymal progenitors and differentiated clusters. Significant differences calculated using Fisher's exact test with an adjusted p -value < 0.001 are marked

with **. (C) Cluster outlines colored according to the log2 (odds ratio) of changes in regeneration at day 2 (left) and day 4 (right) versus day 0, showing enriched clusters in green colors and depleted clusters in magenta colors. Significant changes are indicated by black solid outlines. (D and E) In situ hybridization on sections (D) and quantification (E) of aqp+ parenchymal cells in regenerating planarians after 0 and 4 days of regeneration. Mann Whitney U test p -value $< 10^{-7}$. Scale bar: 100 μ m.

cells to be inferred, even when they are disconnected. Thus, combining these two computational methods results in a robust lineage prediction. This prediction is supported by the continuity of expression of marker genes and the mapping of stem cells and differentiated cells on the tree (Fig. 4A), and validates known differentiation trajectories such as that of the epidermal lineage (Fig. 4B) (34).

We used PAGA (22) to reconstruct in an unbiased way the lineage tree of all major planarian cells. This method, although indirect, allows the lineage information to be reconstructed from the transcriptomic snapshot of individual cells. Additionally, in contrast to high-throughput lineage-tracing methods (9–12), which rely on the use of transgenic or CRISPR-Cas tools, it can be applied to every species provided that single cells can be isolated and sequenced. Using this method, we identified de novo planarian stem cells and predict their differentiation paths to at least 23 different lineages, including several multipotent progenitor populations. Notably, these tools can readily be applied to other organisms to identify de novo stem cells, identify their differentiation trajectories, and estimate the developmental potency of the resulting cell populations.

Pseudotemporal ordering of cells along these lineages allowed us to discover gene sets that are putatively involved in differentiation programs, highlighting the similarities and differ-

ences that exist across tissues and identifying several genes known to be involved in cell differentiation not only in planarians but also in other species, such as *myoD*, *nkx6*, and *pax6*. Further characterization of these gene sets should be the subject of future studies. As we show for the *h2b* RNAi phenotype, Drop-seq also allows profiling of perturbation studies at both the transcriptomic and cellular levels. In general, and beyond planaria, we can foresee that future studies will use single-cell transcriptomics coupled to loss-of-function experiments to unravel the specific developmental functions of genes or regulatory networks.

Furthermore, we used single-cell transcriptomics to profile cellular abundance changes upon regeneration. Our experiments revealed that several of our newly described parenchymal cell types are depleted in regeneration. These cell types had been largely overlooked in molecular studies but had been described, on the basis of microscopy, in the literature decades ago. This is in part due to the unbiased nature of both microscopic and single-cell transcriptomic data. These parenchymal cells are highly enriched in lysosomes and other vacuoles and might be an energy reservoir that regenerating planarians mobilize to fuel regeneration.

To make our data easily accessible, we built an interactive app that allows users to query and interpret all sequencing data ([https://shiny.mdc-](https://shiny.mdc-berlin.de/pasca)

[berlin.de/pasca](https://shiny.mdc-berlin.de/pasca)). We also provide a detailed tutorial for the lineage reconstruction algorithm PAGA that we hope will serve as a reference for future studies (https://github.com/rajewsky-lab/planarian_lineages). Together, our results show that single-cell expression profiling can be used to systematically annotate cell types of entire animals, to reconstruct stem cell differentiation lineages and identify gene sets that likely contain genes involved in programming them, and to study complex processes such as regeneration (and their relation to lineages used in normal development) at single-cell resolution. Our results and methods demonstrate that single-cell approaches will become an indispensable method to study developmental and regeneration biology.

Methods summary

Single-cell transcriptomic profiling of asexual adult planarians from the species *S. mediterranea* was performed with Drop-seq. Single-cell suspensions were prepared by dissociating cells from adult planarians with trypsin. We used FACS to discard broken and dead cells. Cells were either directly processed for Drop-seq or preserved in methanol or DMSO for later processing. For RNAi experiments, animals were injected with double-stranded RNA against the coding region of *h2b* or *gfp* for three consecutive days and kept at 20°C; their cells were prepared for FACS and single-cell transcriptomics 5 days after the third

injection. For regeneration experiments, animals ranging from 4 to 10 mm in size were cut into five to seven pieces; the head pieces were discarded, and the remaining pieces were processed for Drop-seq immediately, or 2 or 4 days after cutting.

