Molecular profile, source and lineage restriction of stem cells in an annelid regeneration model

Alexander W. Stockinger[1,2,3,4,5,*], Leonie Adelmann[1,2,3,4,5,*], Martin Fahrenberger[1,3,4,6,7], Christine Ruta[8], B. Duygu Özpolat[9,#], Guillaume Balavoine[9,@], Florian Raible[1,2,3,@]

1, Max Perutz Labs, Vienna Biocenter Campus (VBC), Vienna, Austria
2, University of Vienna, Center for Molecular Biology, Department of Genetics and Microbiology, Vienna, Austria
3, Research Platform Single-Cell Regulation of Stem Cells (SinCeReSt), University of Vienna, Vienna, Austria
4, Vienna Biocenter PhD Program, a Doctoral School of the University of Vienna and the Medical University of Vienna, Vienna, Austria
5, PhD Programme Stem Cells, Tissues, Organoids – Dissecting Regulators of Potency and Pattern Formation (SCORPION), University of Vienna, Vienna, Austria
6, Center for Integrative Bioinformatics Vienna (CIBIV), University of Vienna and Medical University of Vienna, Austria
7, Medical University of Vienna, Max Perutz Labs, Vienna, Austria.
8, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
9, Université de Paris Cité, CNRS, Institut Jacques Monod, Paris, France

* these authors contributed equally to this work
# present address: Department of Biology, Washington University in Saint Louis, MO, USA

@ correspondence for transgenic work: guillaume.balavoine@ijm.fr
@ correspondence for single-cell analyses: florian.raible@univie.ac.at

Abstract

Regeneration of missing body parts can be observed in diverse animal phyla, but it remains unclear to which extent these capacities rely on shared or divergent principles. Research into this question requires detailed knowledge about the involved molecular and cellular principles in suitable reference models. By combining single-cell RNA sequencing and mosaic transgenesis in the marine annelid *Platynereis dumerilii*, we map cellular profiles and lineage restrictions during posterior regeneration. Our data reveal cell-type specific injury responses, re-expression of positional identity factors, and the re-emergence of stem cell signatures in multiple cell populations. Epidermis and mesodermal coelomic tissue produce distinct posterior stem cells (PSCs) in the emerging blastema. A novel mosaic transgenesis strategy reveals both developmental compartments and lineage restrictions during regenerative growth. Our work supports the notion that posterior regeneration involves dedifferentiation, and reveals molecular and mechanistic parallels between annelid and vertebrate regeneration.

Introduction

The ability of some animals to regenerate missing body parts is a fascinating phenomenon. Whereas the regenerative ability of mammals is generally limited to individual cell types or a few specific organs, other animals are capable of rebuilding their entire body from mere fragments of tissue. Complex tissue regeneration can be observed in almost all clades of bilaterian life and might therefore reflect an ancient capacity [1,2]. This regenerative process usually involves the formation of an epimorphic, proliferative cell mass called blastema [3]. The cellular sources and molecular properties of blastema cells, however, differ between available model systems. For example, whereas some invertebrates like the planarian *Schmidtea mediterranea* form their blastemas from totipotent stem cells [4], other regenerative models employ less potent stem cells, such as uni- or oligopotent progenitors, and make use of dedifferentiation, transdifferentiation or a combination of these processes [3,5–7]. While molecular similarities between these regenerative strategies can be observed across phyla, we still lack comprehensive data on representative species to uncover whether or not these similarities indicate true homologies [3]. Defining molecular signatures of blastema formation in accessible model systems and regeneration paradigms will be a key requirement for such comparisons.

Annelids show particular promise as models to inform mechanisms and pathways of blastema-based regeneration. Many species within this clade exhibit the ability to regenerate large parts of their primary body axis. Due to the evolutionary placement of annelids in the lophotrochozoan superphylum, annelids are well-positioned for long-range comparisons with other lophotrochozoan clades, such as planarians, as well deuterostomes (which include vertebrates). Likewise, comparisons between different annelids provides an avenue to assess modulation of regenerative capacities within the clade [8–12]. This makes annelids ideal models to assess both commonalities and differences in regenerative processes.

The nereidid worm *Platynereis dumerilii* has a long history as a model system for regeneration and regenerative growth [8,9]. Its blastema-based regeneration upon posterior amputation along the primary body axis has been well characterized: in a first, rapid response to injury, the gut seals the wound. In a second step, epidermal cells cover the injury under a wound epithelium, followed by the formation of a largely undifferentiated blastemal cell mass. Proliferation then expands this population, and cells begin differentiating and rebuilding the missing structures [10,13]. Cytological evidence suggested the emergence of new stem cells in differentiated tissues such as the wound-adjacent epidermis. This process has classically been referred to as “re-embryonalisation” [14], suggesting the existence of some type of dedifferentiation mechanism. *Platynereis* has since become a valuable annelid reference model that can be continuously bred in the laboratory, and offers a variety of molecular and genetic tools for analysis. These include transcriptomic profiling, multiplexed detection of RNAs in fixed specimens, and transgenic manipulation [9,15–19].

More recent work has shed light on some molecular factors associated with posterior growth and regeneration. *Platynereis* grows by segmental addition throughout most of its adult life. This process involves posterior stem cells (PSCs) expressing members of an ancestral transcriptomic profile used in both stem cells and germ cells, called germline multipotency program (GMP) [20–22]. It has been suggested, but not demonstrated, that separate, molecularly distinct pools of PSCs produce the germ layers (ectoderm, mesoderm, endoderm) of new segments. After posterior amputation and thus removal of the PSCs, animals close the wound, form a blastema and, within 3 to 4 days, re-establish a proliferating zone of PSCs [13]. Recent bulk transcriptomic profiling and an analysis of epigenetic remodeling underlying this process found additional evidence for the fast upregulation of many GMP-related genes and extensive chromatin remodeling after injury [23,24]. While these results are consistent with cell reprogramming and dedifferentiation, and indicate a strong injury response in all three germ layers of wound-adjacent tissue also observed in other annelids [10,11], they lack the cellular resolution to compare tissue- and cell type specific properties

of regeneration across species. Additionally, no information regarding the differentiation potential / lineage restriction of *Platynereis* PSCs is currently available, further complicating comparative analyses. To gain deeper and unbiased insight into this process and enable cross-species comparisons of blastema based regeneration, molecular profiling at cellular resolution and clonal information of lineage restriction are required.

Here, we follow a dual approach of single-cell RNA sequencing and transgenesis experiments throughout posterior regeneration to address these challenges. By sampling single-cell transcriptomes at multiple regenerative stages and comparing them to wound-adjacent tissue right after injury, we are able to derive a comprehensive, time-resolved map of cellular profiles over the regenerative process. We detect cell-type specific injury responses and re-expression of positional identity factors. We also uncover that multiple wound-adjacent cell populations start expressing stem cell related genes and enter the cell cycle upon injury, consistent with the notion that these cells undergo dedifferentiation. Investigating signature genes for two of these populations, we identify the epidermis and mesodermal coelomic tissue as two likely source tissues that produce distinct PSCs in the segment addition zone. Capitalizing on a novel mosaic transgenesis strategy, we are able to identify both developmental compartments and restrictions in cell lineages throughout posterior growth and regeneration. We demonstrate that the segment addition zone (SAZ) of *Platynereis dumerilii* harbors separate pools of germ layer-restricted PSCs and that these pools are regenerated from cells of the corresponding germ layer. Our combined datasets provide a detailed view of the sources, molecular signatures and differentiation potential of major cell types in the blastema, and reveal molecular and mechanistic similarities between annelid and vertebrate regeneration.

Results

A temporally resolved single cell atlas reveals the dynamic transcriptional landscape of posterior regeneration

To establish a first unbiased, in-depth analysis of the transcriptomic landscape of individual cell populations during posterior regeneration in annelids, we devised a suitable sampling scheme. For this, we induced posterior regeneration by removing approximately ⅓ of the animals' posterior tissue, including the SAZ and its rapidly proliferating progeny, amputating between segments 30 and 31. For sampling, we then isolated the posteriormost segment along with any newly regenerated tissue at distinct time points after amputation (Supplementary Fig. 1a).

We reasoned that inclusion of the last non-amputated segment in these analyses would not only provide us with data on differentiated cell types, but also allow us to detect any molecular signatures associated with the response of this segment to the adjacent wound. To assess whether this sampling scheme captured relevant molecular events in the early phases of blastema formation, we first performed a bulk RNA sequencing experiment, in which the total mRNA of each sampling time point was sequenced from biological triplicates. By using an unbiased gene-clustering approach, we determined seven major categories of gene expression dynamics over the first three days of regeneration, including four categories in which gene expression increased after amputation, with differences in the point of onset and kinetics (Supplementary Fig. 1b). Genes in these categories include known markers for stem cells and the SAZ, as well as proliferation-related transcripts (Supplementary Fig. 1b). These findings are consistent with previous observations [13,23,24] and confirmed that our sampling strategy could be used to capture relevant molecular processes.

Based on these results, we devised a similar sampling scheme to build a comprehensive single cell atlas of posterior regeneration (Fig. 1a). We obtained single cells from multiple dissociated samples, representing five distinct stages of regeneration. These spanned from freshly amputated individuals (0 hpa, equivalent to uninjured trunk segments) to 72 hpa, corresponding to the onset of rapid proliferation in the regenerated SAZ [13] (Fig. 1a). After removing low-quality cells, we obtained a total of 80,298 cells, sampled in two independent biological replicates of 4 and 5 timepoints, respectively (Supplementary Data 1).

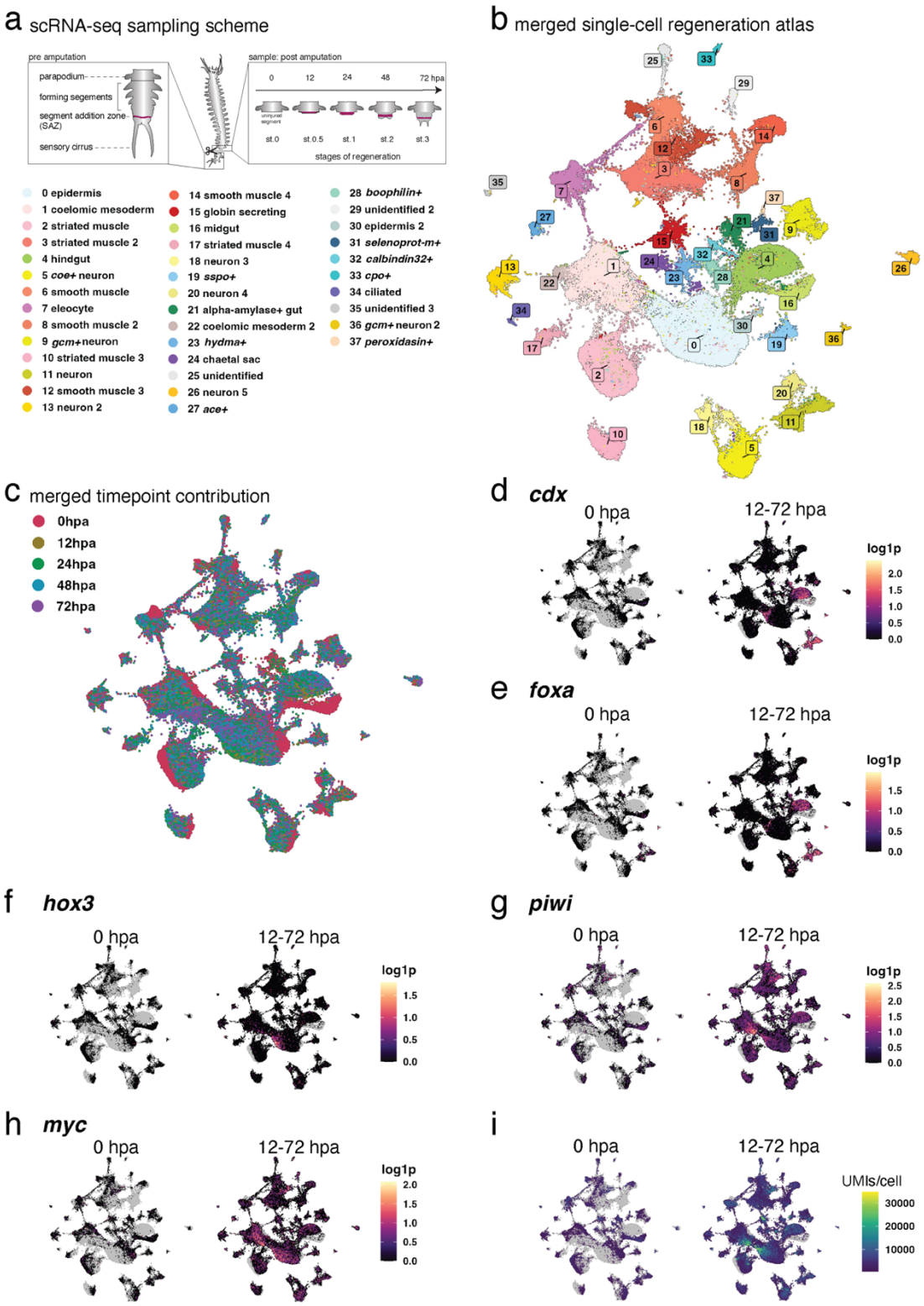


Figure 1. A temporally resolved single cell atlas reveals the dynamic transcriptional landscape of cell populations during posterior regeneration. (a) Sampling scheme illustrating posterior amputation and sampling timepoints, ranging from 0 hours post amputation (0 hpa, equivalent to a regular trunk segment) to 72 hpa, matching morphologically defined stages (st. 0 to 3). (b) UMAP visualization of cells, annotated by tentative cell type / population identity. (c) UMAP visualization showing the regenerative timepoint at which cells were sampled. (d,e) UMAP visualizations showing the expression of posterior identity markers (*cdx, foxa*) on the merged dataset, contrasting the freshly amputated sample (left) with the post-amputation time points (right) (f-i) similar UMAP visualizations, highlighting the changes in expression of stem-cell related genes (*hox3*, *piwi*, *myc*) (f-h) as well as the emergence of hypertranscriptomic cells (UMIs/cell) (i).

Single-cell data comprising multiple replicates or biological samples might suffer from batch effects, where technical differences between sampling rounds could overshadow biologically meaningful differences between samples or cell types [25]. To counter this effect and minimize technical variations, we took advantage of the recent establishment of a combined cell fixation and storage protocol (ACetic-MEthanol/ACME) that is compatible with single-cell sequencing [26]. The adaptation of this protocol for our *Platynereis* regenerate paradigm allowed us to sort, process and sequence cells from all sampled stages in parallel. We subsequently used standard single-cell RNA sequencing (scRNA-seq) analysis methods to process the joined data-set of 80,298 cells (see Methods). Unbiased clustering of the cells resulted in 38 transcriptionally distinct clusters. The comparison between biological replicates did not suggest any batch effect affecting cluster formation (Supplementary Data 1, Supplementary Fig. 2a). The resulting clusters, as illustrated on a uniform manifold approximation and projection (UMAP) visualization [27] (Fig. 1b), correspond to cell types / populations of similar transcriptomic profiles. We annotated these populations based on the identities of cluster-specific marker genes, and their expression levels of a list of known annelid cell-type markers. In total, we annotated 35 of the clusters, either as known cell populations, or based on their most diagnostic marker gene (Fig. 1b, see details in Supplementary Data 2 and 3).

As each sampled tissue contains the segment adjacent to the injury site, we expected to find data for most adult cell types located in annelid trunks. In line with this, we identify a variety of cell types in our dataset. For example, an investigation of genes previously used for assigning different *Platynereis* muscle cell types [28], allowed us to distinguish several populations of smooth (clusters 3, 6, 8, 12 and 14) and striated (clusters 2, 10, and 17) muscle, based on their respective expression of specific myosins [28]. Even less abundant cell types, such as chaetal sac cells (cluster 24) which form the bristle worm's chitinous bristles [29,30] and extracellular globin-secreting cells (cluster 15) [31], were identified as distinct populations. This shows that our approach yielded a high-quality cell atlas containing biologically meaningful clusters of cell populations and with sufficient sensitivity to resolve rare and poorly understood cell types.

Molecular repatterning and emerging stem cell-like properties in distinct cell populations

As outlined above, deconvolving the dynamic injury response to individual cell populations in an annelid is expected to advance our understanding of regeneration in an evolutionary context. By capitalizing on the temporal information embedded in each transcriptome of our dataset (Fig. 1c), we were able to perform comparisons of gene expression within cell populations across time.

A common challenge in complex tissue regeneration is the re-establishment of appropriate positional information, such as the position along the antero-posterior axis. To test whether our dataset could be used to identify the individual cell types involved in repatterning, we analyzed the expression of several transcription factors involved in posterior identity. Bulk RNA sequencing of posterior regeneration (Supplementary Fig. 1b) revealed the presence of genes encoding posteriorly expressed transcription factors such as *caudal* (*cdx*) and *distalless* (*dlx*), in two of the clusters upregulated after injury. This is consistent with previous suggestions that early steps in annelid regeneration include a morphallactic adjustment of positional values [32]. Using the single cell atlas, we were able to add cellular resolution to this process. For example, cells of the hindgut (cluster 4) and a subset of neuronal populations (clusters 11, 20) express *cdx* shortly after injury (Fig. 1d), while two other populations (clusters 0, epithelium; cluster 9, *gcm+* neurons) started to express *dlx* (Supplementary Fig. 3a). Further comparisons between homeostatic and post-injury gut (clusters 4 and 16, respectively) also revealed the re-expression of *foxA*, a marker for the posterior gut, upon regeneration (Fig. 1e). These findings corroborate previous models on gut

198 posteriorization during regeneration [33] and further support the notion that selected wound-adjacent
199 tissues quickly re-establish posterior identities.
200 As these data supported our approach to reconstructing temporal dynamics, we next investigated
201 the expression of stem cell related genes after posterior injury. We reasoned that if stem cells are,
202 at least in part, regenerated by dedifferentiation of wound-adjacent cells, we should detect cell
203 clusters that are already present at 0 hpa, but start to express stem cell and proliferation-related
204 markers after injury. To assess this point, we first investigated the expression of the homeobox
205 gene *hox3*, whose transcripts are rapidly upregulated in posterior regeneration of *Platynereis*
206 *dumerilii* [34]. Prior work has shown *hox3* to be mostly restricted to an ectodermal population of PSCs
207 [13,22]. Whereas homeostatic trunk cells (0 hpa) are almost entirely devoid of *hox3* expression, we
208 could detect a strong and mostly cluster-specific upregulation of this gene in post-injury time points
209 of cluster 0 (Fig. 1f). Likewise, we find that this cluster expresses *Platynereis piwi* (Fig. 1g), a key
210 member of the GMP [35], and *myc* (Fig. 1h), both of which are expressed in *Platynereis* PSCs [22].
211 These data suggest that cluster 0 is a source of ectodermal PSCs.
212 As outlined above, *hox3* is preferentially expressed in ectodermal PSCs. However, additional
213 populations of stem cells contributing to *Platynereis* growth and regeneration have previously been
214 hypothesized, including mesodermal PSCs [13,36]. We therefore systematically queried our single-
215 cell atlas with a combined signature of stem cell and proliferation markers (such as *vasa*, *nanos*,
216 and *proliferating cell nuclear antigen/pcna*). This analysis revealed the combined signature of these
217 genes expressed in hotspots within several additional clusters, hinting at sources of
218 dedifferentiated PSCs in additional populations of wound-adjacent cells after injury (Fig. 1f-h,
219 Supplementary Fig. 3b-d).
220 *Platynereis* PSCs exhibit larger nuclei and nucleoli [22], a feature usually associated with increased
221 transcriptional (and translational) activity [37]. Increased, broad transcription, referred to as
222 "hypertranscription", is frequently observed in active stem cells and progenitors, closely associated
223 with proliferation, and plays a role in stem cell activation and function during growth and
224 regeneration. Recently, absolute scaling of single cell transcriptomes using Unique Molecular
225 Identifier (UMI)-based sequencing data has been shown to identify hypertranscriptomic stem cells
226 and progenitors [38]. We therefore investigated the dynamic changes in transcriptional activity upon
227 injury in our dataset and found evidence for hypertranscription (increased numbers of total UMIs
228 detected per cell (Fig. 1i). Our analysis shows that there is a progressive increase in high-UMI cells
229 over the regeneration process (Supplementary Fig. 3e). Hypertranscriptomic cells are located
230 within hotspots of *myc* and *chd1* gene expression on the UMAP (Fig. 1h, Supplementary Fig. 3f)
231 and can be found in several clusters, consistent with multiple, parallel de-differentiation events
232 occurring in distinct populations of wound-adjacent cells.
233 Taken together, our analysis supports the notion that upon injury, several distinct cell types
234 contribute stem cells to the regenerating blastema. These source cells must already be present in
235 regular body segments, independent of injury state or proximity to the posterior growth zone. To
236 further localize the source cells, we identified strongly expressed marker genes for the two major
237 clusters in question (cluster 0, cluster 1) and designed *in situ* HCR probes for them (Supplementary
238 Data 4). Co-labeling of both genes in posterior parts of uninjured, posteriorly growing animals
239 revealed that they demarcated two spatially distinct tissues, corresponding to the ectodermal
240 epidermis (cluster 0) and a sub-epidermal mesodermal cell type (cluster 1) (Fig. 2a-e). The latter
241 population covers the sub-epidermal region and extends to the gut epithelium of the animals, but
242 does not include muscle. We therefore refer to this population as coelomic mesoderm.