Computational analysis of the sequenced samples was done with Drop-seq tools and the Seurat package (3). Briefly, reads were mapped to the *S. mediterranea* dd_Smed_v6 transcriptome and processed with Drop-seq tools and custom perl scripts to generate Digital Gene Expression (DGE) matrices for each sample. Finally, all DGE matrices were joined. Variable genes across all clusters were used to perform a PCA. The first 50 principal components obtained were then tested for statistical significance, and those with a p -value $< 10^{-5}$ were used to perform clustering. The robustness of the obtained clusters was assessed, and spurious clusters were merged to obtain a final set of 51 clusters.

Cell type identification was performed by calculating marker genes for each cluster. Manual inspection, comparison to previously published single-cell data, and experimental validation using in situ hybridizations of the marker genes reported allowed the identification of the different cell populations. Additional characterization of the identified cell types was performed by characterizing GO-term-based gene sets with PAGODA (40). Experimental validation of cell types was done by whole-mount in situ hybridization and in situ hybridization on histological sections as previously described and using probes complementary to marker genes (table S9).

Lineage reconstruction was done by combining the unsupervised graph obtained with the PAGA algorithm (22) with velocity (44), gene expression analysis, and experimental data from *h2b(RNAi)* and X1 FACS-sorted cells. To calculate RNA velocity with velocity, we mapped the reads from all data sets to the planarian genome to extract spliced and unspliced reads. These analyses allowed pseudotemporal ordering of cells that was used to identify gene sets that change during stem cell differentiation by using self-organizing maps.

To perform cell counting of regenerating planarians, positive cells were automatically counted by using a custom script for ImageJ (<https://imagej.net>).