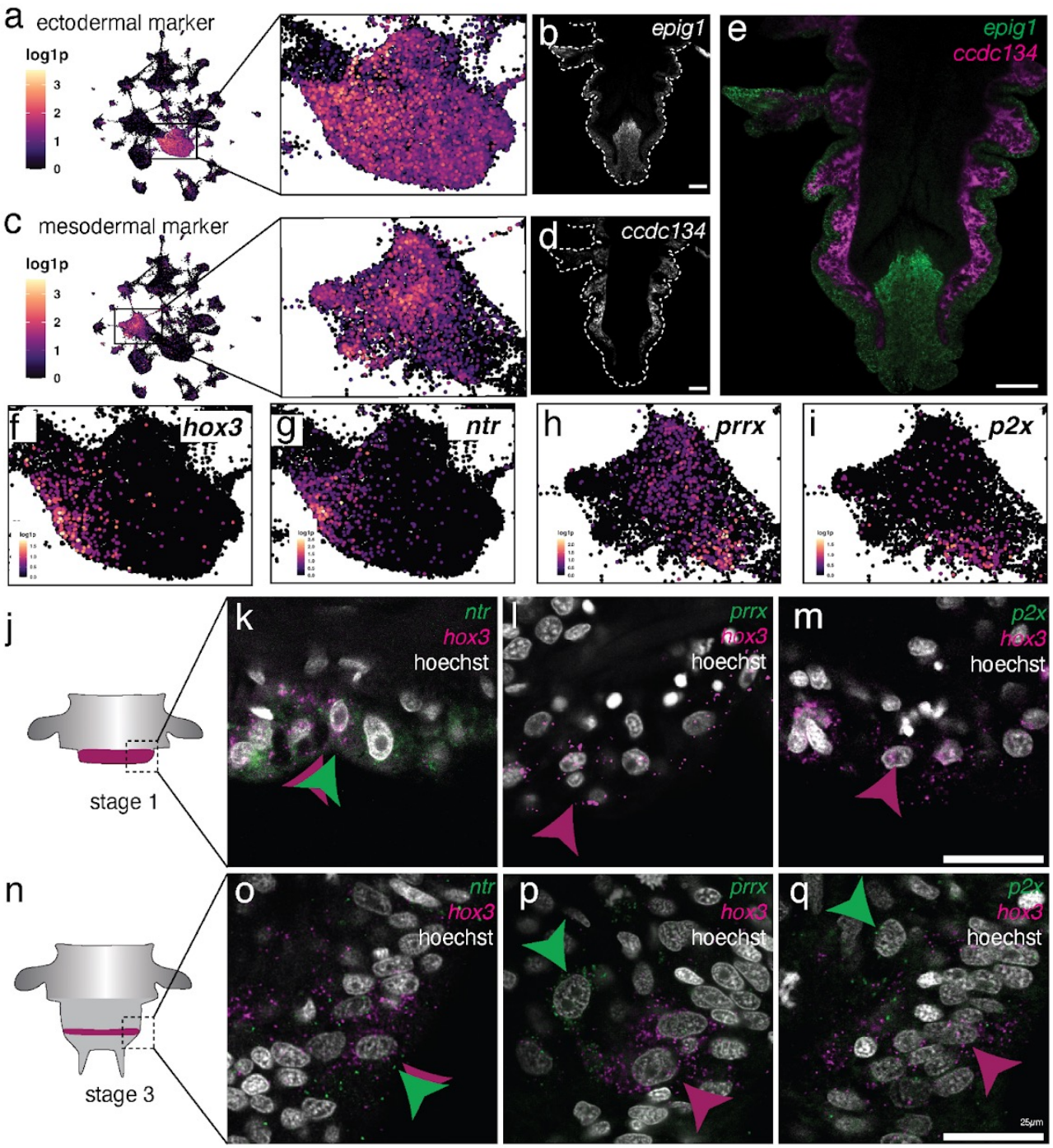


Figure 2. Distinct wound-adjacent cell populations acquire stem cell properties upon amputation.
(a,c) UMAP visualization of markers specific to cluster 0 (*epig1*) and 1 (*ccdc134*). (**b,d,e**) *in situ* HCR (uninjured animal, posterior end), showing mutually exclusive expression of *epig1* and *ccdc134* in the epidermis and coelomic mesoderm, respectively. **(f-j)** UMAP visualization of known (*hox3*) and novel (*ntr*) markers of the stem cell like subpopulation of epidermis (cluster 0), and of novel markers (*prrx*, *p2x*) for the corresponding subpopulation of coelomic mesoderm (cluster 1). **(j-q)** Overview scheme and *in situ* HCR of putative subpopulation-specific stem cell markers at stage 1 **(j-m)** and stage 3 **(n-q)** of posterior regeneration; arrowheads in the respective color indicate cells with enlarged nucleoli. Scale bars: 50 µm (e); 25 µm (m,q).

Ecto- and mesodermal PSCs exhibit shared and distinct molecular signatures

Having found evidence for distinct pools of dedifferentiated progenitors in our single cell data using candidate marker genes, we next aimed to identify the molecular properties of these populations in an unbiased way. We reasoned that if these PSC-like sub-populations are distinct from each other, we should be able to find genes enriched in either of them. As mentioned above, *hox3* has previously been described as a marker predominantly expressed in ectodermal PSCs. In line with this, *hox3* expression is mostly restricted to a subset of cluster 0 corresponding to the ectodermal group of GMP-expressing, high-UMI cells. Based on this observation, we sub-clustered cells of both the epidermal (cluster 0) and the coelomic mesodermal (cluster 1) populations to define their respective PSC-like sub-populations. We used UMIs and the expression of GMP, SAZ, proliferation and epigenetic remodeling-related genes to define the respective subclusters (Supplementary Fig. 4) and identified novel molecular markers unique to these cells (Fig. 2f-q; Supplementary Fig. 5a-t). These markers include genes encoding putative receptors, as well as proteins with DNA binding motifs such as transcription factors, thus establishing a set of molecules with possible regulatory functions (Supplementary Data 3).

In addition to the previously described genes *hox3* and *evenskipped (evx)*, our analysis predicts novel markers for ectodermal PSCs such as a *neurotrophin receptor* gene (*ntr,* Supplementary Fig. 5f-j) and a gene (*sp/btd*) encoding the *Platynereis* homolog of the transcription factor Sp9 [39] This population of cells further expresses early neuronal progenitor genes and patterning factors, such as the transcription factor gene *soxb1* (Supplementary Data 5) and the gene *four-jointed* that is involved in planar cell polarity, and was previously demonstrated to be expressed in developing medial neuroectoderm [40]. These data are consistent with the concept that these cells are the source of new neurons in post-regenerative growth. For the mesoderm-derived population of PSCs, our analysis also predicts distinct marker genes. These include the gene *chd3/4/5b* that encodes a chromodomain helicase DNA-binding protein, and has previously been detected in regenerating mesoderm [24] as well as a gene we identify as *Platynereis paired-related homeobox* gene (*prrx*) (Fig. 2h, Supplementary Data 5; Supplementary Fig. 5k-o). The putative purinoreceptor gene *p2x* (Fig. 2i) and the *Platynereis* orthologue of the mesoderm related homeobox factor *msx* [41,42] are also predicted to be expressed in mesoderm derived PSCs, albeit less exclusively than *prrx* (Supplementary Fig. 5k-t; Supplementary Data 3).

If these genes are novel markers of distinct ecto- and mesodermal stem cells respectively, they should be expressed in distinct injury-adjacent populations of cells and exhibit morphological properties of stem cells. To test this prediction, we designed specific *in situ* HCR probes (Supplementary Data 4) and used these to analyze the expression of both genes in posterior regenerates (Fig. 2j-q). In agreement with our digital data, *ntr* transcripts were co-expressed with *hox3* in cells of the epidermal layer, both at stage 1 (Fig. 2k) and stage 3 (Fig. 2o) samples. By contrast, the predicted mesodermal stem cell marker *prrx* labeled cells at a deeper layer. Consistent with the time-resolved atlas (Supplementary Fig. 5k-o), *prrx* was not yet detectable at stage 1 (Fig. 2l), but at stage 3 (Fig. 2p). Finally, *p2x* showed a similar expression pattern as *prrx* (Fig. 2m,q), but less restricted to the mesodermal layer. In all cases, a subset of labeled cells shows enlarged nucleoli as described for PSCs (arrowheads in Fig. 2k,o,p,q). Taken together, these results are consistent with the notion that, during regeneration, distinct populations of mesodermal and ectodermal PSCs derive from existing cells by dedifferentiation. Our single-cell atlas allows the identification of novel markers of these cells.

Clonal analysis by mosaic transgenesis reveals germ-layer based lineage restriction of posterior growth and regeneration

Whereas our data argued for PSCs of distinct ecto- and mesodermal origin in the blastema, it still remained unclear if these cells had identical potency, contributing to derivatives in all of the regenerate, or if they were more restricted in their developmental potential. We therefore turned towards a transgenic strategy that would allow us to address this question at least on the broad level of germ layers.

Several other lines of evidence from previous studies suggest the existence of lineage-restricted ectodermal and mesodermal stem cells during larval and juvenile posterior growth as well as embryogenesis in *Platynereis* [22,43–45]. The early embryogenesis of *Platynereis* follows a stereotypical programme known as spiral cleavage [46,47]. Highly asymmetric cell divisions (unequal cleavage) in the early embryo produce blastomeres of characteristic sizes and positions, whose fate are strictly determined [43]. Micro-injection of a fluorescent lineage-tracing dye in individual blastomeres at the earliest stages of the spiral cleavage process, shows that ectodermal, mesodermal, and endodermal trunk tissues of the 4-day, three segmented larva are produced, respectively, by micromere 2d, micromere 4d, and the macromeres 4A-4D of the early embryo [43]. In another study, individual cells were tracked via live imaging from early embryogenesis into early larval stages to identify the fates of the mesodermal 4d blastomere [45]. This work revealed that the mesodermal bands and the primordial germ cells are produced by asymmetric divisions of the 4d lineage. In addition, the final divisions of the lineage during embryogenesis forms a group of undifferentiated cells at the posterior end of the hatched larvae, which will possibly become the mesodermal PSCs in later stages. Due to the transient nature of the signal (mRNA or dye injections), tracking the fate of putative PSCs into later juvenile stages was not feasible. However, molecular profiling [22] suggests but does not demonstrate the existence of at least two pools of PSCs with specific signatures, ectodermal and mesodermal, organized as two concentric rings anterior to the pygidium, the terminal piece of the *Platynereis* trunk (Supplementary Data 6, part A). So far, no transgenic lineage tracing technique has been used to clarify the origin of tissues in the posteriorly growing or regenerating juvenile.

To address this gap, we devised a mosaic transgenesis strategy using previously-established Tol2 transgenesis methods [17]. We constructed a Tol2 transgenesis construct with a nuclear mCherry and a membrane EGFP [45], under the control of the ribosomal protein rps9 promoter for ubiquitous expression [17] (Fig. 3a). We injected several batches of zygotes at the one-cell stage with the donor plasmid containing rps9:H2A:mCherry:F2A:mGFP transgene and transposase mRNA. These F0 worms typically show mosaic integration of the transgene. We raised the F0 batches that showed high numbers of surviving juveniles (Supplementary Data 6) (Fig. 3b-j). To screen these individuals for fluorescence patterns and identify which germ layers had the transgene integration, we amputated juvenile worms when they reached 6 weeks. These original tails (pygidium + a few growing segments) were imaged via confocal microscopy from both the dorsal and ventral sides. The amputated worms were further raised in individual containers and allowed to regenerate their posterior parts for three weeks. They were then amputated again to collect the regenerated posterior parts for imaging, taking care of cutting one segment anterior to the regenerated part. The whole cycle was repeated for a third amputation for imaging. For each transgenic individual, we thus collected pictures of the primary clones derived from transgenic blastomeres as a result of normal development, as well as pictures of two reiterative regenerations from the same primary clones originating from the non-regenerated trunk (Supplementary Data 6, part B).

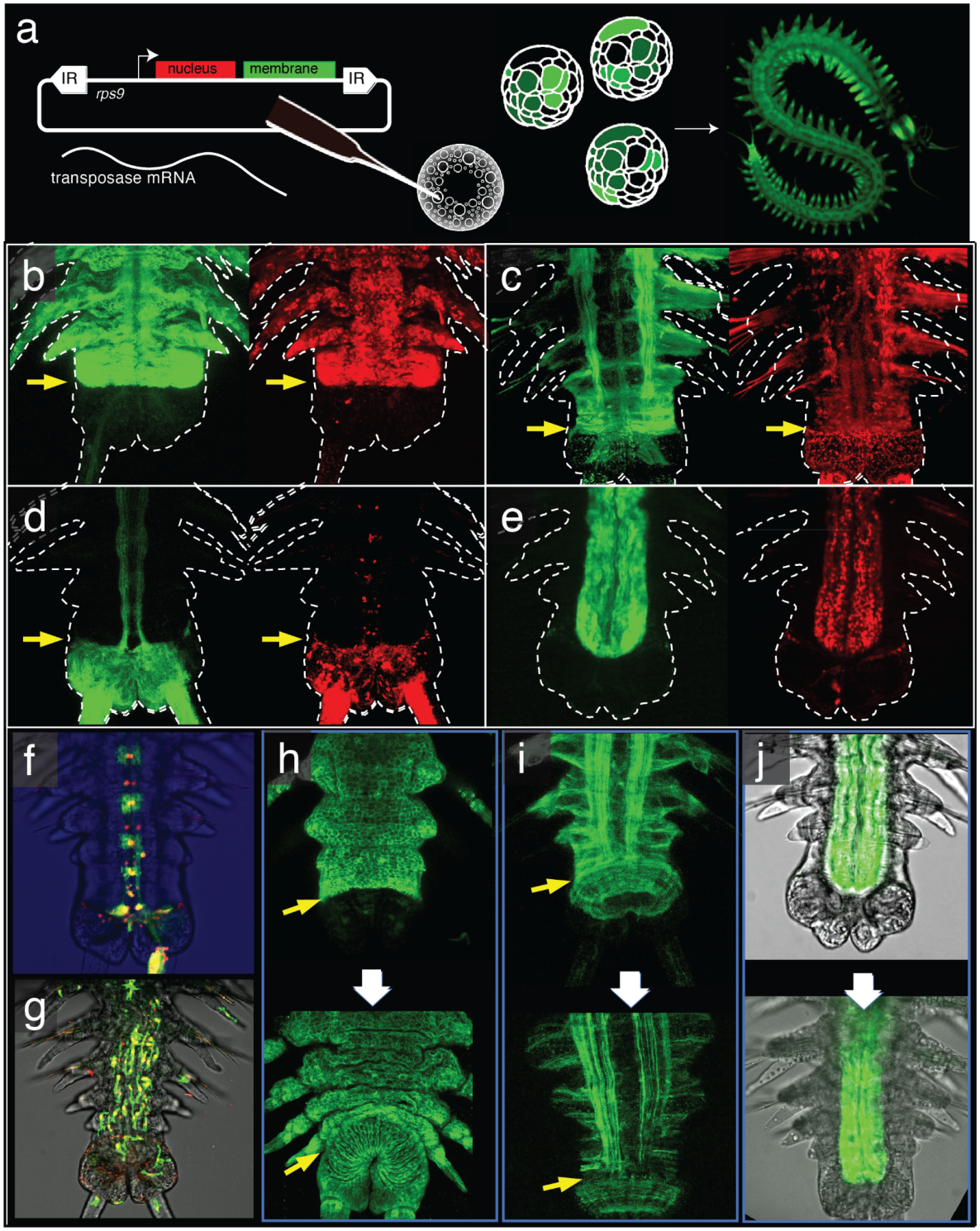


Figure 3. Mosaic transgenesis reveals developmental compartments and germ-layer restrictions of *Platynereis* regeneration. (a) Summary sketch of the protocol for creating large embryonic clones with transposase. (b-g) Examples of simple germ layer– or tissue-specific primary clones observed on posteriorly growing worms; 6 weeks, ventral views. (b) Ectodermal clone, no pygidial or internal cell labeled. (c) Mesodermal clone, only the muscles are clearly visible (d) Pygidial and median neural lineage clones. Median neurites are emanating from pygidial sensory neurons. (e) Endodermal clone. (f) Median neural lineage. A small part of the pygidium is independently labeled (g) Amoeboid cell lineage. (h-j) Regeneration experiments on animals bearing simple clones. Dorsal views, with pre-amputation views on top and matching regenerates on the bottom. (h) Trunk ectoderm clone. Both regenerated trunk and pygidial ectoderm are labeled. (i) Mesoderm clone. (j) Endoderm clone. For all panels, green labeling is cell membranes, red labeling are cell nuclei. Yellow arrow: position of the PSC rings.

354 Overall, we found that most individuals showed complex patterns of fluorescent primary clones
355 due to multiple reasons. Firstly, multiple blastomeres were transformed (Supplementary Data 6,
356 part B, N09 and N35 for examples) resulting in two or even three germ layers labeled. Secondly,
357 only a part of a germ layer may be labeled. This is most evident in cases where only a bilateral half
358 of the tissues is fluorescent because transgenesis happened in only one of the bilateral
359 descendants of the germ layer founding blastomere (e.g. 4d divides bilaterally to give the
360 precursors of the right and left mesoderm, Supplementary Data 6, part B, M24 and N23). Thirdly,
361 some tissues were labeled in a stochastic, salt-and-pepper manner. This phenomenon is known
362 as variegation [48] and presumably happens when a transgene is inserted near or within a
363 heterochromatic region that imposes unstable transcriptional repression on it. This was particularly
364 recurrent at the level of ectodermal tissues (Supplementary Data 6, part B, N25 and N33 for
365 examples).
366 Despite this complexity, simpler patterns were also recovered in several individuals corresponding
367 to the labeling of the whole trunk ectoderm (Fig. 3b), the whole trunk mesoderm (Fig. 3c) and the
368 entire gut endoderm (Fig. 3e). These primary clones support the aforementioned concept that
369 separate pools of precursor cells generate these germ layers during the life-long process of
370 posterior addition of segments. As for the endoderm, no endodermal PSCs have been identified
371 by molecular signature, and it is possible that endodermal precursors or stem cells are spread in
372 a diffuse way along the length of the trunk.
373 In addition to the trunk germ layers, several primary clonal patterns were obtained repeatedly either
374 alone (Fig. 3d, f, g) or in combination with others (Supplementary Data 6). The pygidial ectoderm
375 was often labeled independently of the trunk ectoderm (Fig. 3d). This demonstrates that the
376 pygidial ectoderm is derived from blastomeres different from the trunk ectoderm and that the
377 anterior border of the pygidial ectoderm with the trunk ectoderm is a compartment border with no
378 contribution of the pygidial cells to the growth of the trunk ectoderm (Fig. 3b, d). The one exception
379 to the pygidium/trunk compartmentalization is the presence of a median neural lineage (Fig. 3d),
380 composed of two pairs of cells per new segment, that is also identified alone (Fig. 3f), sometimes
381 unilaterally (Supplementary Data 6, part B, M09, M10, N04, N11, N14, N26, N40 and O15). These
382 cells are probably produced by independent median specialized posterior stem cells that segregate
383 from the 2d lineage in the early embryo. Lastly, a lineage of amoeboid, presumably phagocytic
384 cells (Fig. 3g), possibly derived from anterior embryonic mesoderm, was observed several times.
385 Most importantly, germ-layer compartmentalization is fully conserved during regenerative events
386 (Fig. 3h-j), with each germ layer of the regenerate originating exclusively from cells of its kind in
387 the neighboring non-regenerated trunk. This is in agreement with the distinct source populations
388 of stem cells suggested by our scRNAseq analyses. A diagram of the whole set of primary clones
389 obtained fully supports this interpretation of compartmentalization (Supplementary Data 6, part C).
390 Pygidial ectoderm, entirely removed upon amputation, is regenerated exclusively from trunk
391 ectoderm precursors (Fig. 3h).
392 While these transgenic clones do not demonstrate the embryonic germ layer origins, they show
393 that germ layers remain strictly compartmentalized during posterior segment addition, similar to
394 embryonic/larval development. Taken together, all these results are compatible with the presence
395 of the two rings of ectodermal and mesodermal PSCs immediately anterior to the pygidium/trunk
396 border, while the unsegmented endoderm may grow diffusely or through the activity of specific
397 endodermal PSCs yet to be identified. After amputation (which removes all PSCs), ectodermal and
398 mesodermal PSCs, as well as endoderm, are regenerated exclusively from precursors of their kind
399 in the uncut segments, either from dedifferentiating cells, or from unknown resident lineage
400 restricted precursor cells, in complete agreement with the single-cell transcriptomics clustering.

TOR signaling is required for successful activation of PSCs and regeneration

The protein kinase Target of Rapamycin (TOR) has been implied in wound response and blastemal signaling in planarians, zebrafish and axolotl [49–53], reviewed in [54]. Moreover, increased TOR signaling has also been observed in hypertranscriptomic cells [38], as we observe them in the *Platynereis* regeneration process. We therefore investigated whether posterior regeneration in *Platynereis dumerilii* also required a functional TOR signaling system.

Our bioinformatic analysis identified a gene encoding a *Platynereis* TOR ortholog (Supplementary Data 5). In our single-cell atlas, we found this gene to be broadly expressed, including in and around the tentative PSC subpopulations (Fig. 4a-c). To assess whether or not TOR signaling was required for regenerating PSCs after amputation, we treated amputated animals with the ATP competitive TOR inhibitor AZD8055 and compared their regenerative success to DMSO-treated controls (Fig. 4d-l). Already at 24 hpa, so before the formation of a significant blastema or any detectable increase in cell proliferation [13], *in situ* HCR revealed that *myc* expression was strongly reduced and *hox3* expression was completely undetectable in treated animals (Fig. 4e,f), whereas control animals successfully established a zone of *myc* and *hox3*+ cells (Fig. 4i,j). Additionally, treated animals did not reach a stage 3 regenerate at 72 hpa (Fig. 4g,h). Carrier (DMSO) control animals progressed normally and regenerated a blastema and early developing anal cirri within the same timespan (Fig. 4k,l). We therefore conclude that, upon TOR inhibition, *Platynereis* fails to regenerate PSCs, and subsequently does not develop a blastema or differentiated posterior tissues.

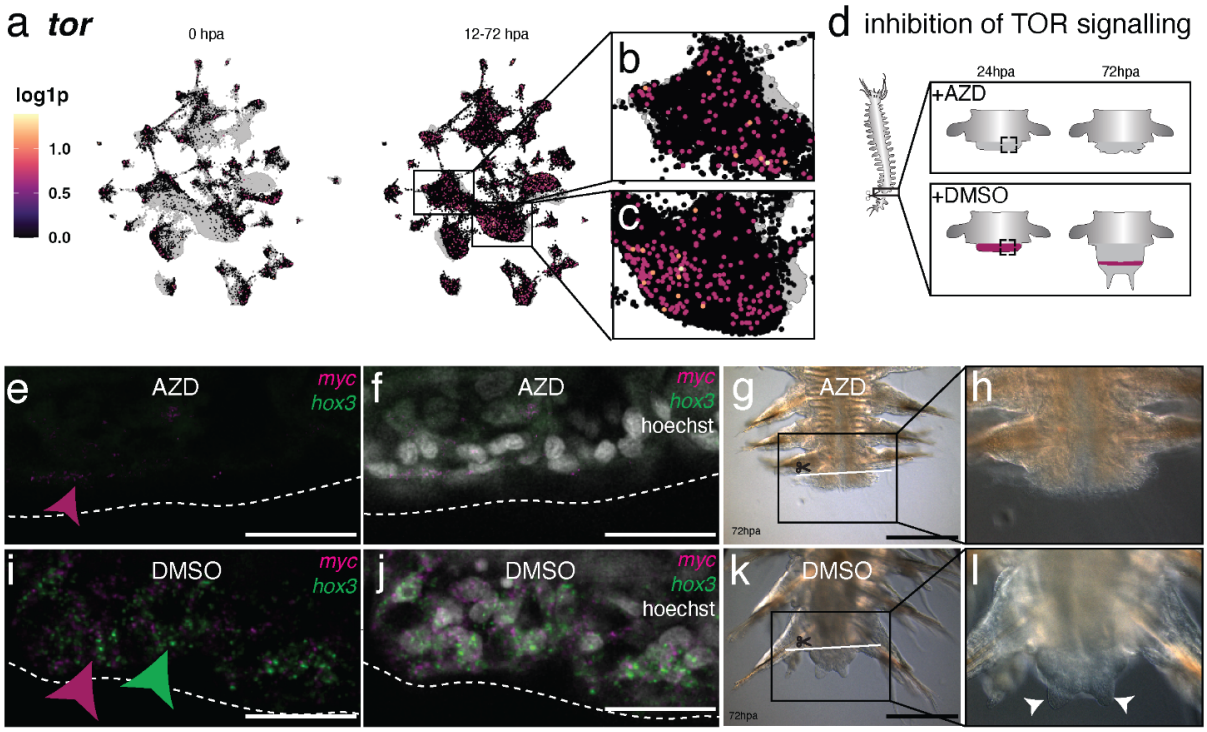


Figure 4. TOR signaling is necessary for re-establishing stem cell gene expression profiles and morphological regeneration upon injury.
(a-c) UMAP visualization of *tor* expression; (**a**) comparison between 0 hpa and 12-72 hpa; (**b,c**) enlarged views of cluster 0 and cluster 1 in 12-72 hpa samples. **(d)** Scheme of posterior amputation, TOR inhibition and posterior regeneration after 24 and 72 hours, highlighting region used for assessing stem cell gene expression 24 hpa. **(e-h)** Analysis of amputated animals (n=6) treated with AZD8055 TOR inhibitor. (**e,f**) Confocal images of *in situ* HCR stainings detecting expression of *hox3* and *myc* at 24 hpa; (**g,h**) brightfield images at 72 hpa; **(i-l)**. Equivalent analyses in DMSO-treated controls. (**i,j**) Confocal images of *in situ* HCR stainings for *hox3* and *myc* at 24 hpa; (**k,l**) brightfield images of posterior regenerates at 72 hpa. Scale bars: 25 µm (**e,f,i,j**); 250 µm (**g,h,k,l**).

Discussion

As outlined above, our work is consistent with the classical proposal that formation of the regenerative blastema of *Platynereis dumerilii* involves a process of injury-induced "re-embryonalisation" of wound-adjacent cells [14]. We advance this model to cell-type resolution, find molecular similarities to vertebrate blastema formation and present a clonal analysis furthering our understanding of lineage restriction during this process. Our data are consistent with both morphallactic repatterning processes and multiple parallel dedifferentiation events, and provide the molecular fingerprint of distinct groups of stem cells in posterior regeneration. These findings complement observations based on candidate genes and transcriptomic analyses in *Platynereis* and other annelid systems [13,55], and offer a fresh perspective on fundamental regenerative processes.