REFERENCES AND NOTES

1. A. Regev et al., The Human Cell Atlas. *eLife* **6**, e27041 (2017). doi: [10.7554/eLife.27041](https://doi.org/10.7554/eLife.27041); pmid: 29206104
2. D. A. Jaitin et al., Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014). doi: [10.1126/science.1247651](https://doi.org/10.1126/science.1247651); pmid: 24531970
3. E. Z. Macosko et al., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015). doi: [10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002); pmid: 26000488
4. K. Shekhar et al., Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323.e30 (2016). doi: [10.1016/j.cell.2016.07.054](https://doi.org/10.1016/j.cell.2016.07.054); pmid: 27565351
5. A. C. Villani et al., Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). doi: [10.1126/science.aah4573](https://doi.org/10.1126/science.aah4573); pmid: 28428369
6. D. Grün et al., Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015). doi: [10.1038/nature14966](https://doi.org/10.1038/nature14966); pmid: 26287467
7. N. Karaikos et al., The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017). doi: [10.1126/science.aan3235](https://doi.org/10.1126/science.aan3235); pmid: 28860209
8. J. Cao et al., Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017). doi: [10.1126/science.aam8940](https://doi.org/10.1126/science.aam8940); pmid: 28818938
9. A. McKenna et al., Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016). doi: [10.1126/science.aaf7907](https://doi.org/10.1126/science.aaf7907); pmid: 27229144
10. K. L. Frieda et al., Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017). doi: [10.1038/nature20777](https://doi.org/10.1038/nature20777); pmid: 27869821
11. B. Spanjaard et al., Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat. Biotechnol.* **2018**, 24 (2018). doi: [10.1038/nbt.4124](https://doi.org/10.1038/nbt.4124); pmid: 29644996
12. A. Alemany, M. Florescu, C. S. Baron, J. Peterson-Maduro, A. van Oudenaarden, Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018). doi: [10.1038/nature25969](https://doi.org/10.1038/nature25969); pmid: 29590089
13. L. Haghighi, M. Büttner, F. A. Wolf, F. Büttner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016). doi: [10.1038/nmeth.3971](https://doi.org/10.1038/nmeth.3971); pmid: 27571553
14. M. Setty et al., Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637–645 (2016). doi: [10.1038/nbt.3569](https://doi.org/10.1038/nbt.3569); pmid: 27136076
15. X. Qiu et al., Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). doi: [10.1038/nmeth.4402](https://doi.org/10.1038/nmeth.4402); pmid: 28825705
16. D. Grün et al., De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016). doi: [10.1016/j.stem.2016.05.010](https://doi.org/10.1016/j.stem.2016.05.010); pmid: 27345837
17. F. Notta et al., Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016). doi: [10.1126/science.aab2116](https://doi.org/10.1126/science.aab2116); pmid: 26541609
18. J. Shin et al., Single-cell RNA-Seq with Waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17**, 360–372 (2015). doi: [10.1016/j.stem.2015.07.013](https://doi.org/10.1016/j.stem.2015.07.013); pmid: 26299571
19. B. Treutlein et al., Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014). doi: [10.1038/nature13173](https://doi.org/10.1038/nature13173); pmid: 24739965
20. L. Velten et al., Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017). doi: [10.1038/ncb3493](https://doi.org/10.1038/ncb3493); pmid: 28319093
21. D. E. Wagner, I. E. Wang, P. W. Reddien, Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* **332**, 811–816 (2011). doi: [10.1126/science.1203983](https://doi.org/10.1126/science.1203983); pmid: 21566185
22. F. A. Wolf et al., Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *bioRxiv* 208819 [Preprint]. 25 October 2017.
23. T. Hayashi, M. Asami, S. Higuchi, N. Shibata, K. Agata, Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev. Growth Differ.* **48**, 371–380 (2006). doi: [10.1111/j.1440-169X.2006.00876.x](https://doi.org/10.1111/j.1440-169X.2006.00876.x); pmid: 16872450
24. P. Önal et al., Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO J.* **31**, 2755–2769 (2012). doi: [10.1038/emboj.2012.110](https://doi.org/10.1038/emboj.2012.110); pmid: 22543868
25. R. M. Labbé et al., A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**, 1734–1745 (2012). doi: [10.1002/stem.1144](https://doi.org/10.1002/stem.1144); pmid: 22696458
26. J. C. van Wolfswinkel, D. E. Wagner, P. W. Reddien, Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell* **15**, 326–339 (2014). doi: [10.1016/j.stem.2014.06.007](https://doi.org/10.1016/j.stem.2014.06.007); pmid: 25017721
27. J. Solana et al., Conserved functional antagonism of CELF and MBNL proteins controls stem cell-specific alternative splicing in planarians. *eLife* **5**, e16797 (2016). doi: [10.7554/eLife.16797](https://doi.org/10.7554/eLife.16797); pmid: 27502555