Both our transcriptomic and clonal analyses argue that, unlike planarians that exhibit pluripotent stem cells capable of regenerating all cell types of the adult [56], *Platynereis* regeneration relies on cells with limited potency that respect the distinction between different germ layers. In both of these spiralian species, however, there appears to be a continuity of potency mechanisms between normal growth and regeneration processes: In planarians, pluripotent neoblasts are not only relevant for regeneration, but also homeostasis and growth [57]. Likewise, in *Platynereis*, lineage restriction applies to transverse growth, posterior growth, and regeneration. This dissimilarity within the clade of spiralians is reminiscent of the diversity of growth and regeneration mechanisms also found in cnidarians, where pluripotent stem cells are found in *Hydractinia* [58], whereas *Hydra* employs lineage-restricted progenitors [59]. As single-cell analyses in other annelid species are becoming available [55,60], we expect that the availability of time-resolved cellular and molecular data in the *Platynereis* model will help to more easily delineate differences and commonalities for regeneration-relevant stem cell mechanisms also in other spiralian species.

As the labeled clones we obtain by zygotic injections are primarily large, they are well suited to provide a clear view of germ-layer-restricted lineages, but cannot yield experimental access to smaller lineages in both posterior growth and regeneration. Our single-cell transcriptomic atlas corroborates multiple distinct clusters of progenitors that are confined to individual germ layers, but also exhibit further subdivisions. It is thus possible that there are even more restricted lineage compartments currently not accessible to experimental validation. In turn, not every transgenically labeled compartment contributes to its regenerated counterpart. One example is the bilateral neural lineage that we identify, which is continuous with the pygidium before amputation, but does not visibly contribute to the regenerate. The existence of such a developmental compartment is consistent with the idea that there might be distinct subsets of *hox3*-positive PSCs in the ring-like segment addition zone, possibly reflecting the existence of distinct neurogenic columns in posterior growth [61]. More refined mapping techniques will be required to assess if such subsets exist in regular development, and if they are reconstituted in the process of regenerative growth.

Our data not only provide insight into the likely source and restriction of PSCs in *Platynereis* regeneration, but also into molecular factors involved in the emergence of blastemal cells and their possible conservation. Based on our findings and existing literature, we propose a working model of regeneration in *Platynereis* consisting of the activation and dedifferentiation of wound-adjacent cells, their acquisition of distinct, lineage restricted stem cell properties and ultimately their proliferation and contribution to regenerating tissue. While dedifferentiation as a source of blastemal stem cells in annelids has previously been proposed [8,10,13,14,23,62], we report cell-based molecular data consistent with this concept. We observed multiple distinct populations of cells responding in similar ways: they start expressing *myc* which, besides its general role in stem cell activity, has been shown to be involved in mammalian dedifferentiation and pluripotency [63,64], as well as several genes of the germline multipotency programme (e.g. *piwi*, *vasa*, *nanos*), which are also expressed in *Platynereis* PSCs during regular growth. These cells show hypertranscription, a

481 known feature of active adult stem cells [37,38], strong expression of genes involved in epigenetic
482 reprogramming (e.g. *chd3/4/5b*, *dnmt1*, *chd1*), and start to proliferate. Both *myc* and *chd1* have
483 previously been shown to play a central role in stem cell activation and hypertranscription in many
484 species, such as the endothelial to hematopoietic stem and progenitor transition in mouse
485 development [37,65].
486 Together, these findings strengthen the argument for multiple, parallel dedifferentiation events
487 underlying regeneration in *Platynereis*. Our data do not exclude the possibility that resident stem
488 cells or dedicated progenitors among heterogeneous pools of differentiated cells could also
489 contribute to the regenerate, as has been proposed or observed in multiple regeneration models
490 including annelids [5,8,10,66]. However, the fast *de novo* establishment of the molecular stem cell
491 signature, in the absence of obvious proliferation, argues that such a contribution may only account
492 for a limited set of blastemal cells.
493 Apart from those more general processes involved in the regeneration programme, the notion that
494 PSCs resulting from epidermal and coelomic mesodermal cells are molecularly distinct will also be
495 helpful in more clearly delineating possible parallels to established, dedifferentiation-based
496 regeneration mechanisms in other systems. Indeed, the expression of the *sp9* homologue *sp/btd*
497 in epidermis-derived PSCs of *Platynereis* is reminiscent of the expression of *sp9* in the
498 dedifferentiating epidermis of axolotl blastemas [67]. In turn, similar to the axolotl *msx-2* gene that
499 becomes refined to the mesenchymal part of the blastema [68], we identify its *Platynereis* ortholog
500 *msx* to be present in the *bona fide* mesodermal PSC clusters. Likewise, our work revealed a
501 previously uncharacterized *Platynereis prrx1* orthologue that is expressed in the mesodermal
502 PSCs. Vertebrate *prrx1* genes are prominently expressed in mesenchymal cells during limb
503 development of chicken [69], and mouse [70], and have also been characterized as part of the
504 connective tissue progenitor cells in axolotl and frog blastemas [68,71,72]. It has previously been
505 argued that evolutionary understanding of blastema-based regeneration will require a more
506 detailed understanding of the underlying genetic circuitry [3]. Our work establishes *Platynereis* as a
507 promising candidate for such analyses and provides a first comprehensive set of data towards this
508 end.
509 The prevalent expression of ribosome- and cellular growth related genes in PSC-like cells during
510 regeneration warranted an investigation of the role of TOR signaling in the activation or
511 dedifferentiation or activity of stem cells in our model. As discussed above, TOR signaling has
512 previously been implied in regeneration in many species, with a recent article demonstrating its
513 crucial, regulatory role upstream of axolotl limb regeneration [49]. While the exact role this pathway
514 plays in this process is not well understood, its activity seems to generally be required for
515 regenerative stem cell proliferation. Interestingly, the formation of stem cells through
516 dedifferentiation occurs during a period of lower TOR activity and high autophagy (which TOR
517 usually inhibits) [54]. We found that impaired TOR signaling in *Platynereis* not only affects
518 proliferation and the morphological formation of a blastema, but directly blocks the otherwise
519 reliable establishment of PSCs as early as 24 hpa, indicating an early, central role of TOR kinase
520 activity in regulating the dedifferentiation of cells in response to injury.
521 Lastly, we expect our methodological advances presented in this manuscript to be of broader use
522 in the establishment of additional resources for comparative regeneration biology. Using a widely
523 applicable cell fixation method and parallel processing of all samples allowed us to generate a
524 merged, temporally resolved dataset without having to rely on computational batch effect removal
525 strategies. Consequently, we uncovered the hypertranscriptomic state of stem cells at later
526 regenerative timepoints, a feature which otherwise might have been lost during pre-processing.
527 Our experimental strategy should be compatible with a wide range of species. Additionally, our
528 clonal analysis approach offers a direct way to observe clonogenic lineages in development and
529 regeneration as well as new insights into *Platynereis* lineage restriction, without requiring stable
530 expression of transgenic constructs beyond F0. Taken together, our manuscript therefore provides

a framework and methodological toolkit for future projects aimed at acquiring and analyzing the data required to compare regeneration across species and ultimately advance our understanding of blastema-based regeneration.

Methods

Animal culture

Platynereis dumerilii were kept in laboratory culture at temperatures between 18° and 20°C in a 16:8-hour light-5 dark (LD) cycle, and maintained under previously published conditions [19,73], adhering to the applicable national legislation.

Posterior amputation surgery

To perform posterior trunk amputation surgery, animals were anesthetized in 7.5% $MgCl_2$ diluted 1:1 with sea water. For bulk- and single cell transcriptomics and for all *in situ* HCR labelings shown, animals of 40-60 segments size and 3-6 months age showing no signs of sexual maturation and no prior injuries were sampled. Amputations were done by performing a transverse section posterior of segment 30, using surgical scalpels (Swann-Morton, Type 11). For regeneration time courses, animals were then rinsed in sea water and transferred to fresh culture boxes for recovery.

Bulk RNA sequencing of regenerating tissue

Posterior surgeries were performed as described above. Tissue was then harvested at each sampling time point by anesthetizing animals again as described above, followed by a transverse cut anterior of the wound-adjacent segment, right at the segment boundary. Tissue pieces were transferred to sea water, pooling 8 pieces per replicate, on ice. After tissue pieces sank to the bottom of the reaction tube, the supernatant was removed and the samples were frozen in liquid nitrogen and stored at -80° C.

RNA extractions were performed using a commercial kit (Direct-zol RNA MiniPrep, Zymo Research, USA), following the manufacturer's guidelines. RNA was eluted in 30 µl RNAse-free H2O. Illumina sequencing libraries were prepared at the Vienna Biocenter Next Generation Sequencing Facility, using NebNext Dual adaptors with dual indexing. Libraries were then sequenced using the Illumina NovaSeq platform with S1 PE100 flowcells.

Bulk RNA sequencing analysis

Sequences were trimmed using cutadapt (v. 1.12) [74] and analyzed for sequence quality using fastQC (v. 0.11.9) [75] and multiQC (v. 1.14) [76]. We used STAR aligner (v. 2.7.10b) [77] to generate a reference file from the *Platynereis* genome assembly draft (*genomeGenerate*) and align sequencing reads (*alignReads*), then extracted a counts matrix using featurecounts (v. 2.0.1) [78]. Data were then processed using DESeq2 (v. 1.36.0) [79] and mfuzz (v. 2.56.0) [80], estimating ideal clustering variable and cluster number following the software package guidelines. Cluster members (membership cutoff = 0.5) were then plotted using ggplot2 (v. 3.4.2) [81], with select genes of interest plotted in color (supplementary Fig. 1).

Single cell transcriptome sequencing

For single-cell RNA sequencing, we performed posterior amputations and regeneration time courses similar to those described above. However, after sampling the wound-adjacent segments, we transferred those to an acetic acid/ methanol (ACME) solution with 1% NAC for dissociation [26]. Cells were then dissociated over the course of 45 minutes, interspersed with rigorous pipetting every 15 minutes using a P1000 pipette pre-coated in 1% BSA in PBS to reduce cell loss due to stickiness. Dissociated cells were centrifuged (1000 g for 5 min at 4°C), resuspended in storage buffer (1% BSA in PBS, with 10% DMSO and 1U/µl recombinant RNAse inhibitor, Takara Bio) and stored at -20° C for further processing.

To remove debris and concentrate cells, we performed FACS on freshly thawed samples. We first labeled nuclei (Hoechst 33342 at 5 µg/ml, for 15 minutes at room temperature), then sorted 15,000 cells directly into 10X genomics chromium buffer, using FACS gates (FSC-A vs FSC-H; DAPI-A) to exclude debris and clumped cells (BD FACS Aria IIIu).

Single cell barcoding droplet (GEM) production and library preparation was performed at the Vienna Biocenter Next Generation Sequencing facility using the 10X genomics Chromium platform according to manufacturer's instructions (10X 3' v3 chemistry, 10X Dual Index Kit). Two

584 independent experiments were performed (replicates a and b), following the same strategy: all
585 animals for one experiment were sampled from sibling- or closely related batches of animals.
586 Amputations were performed in parallel, and by freezing ACME-dissociated cells, we were able to
587 store sampled cells throughout the regeneration time course. All samples were then sorted, and
588 barcoded in parallel. Libraries were sequenced on the Illumina NovaSeq platform (S4 lanes,
589 PE150).
590 Single cell data processing
591 We used Cellranger software by 10X genomics (v. 7.0.1) to generate a custom genome reference
592 (*mkref* function) from the *Platynereis* genome assembly v021 (https://4dx.embl.de/platy), provided
593 by Kevin Nzumbi Mutemi and Detlev Arendt (EMBL, Heidelberg). We then assigned reads to
594 cellular barcodes, aligned them to the custom reference and generated a read barcode matrix for
595 each sample individually using the *cellranger count* function with *expect-cells* set to 10000.
596 Our main processing pipeline was based on the "Seurat" package for R [82], with specific
597 modifications for merging multiple datasets. In brief, we imported the barcode matrices (*Read10X*
598 with min.cells = 3 and min.features = 200) and followed standard pre-processing steps. Outlier
599 cells were removed based on manual inspection of scatterplots (counts vs features), removing
600 between 30 and 108 cells per dataset. Quality metrics and cutoff values for all samples are
601 available in (Supplementary Data 1).
602 We merged all datasets following an approach employed elsewhere [83]. We Normalized the data
603 (LogNormalize with scale.factor = 10000) and identified the variable features for each dataset
604 individually (using FindVariableFeatures). We then used the union of all variable features from all
605 different timepoints to conserve features which might be variable only at a certain timepoint, and
606 used this set of features for the merged dataset. Next we scaled the merged dataset and performed
607 dimensionality reduction (first 50 dimensions in PCA space). We then calculated the UMAP
608 embedding and performed cell-type clustering. After trying multiple clustering resolutions, we
609 settled on a resolution of 0.5 as it best reflected known biological cell types and subpopulations.
610 Single cell cluster annotation
611 Marker genes were identified for each cluster using the Seurat *FindAllMarkers* function (only
612 returning positive markers at a min.pct of 0.25 and a log-fold change threshold of 1). As the
613 reference genome used in this manuscript lacks a gene name annotation, we annotated marker
614 genes here and below using a table of best BLAST hits for each coding sequence identified on the
615 genome (see "transcriptome annotation" below). The annotated table of markers for each cluster
616 (Supplementary Data 3), combined with screening expression of an array of genes with known
617 expression patterns in *Platynereis* (Supplementary Data 2) were then used to annotate the
618 clusters.
619 Analysis of UMIs per cell
620 UMI values of cells were compared between timepoints post amputation using the Wilcoxon rank
621 sum test with continuity correction (suppl. Fig. 3). To ensure these results were not affected by
622 outlier removal (see above) we performed the same test on cells without removing outliers, yielding
623 comparable results (0 hpa vs 24 hpa: $p < 2.2e-16$; 24 hpa vs 72 hpa: $p < 1.227e-11$).
624 Gene annotation
625 Top genes of every cluster were annotated using an automated pipeline. For this, all transcript
626 sequences predicted from a given gene locus (XLOC ID) were used for sequence searches using
627 BLASTX [84] against two protein databases: a version of the NCBI Uniprot/Swissprot repository
628 (accessed on June 16, 2021) and a database combining entries of the more inclusive NCBI RefSeq
629 repository (accessed on November 25, 2021). For the latter one, all protein sequences available
630 for a set of representative landmark species and taxa were used: Annelida, the mollusks *Mytilus*
631 *galloprovincialis, Crassostrea gigas, Mizuhopecten yessoensis, Octopus bimaculoides, Pomacea*
632 *canaliculata, Lymnaea stagnalis, Biomphalaria glabrata, Aplysia californica*; the insects *Drosophila*
633 *melanogaster, Apis mellifera, Clunio marinus, Nasonia vitripennis*; and the chordates

634 *Branchiostoma, Mus musculus, Oryzias latipes, Danio rerio and Gallus gallus*. Best hits for each
635 of these searches were tabulated. For gene loci with multiple transcripts, the results for the
636 transcript that retrieved the highest score in the analyses was retained as reference, so that each
637 gene locus retrieved one annotation.
638 For investigation of specific genes, we assembled *bona fide* full length sequences independently
639 of the genome annotation, using available RNAseq data generated for the laboratory strains (PIN,
640 VIO) maintained in the laboratory [9]. Sequences from individual libraries were assembled using the
641 CLC Main Workbench Software package (version 23.0.2), and predicted protein sequences
642 subjected to domain analysis using SMART [85]. Assembled gene sequences for *Platynereis prrx,*
643 *ntr, p2x, epig1, ccdc134, tor* and *smg1* were submitted to the NCBI Genbank repository.
644 Phylogenetic analyses
645 For phylogenetic analyses of selected *Platynereis* proteins whose phylogeny was not yet
646 previously reported, we used the following strategy: We identified the top hits in selected reference
647 species representing key vertebrate phyla (mammals, birds, amphibians) as well as invertebrate
648 phyla (mollusks, insects) by performing BLASTP searches against the NCBI clustered nr database.
649 For identifying / completing matching sequences in the axolotl (*Ambystoma mexicanum*), we made
650 use of the Axolotl-omics web resource (https://www.axolotl-omics.org/) that allowed access to the
651 latest axolotl transcriptome assembly (AmexT_v47). Proteins were aligned using the CLC Main
652 Workbench Software package (version 23.0.2) and cleaned of short sequences. Subsequently,
653 alignments were exported and used for phylogenetic analyses using IQ-Tree 1.6.12 [86,87], allowing
654 for the choice of the most suitable substitution model. Ultrafast bootstrap analysis [88] with 1000
655 repetitions was used to assess confidence for individual branches. The most likely tree topology
656 was then visualized using the iTOL suite [89], available at https://itol.embl.de .
657 Sub-clustering populations with stem cell like properties
658 To identify markers specific to sub-populations of clusters with stem cell like properties, we subset
659 clusters 0 and 1 (epidermis and coelomic mesoderm) and re-processed these transcriptomes as
660 described above, but without identifying variable features for each sample individually. Newly
661 calculated clusters were then analyzed for their expression of GMP and SAZ related transcripts as
662 well as their UMIs per cell. Strongly positive subpopulations were identified for both clusters.
663 Marker genes for those subpopulations were then calculated against the entire dataset (suppl. Fig.
664 4, 5; full marker gene table in Supplementary Data 3).
665 Plotting and visualization
666 All plots displaying single cell data were generated using the R package SCpubr (v. 1.1.2) [90]. The
667 displayed values on gene expression UMAPs are log-transformed transcript counts normalized by
668 UMI.
669 in situ HCR
670 For *in situ* visualization of gene expression, we sampled tissues of either growing or regenerating
671 animals as described above. Labeling was done following our previously published protocol for
672 *Platynereis* [18], with probes designed using the publicly available algorithm HCR3.0 Probe Maker
673 (v. 2021_0.3.2, https://github.com/rwnull/insitu_probe_generator) described in [91]. All HCR probe
674 sequences are available (Supplementary Data 4).
675 Microscopy and image analysis
676 Confocal images of in situ HCR labelings were taken using a Zeiss LSM700 confocal microscope
677 with a Plan-Apochromat 40x/1.3 Oil DIC, WD 0.21 mm lense, using Zeiss Zen software.
678 Images were then processed using FIJI software [92] for adjustment of contrast, LUT selection and
679 creation of overlay images.
680 Molecular Cloning
681 The transgene rps9::H2A::mCherry::F2A::GFP::CAAX (simplified as pHCX) was engineered using
682 the Gibson assembly protocol [93]. The donor plasmids for this construction were pEXPTol2-EGFP-
683 CAAX, pEXPTol2-H2A-mCherry [45] and pTol2{rp-s9::egfp} [17]. The ribosome skip site coding

684 sequence F2A was inserted between the two recombinant fluorescent protein coding sequences
685 by adding it to the cloning primer. For the Gibson reaction, Gibson assembly master mix (NEB,
686 France, E2611S) and NEB 5-alpha Competent E.coli (NEB C2987H) were used, following the
687 manufacturer's protocol.
688 Generation and analysis of transgenic animals
689 *Tol2* transposase mRNA was synthetized using the SP6 mMessage kit from Ambion (AM1340).
690 The manufacturer's protocol was followed until the end of DNase step (1ml DNase 15 mins at
691 37°C). For purification of mRNA, MEGAclear kit from Ambion (AM1908) was used, following RNA
692 elution option two with the following modifications: elution buffer was heated to 72°C, this warmed
693 elution buffer was applied to filter cartridge containing mRNA, the tubes were kept at 72°C heated
694 plate for 5 min before centrifuging for elution.
695 For micro-injections, previous protocols were used [45,94]. Briefly, fertilized eggs were dejellified at
696 45 mins post fertilization, using a 80 mm nylon mesh. The egg cuticle was softened with a flash
697 treatment with proteinase K diluted in seawater (100 mg/ml). After abundant rinsing with sea water,
698 the eggs were micro-injected with a mix of Tol2 transposase mRNA (300 ng/ml) and plasmid pHCX
699 (100 ng/ml). Injected eggs were incubated overnight at 28°C. Injected batches of larvae and
700 juveniles were raised in a common polypropylene container with 800 ml of sea water until they
701 reached 6 weeks. Worms were then relaxed and amputated as described before. Amputated
702 posterior parts were kept in 7.5% MgCl2 diluted 1:1 with sea water and mounted on slides using
703 three layers of tapes as spacer. Confocal images were acquired on a Zeiss LSM 780 confocal
704 scanning microscope. Amputated animals were dispatched in individual boxes and fed carefully to
705 avoid fouling of the small amount of sea water. The operation was repeated twice after three weeks
706 of regeneration. Some individuals however were not documented for two rounds of regeneration
707 because they underwent sexual maturation, that stops regeneration.
708 AZD8055 treatment
709 Animals were surgically amputated as described above. Regenerating animals were then kept in
710 glass beakers in artificial sea water, either treated with 10µM AZD8055 (MedChemExpress, USA)
711 or with an equal amount of carrier control (DMSO).

712

Acknowledgements

The authors are grateful to members of their respective research groups for continuous input into this project; Alison Cole (University of Vienna) and Simon Haendeler (CIBIV, University of Vienna) for advice on single-cell data processing; Kevin Nzumbi Mutemi and Detlev Arendt (EMBL, Heidelberg) for access to the *Platynereis dumerilii* genome sequence and primary annotations; Margaryta Borisova, Andrij Belokurov and Netsanet Getachew (University of Vienna) for animal supply; the Max Perutz Labs BioOptics Facility for access to advanced light microscopy and FACS; the ImagoSeine core facility of Institut Jacques Monod, a member of France BioImaging (ANR-10-INBS-04) and certified IBiSA platform and the Next Generation Sequencing services at the Vienna BioCenter Core Facilities for service and advice.

Funding

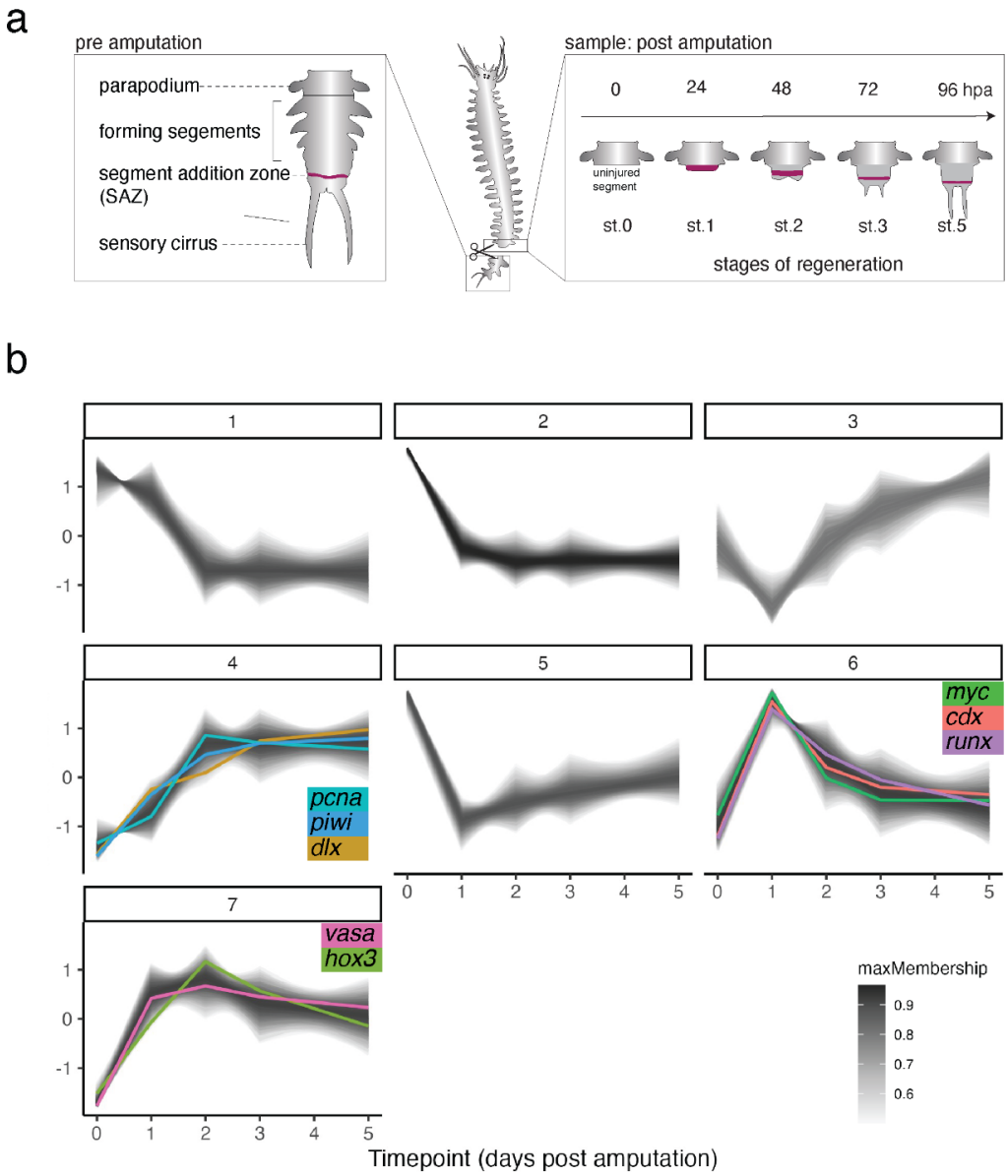
This research has been supported by a joint grant from the Agence Nationale de Recherche (ANR) ANR-16-CE91-0007 (G.B.) and the Austrian Science Funds (FWF) I2972 (F.R.), the FWF Special Research Project F78 (F.R.), the Fondation ARC pour la recherche sur le cancer (grant LSP 190375), the FWF doc.Funds PhD programme DOC 72 "Stem Cells, Tissues, Organoids – Dissecting Regulators of Potency and Pattern Formation (SCORPION)" (F.R.), the Austrian Academy of Sciences (ÖAW) DOC fellowship programme (A.W.S.), the German Academic Scholarship Foundation (L.A.), and the University of Vienna Research Platform "Single Cell Regulation of Stem Cells (SinCeReSt)" (F.R.).