28. J. Alles et al., Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017). doi: [10.1186/s12915-017-0383-5](https://doi.org/10.1186/s12915-017-0383-5); pmid: 28526029
29. A. Guillaumet-Adkins et al., Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* **18**, 45 (2017). doi: [10.1186/s13059-017-1171-9](https://doi.org/10.1186/s13059-017-1171-9); pmid: 28249587
30. J. Baguña, R. Romero, Quantitative analysis of cell-types during growth, Degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia Tigrina*. *Hydrobiologia* **84**, 181–194 (1981). doi: [10.1007/BF00026179](https://doi.org/10.1007/BF00026179)
31. F. Cebrià, P. A. Newmark, Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. *Development* **134**, 833–837 (2007). doi: [10.1242/dev.02794](https://doi.org/10.1242/dev.02794); pmid: 17521262
32. J. C. Rink, H. T.-K. Vu, A. Sánchez Alvarado, The maintenance and regeneration of the planarian excretory system are regulated by EGFR signaling. *Development* **138**, 3769–3780 (2011). doi: [10.1242/dev.066852](https://doi.org/10.1242/dev.066852); pmid: 21828097
33. S. W. Lapan, P. W. Reddien, Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell Reports* **2**, 294–307 (2012). doi: [10.1016/j.celrep.2012.06.018](https://doi.org/10.1016/j.celrep.2012.06.018); pmid: 22884275
34. O. Wurtzel, I. M. Odenberg, P. W. Reddien, Planarian epidermal stem cells respond to positional cues to promote cell-type diversity. *Dev. Cell* **40**, 491–504.e5 (2017). doi: [10.1016/j.devcel.2017.02.008](https://doi.org/10.1016/j.devcel.2017.02.008); pmid: 28292427
35. K. J. Pedersen, Studies on the nature of planarian connective tissue. *Z. Zellf. Mikrosk. Anat.* **53**, 569–608 (1961). doi: [10.1007/BF00339508](https://doi.org/10.1007/BF00339508)
36. B. M. Stubenhaus et al., Light-induced depigmentation in planarians models the pathophysiology of acute porphyrias. *eLife* **5**, e14175 (2016). doi: [10.7554/eLife.14175](https://doi.org/10.7554/eLife.14175); pmid: 27240733
37. C. Wang et al., Forkhead containing transcription factor Albino controls tetrapyrrole-based body pigmentation in planarian. *Cell Discov.* **2**, 16029 (2016). doi: [10.1038/celldisc.2016.29](https://doi.org/10.1038/celldisc.2016.29); pmid: 27551436
38. I. E. Wang, S. W. Lapan, M. L. Scimone, T. R. Clandinin, P. W. Reddien, Hedgehog signaling regulates gene expression in planarian glia. *eLife* **5**, e16996 (2016). doi: [10.7554/eLife.16996](https://doi.org/10.7554/eLife.16996); pmid: 27612382
39. R. H. Roberts-Galbraith, J. L. Brubacher, P. A. Newmark, A functional genomics screen in planarians reveals regulators of whole-brain regeneration. *eLife* **5**, e17002 (2016). doi: [10.7554/eLife.17002](https://doi.org/10.7554/eLife.17002); pmid: 27612384
40. J. Fan et al., Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016). doi: [10.1038/nmeth.3734](https://doi.org/10.1038/nmeth.3734); pmid: 26780092
41. J. Solana et al., Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome Biol.* **13**, R19 (2012). doi: [10.1186/gb-2012-13-3-r19](https://doi.org/10.1186/gb-2012-13-3-r19); pmid: 22439894
42. O. Wurtzel et al., A generic and cell-type-specific wound response precedes regeneration in planarians. *Dev. Cell* **35**, 632–645 (2015). doi: [10.1016/j.devcel.2015.11.004](https://doi.org/10.1016/j.devcel.2015.11.004); pmid: 26651295
43. A. M. Molinaro, B. J. Pearson, In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome Biol.* **17**, 87 (2016). doi: [10.1186/s13059-016-0937-9](https://doi.org/10.1186/s13059-016-0937-9); pmid: 27150006
44. G. La Manno et al., RNA velocity in single cells. *bioRxiv* 206052 [Preprint]. 19 October 2017.
45. G. T. Eisenhoffer, H. Kang, A. Sánchez Alvarado, Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* **3**, 327–339 (2008). doi: [10.1016/j.stem.2008.07.002](https://doi.org/10.1016/j.stem.2008.07.002); pmid: 18786419
46. T. Kohonen, The self-organizing map. *Proc. IEEE* **78**, 1464–1480 (1990). doi: [10.1109/5.58325](https://doi.org/10.1109/5.58325)
47. P. Tamayo et al., Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2907–2912 (1999). doi: [10.1073/pnas.96.6.2907](https://doi.org/10.1073/pnas.96.6.2907); pmid: 10077610
48. M. L. Scimone, L. E. Cote, P. W. Reddien, Orthogonal muscle fibres have different instructive roles in planarian regeneration. *Nature* **551**, 623–628 (2017). doi: [10.1038/nature24660](https://doi.org/10.1038/nature24660); pmid: 29168507
49. M. W. Cowles, K. C. Omuro, B. N. Stanley, C. G. Quintanilla, R. M. Zayas, COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians. *PLoS Genet.* **10**, e1004746 (2014). doi: [10.1371/journal.pgen.1004746](https://doi.org/10.1371/journal.pgen.1004746); pmid: 25356635
50. K. Agata, Y. Saito, E. Nakajima, Unifying principles of regeneration I: Epimorphosis versus morphallaxis. *Dev. Growth Differ.* **49**, 73–78 (2007). doi: [10.1111/j.1440-169X.2007.00919.x](https://doi.org/10.1111/j.1440-169X.2007.00919.x); pmid: 17335428