Competing interests

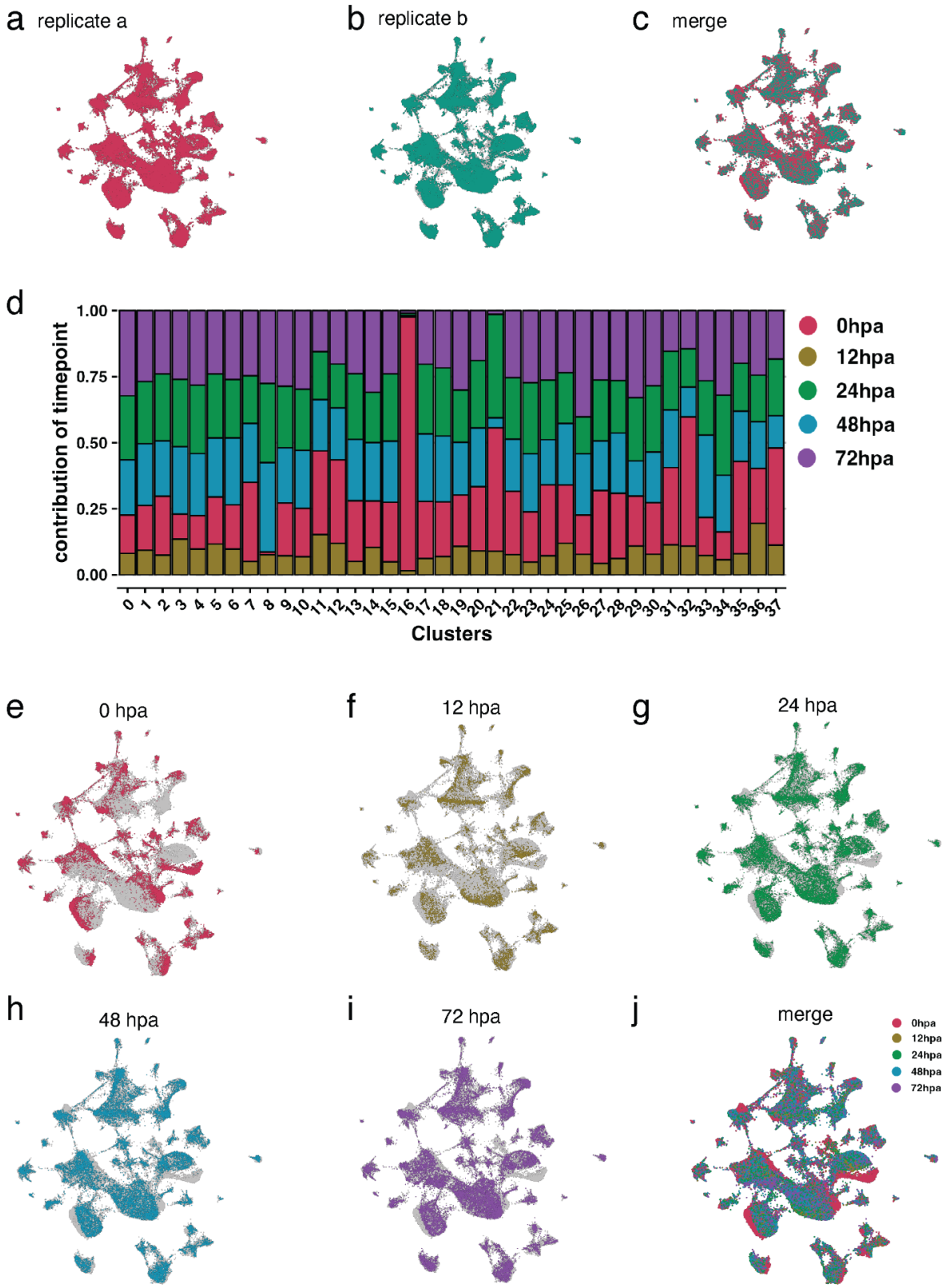
The authors declare no competing interests.

Supplementary Information

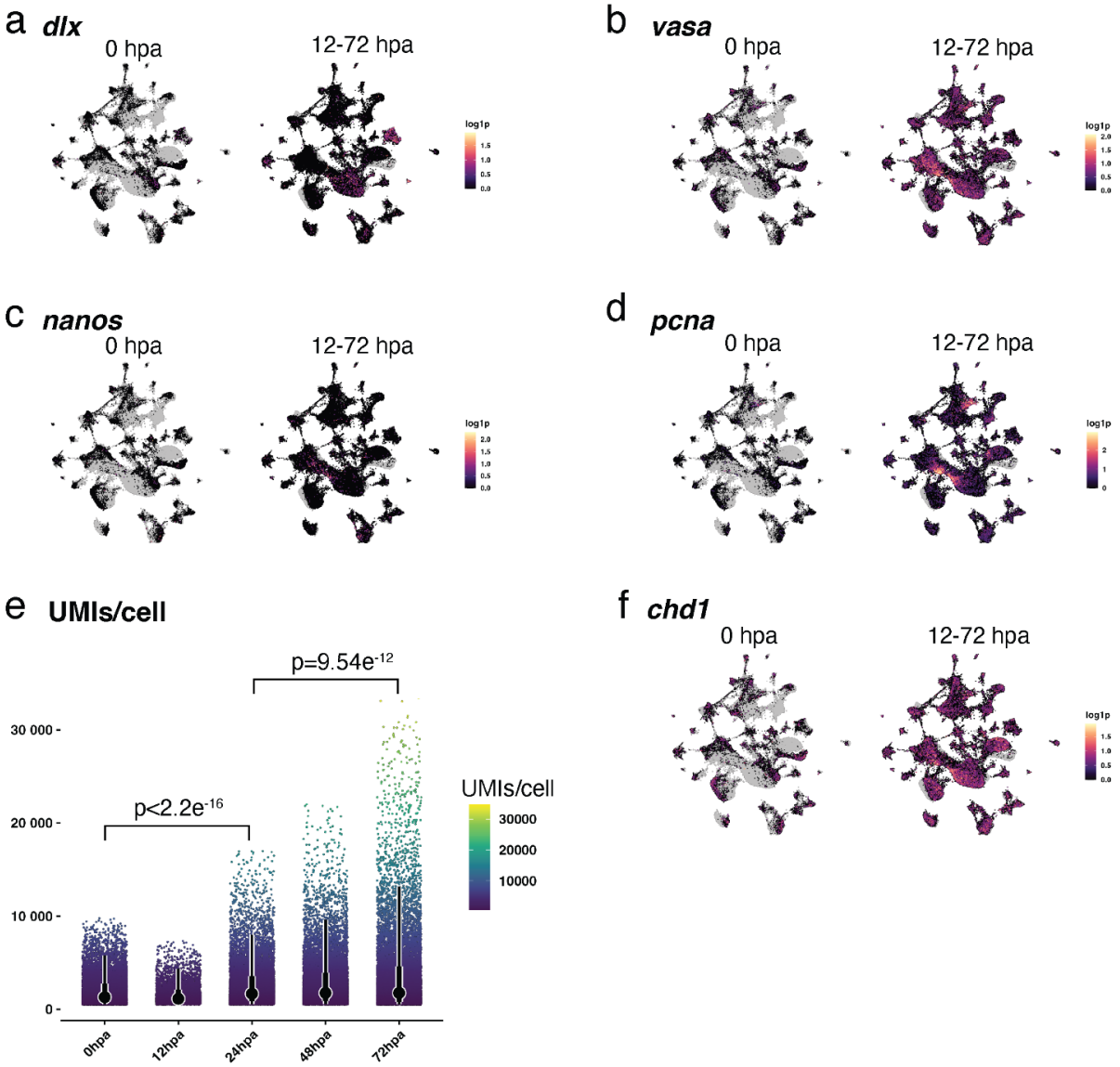
Supplementary Figures



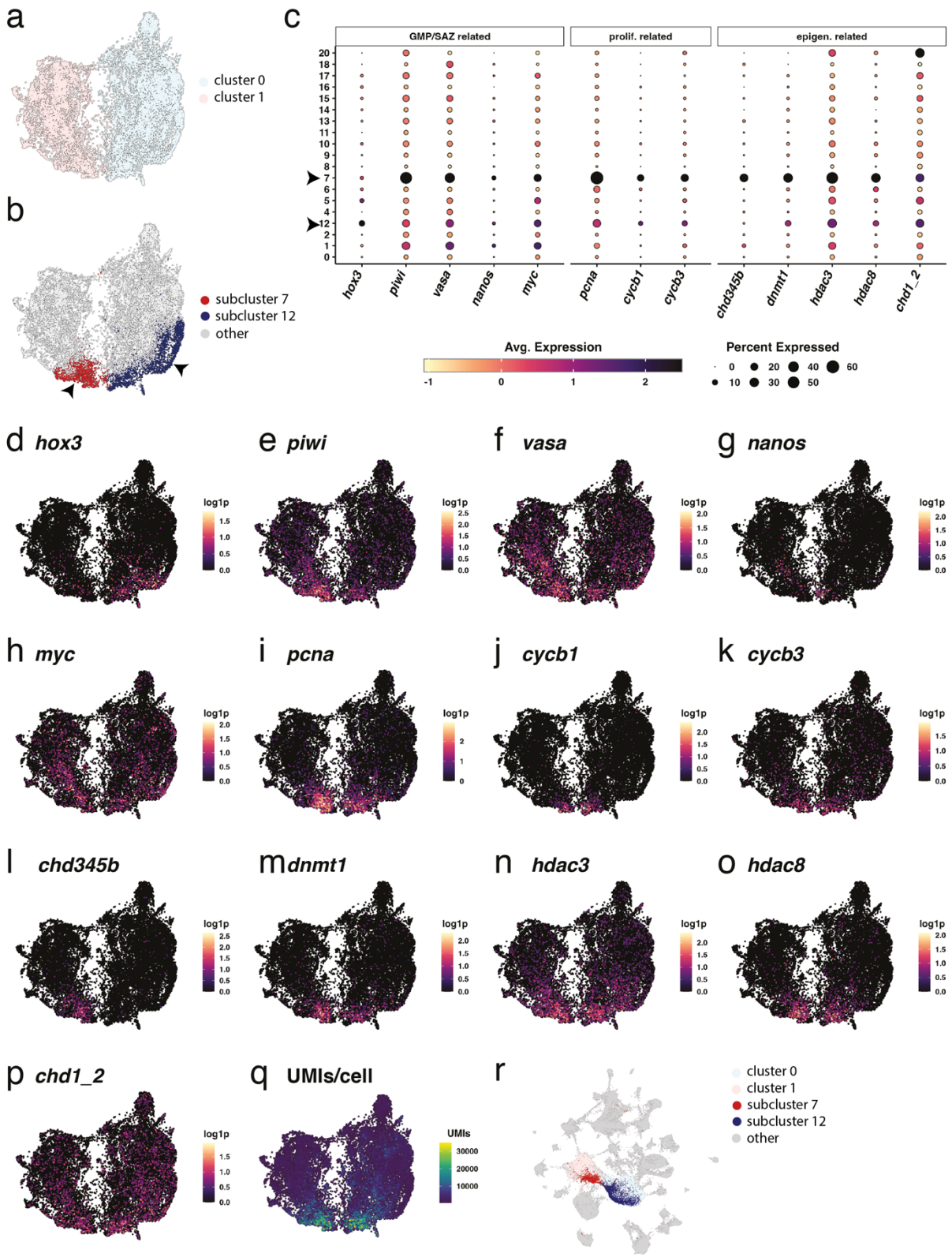
Supplementary Figure 1. Bulk RNA sequencing captures dynamics of gene expression during regeneration. **(a)** Sampling scheme illustrating posterior amputation and sampling timepoints 0 to 5, ranging from 0 hours post amputation (0 hpa, equivalent to a regular trunk segment) to 96 hpa; **(b)** distinct mfuzz clusters grouping major variants of gene expression dynamics over the sampled time points. Colored lines indicate the assignment of individual candidate genes (*pcna, piwi, cdx, myc, runx, hox3, vasa*) to these clusters. All transcripts with an mfuzz cluster membership > 0.5 were included (gray colors).