51. D. Wenemoser, P. W. Reddien, Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Dev. Biol.* **344**, 979–991 (2010). doi: [10.1016/j.ydbio.2010.06.017](https://doi.org/10.1016/j.ydbio.2010.06.017); pmid: [20599901](https://pubmed.ncbi.nlm.nih.gov/20599901/)
52. C. González-Estévez, D. A. Felix, A. A. Aboobaker, E. Saló, Gtdap-1 promotes autophagy and is required for planarian remodeling during regeneration and starvation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13373–13378 (2007). doi: [10.1073/pnas.0703588104](https://doi.org/10.1073/pnas.0703588104); pmid: [17686979](https://pubmed.ncbi.nlm.nih.gov/17686979/)
53. N. J. Oviedo, P. A. Newmark, A. Sánchez Alvarado, Allometric scaling and proportion regulation in the freshwater planarian *Schmidtea mediterranea*. *Dev. Dyn.* **226**, 326–333 (2003). doi: [10.1002/dvdy.10228](https://doi.org/10.1002/dvdy.10228); pmid: [12557210](https://pubmed.ncbi.nlm.nih.gov/12557210/)
54. L. Van Der Maaten, G. Hinton, Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

ACKNOWLEDGMENTS

We thank J. Alles and A. Boltengagen for help with single-cell sequencing and members of the Rajewsky lab for discussions. We

thank S. Linnarsson and G. La Manno for help running velocity. We also thank N. Friedman and A. García-Pérez for helpful comments. **Funding:** Work on this project was funded by the German Center for Cardiovascular Research (DZHK BER 1.2 VD), the DFG (grant RA 838/5-1), the BMBF (grant 01ZX1711A), and the Helmholtz Association (Incubator grant sparse2big, grant ZT-I-0007). F.A.W. acknowledges the support of the Helmholtz Postdoc Program, Initiative and Networking Fund of the Helmholtz Association. **Author contributions:** J.S., M.P., C.K., and N.R. designed the project. J.S., C.K., and N.R. designed all experiments. S.A., J.S., and C.K. performed all experiments. M.P. performed and coordinated computational analyses. F.A.W. performed the PAGA analysis with the supervision of F.J.T. A.M. performed velocity analyses. P.G. developed the Shiny app and performed image analysis. B.O. developed the visualization of gene expression along the lineage tree. All authors discussed and interpreted the data. M.P., J.S., and N.R. wrote the manuscript with input from all other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The sequencing

data generated are available in Gene Expression Omnibus under the accession GSE103633. The single-cell data generated can be interactively accessed at <https://shiny.mdc-berlin.de/psca>. Manuals to run and reproduce velocity and PAGA are available on Github (https://github.com/rajewsky-lab/planarian_lineages).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6391/eaq1723/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S22

Tables S1 to S10

References (55–74)

9 October 2017; resubmitted 14 February 2018

Accepted 12 April 2018

Published online 19 April 2018

10.1126/science.aaq1723

Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics

Mireya Plass, Jordi Solana, F. Alexander Wolf, Salah Ayoub, Aristotelis Misios, Petar Glazar, Benedikt Obermayer, Fabian J. Theis, Christine Kocks and Nikolaus Rajewsky

Science **360** (6391), eaaq1723.

DOI: 10.1126/science.aag1723 originally published online April 19, 2018

Mapping the planarian transcriptome

A cell type's transcriptome defines the active genes that control its biology. Two groups used single-cell RNA sequencing to define the transcriptomes for essentially all cell types of a complete animal, the regenerative planarian *Schmidtea mediterranea*. Because pluripotent stem cells constantly differentiate to rejuvenate any part of the body of this species, all developmental lineages are active in adult animals. Fincher *et al.* determined the transcriptomes for most, if not all, planarian cell types, including some that were previously unknown. They also identified transition states and genes governing positional information. Plass *et al.* used single-cell transcriptomics and computational algorithms to reconstruct a lineage tree capturing the developmental progressions from stem to differentiated cells. They could then predict gene programs that are specifically turned on and off along the tree, and they used this approach to study how the cell types behaved during regeneration. These whole-animal transcriptome "atlases" are a powerful way to study metazoan biology.

Science, this issue p. eaaq1736, p. eaaq1723

ARTICLE TOOLS

<http://science.sciencemag.org/content/360/6391/eaaq1723>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/04/18/science.aag1723.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/360/6391/eaaq1736.full>

REFERENCES

This article cites 71 articles, 12 of which you can access for free
<http://science.sciencemag.org/content/360/6391/eaaq1723#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)