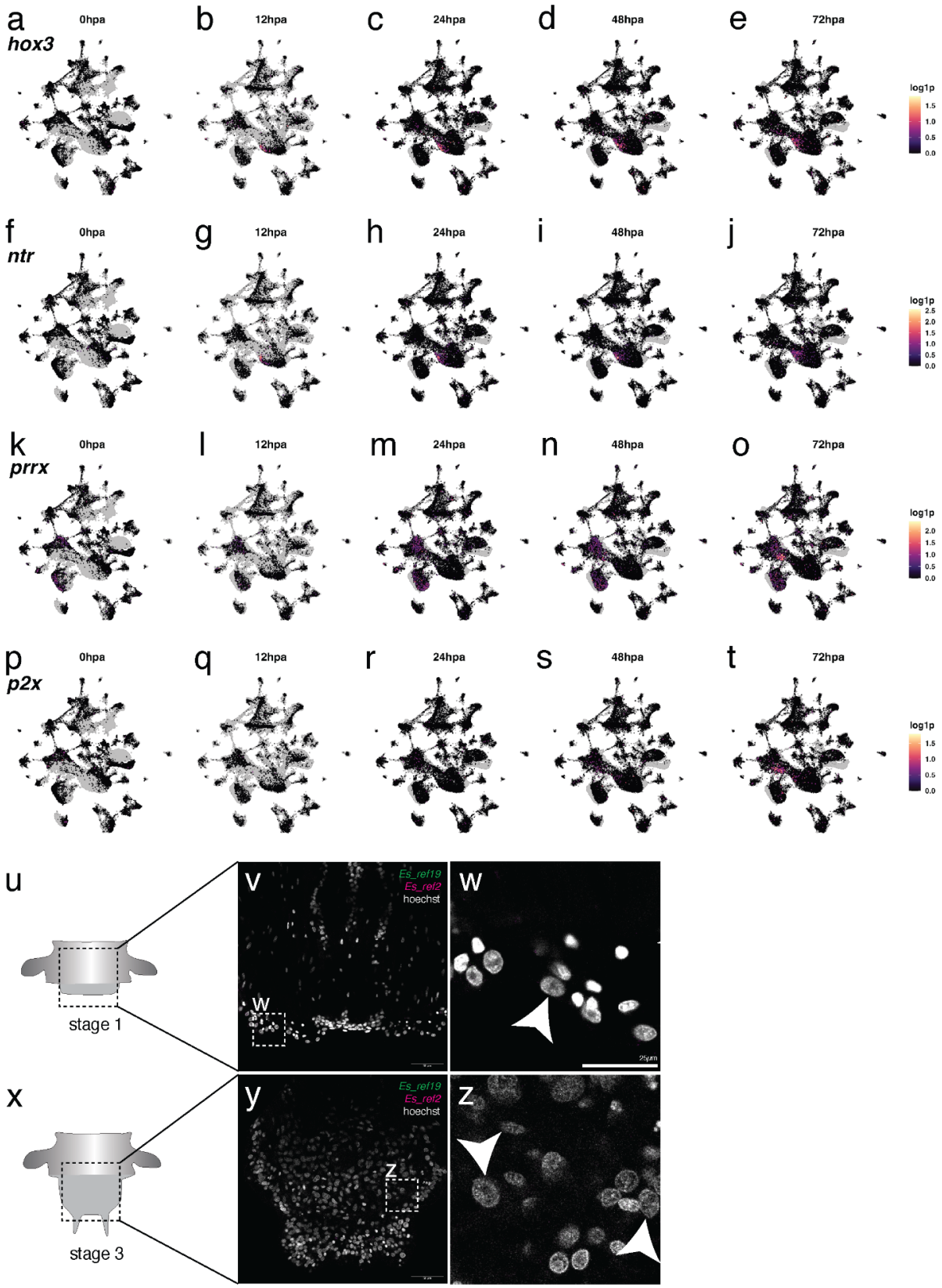
Supplementary Figure 2: Sample and replicate contribution to single cell UMAP and clusters. (**a-c**) UMAP visualization of independently replicated scRNAseq time series (merged timepoints 0 hpa - 72 hpa) both individually (**a-b**) and replicate a and b merged (**c**); (**d**) relative contribution of different sampling timepoints to the clusters of this dataset; (**e-j**) UMAP visualization of different sampling timepoints, shown individually (**e-i**) and merged (**j**).



Supplementary Figure 3: UMAP visualizations and UMI analysis of gene expression dynamics during regenerative timepoints. (**a-d**) gene expression UMAP visualizations of select PSC-related genes, splitting the dataset into 0 hpa (right after posterior amputation) and 12-72 hpa (post-injury timepoints). (**e**) geyser plot of UMIs per cell across sampled timepoints showing the progressive emergence of cells with significantly higher UMIs as regeneration proceeds (Wilcoxon rank sum test with continuity correction). (**f**) UMAP visualization of UMIs/cell, splitting the dataset the same way as above.



Supplementary figure 4. Subsetting clusters 0 and 1 to identify strongly PSC-related subclusters (**a-b**) UMAP visualizations of subset clusters 0 and 1, with the newly identified PSC-related subclusters highlighted (**b**). (**c**) dotplot representation of gene expression in subclusters of clusters 0 and 1, showing genes related to the GMP/SAZ, proliferation and epigenetic remodeling. (**d-p**) UMAP visualizations of all genes shown in the dotplot (**c**) individually. (**q**) UMAP visualization of UMIs per cell. (**r**) UMAP visualization of the entire dataset, highlighting clusters 0 and 1, and their newly identified respective PSC-like subclusters 12 and 7.



Supplementary figure 5: Synopsis of subcluster-enriched stem cell signatures.
(**a-t**) Time-resolved UMAP visualizations showing the expression of (**a-e**) hox3, (**f-j**) ntr, (**k-o**) *prrx* and (**p-t**) *p2x* over the course of regeneration. (**u,v**) imaging of samples labeled with control probes for *in situ* HCR analyses (Fig. 4).

References

1. Bely, A. E. & Nyberg, K. G. Evolution of animal regeneration: re-emergence of a field. *Trends Ecol. Evol.* 25, 161–170 (2010).
2. Lai, A. G. & Aboobaker, A. A. EvoRegen in animals: Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes. *Dev. Biol.* 433, 118–131 (2018).
3. Srivastava, M. Beyond Casual Resemblances: Rigorous Frameworks for Comparing Regeneration Across Species. *Annu. Rev. Cell Dev. Biol.* 37, 1–26 (2021).
4. Reddien, P. W. & Alvarado, A. S. FUNDAMENTALS OF PLANARIAN REGENERATION. Annu. Rev. Cell Dev. Biol. 20, 725–757 (2004).
5. Tanaka, E. M. & Reddien, P. W. The Cellular Basis for Animal Regeneration. *Dev. Cell* 21, 172–185 (2011).
6. Tanaka, E. M. The Molecular and Cellular Choreography of Appendage Regeneration. Cell 165, 1598–1608 (2016).
7. Tajer, B., Savage, A. M. & Whited, J. L. The salamander blastema within the broader context of metazoan regeneration. *Front. Cell Dev. Biol.* 11, 1206157 (2023).
8. Özpolat, B. D. & Bely, A. E. *Developmenta*l and molecular biology of annelid regeneration: a comparative review of recent studies. *Curr. Opin. Genet. Dev.* 40, 144–153 (2016).
9. Özpolat, B. D. *et al.* The Nereid on the rise: *Platynereis* as a model system. *EvoDevo* 12, 10 (2021).
10. Kostyuchenko, R. P. & Kozin, V. V. Comparative Aspects of Annelid Regeneration: Towards Understanding the Mechanisms of Regeneration. *Genes* 12, 1148 (2021).
11. Nikanorova, D. D., Kupriashova, E. E. & Kostyuchenko, R. P. Regeneration in Annelids: *Cell* Sources, Tissue Remodeling, and Differential Gene Expression. *Russ. J. Dev. Biol.* 51, 148–161 (2020).
12. Bely, A. E., Zattara, E. E. & Sikes, J. M. Regeneration in spiralians: evolutionary patterns and developmental processes. *Int. J. Dev. Biol.* 58, 623–634 (2015).
13. Planques, A., Malem, J., Parapar, J., Vervoort, M. & Gazave, E. Morphological, cellular and molecular characterization of posterior regeneration in the marine annelid *Platynereis dumerilii.* Dev. Biol. 445, 189–210 (2019).
14. Hofmann, D. K. Untersuchungen zur Regeneration des Hinterendes bei *Platynereis dumerilii. Zool Jb Abt Allg Zool Physiol* 374–430 (1966).
15. Zantke, J., Bannister, S., Rajan, V. B. V., Raible, F. & Tessmar-Raible, K. Genetic and Genomic Tools for the Marine Annelid *Platynereis dumerilii. Genetics* 197, 19–31 (2014).
16. Bannister, S. et al. TALENs Mediate Efficient and Heritable Mutation of Endogenous Genes in the Marine Annelid *Platynereis dumerilii.* Genetics 197, 77–89 (2014).
17. Backfisch, B., Kozin, V. V., Kirchmaier, S., Tessmar-Raible, K. & Raible, F. Tools for Gene-Regulatory Analyses in the Marine Annelid *Platynereis dumerilii. PLoS ONE* 9, e93076 (2014).
18. Ćorić, A. et al. A Fast And Versatile Method for Simultaneous HCR, Immunohistochemistry And Edu Labeling (SHInE). *Integr. Comp. Biol.* 63, 372–381 (2023).
19. Hauenschild, C. & Fischer, A. Platynereis dumerilii. Mikroskopische Anatomie, Fortpflanzung, Entwicklung. *Großes Zoologisches Praktikum* 10b, (1969).
20. Fierro-Constaín, L. et al. The Conservation of the Germline Multipotency Program, from Sponges to Vertebrates: A Stepping Stone to Understanding the Somatic and Germline Origins. *Genome Biol. Evol.* 9, evw289 (2017).
21. Rebscher, N., Zelada-González, F., Banisch, T. U., Raible, F. & Arendt, D. Vasa unveils a common origin of germ cells and of somatic stem cells from the posterior growth zone in the polychaete *Platynereis dumerilii.* Dev. Biol. 306, 599–611 (2007).
22. Gazave, E. et al. Posterior elongation in the annelid Platynereis dumerilii involves stem cells

molecularly related to primordial germ cells. Dev. Biol. 382, 246–267 (2013).
23. Paré, L. et al. Transcriptomic landscape of posterior regeneration in the annelid *Platynereis dumerilii*. *BMC Genom.* 24, 583 (2023).
24. Planques, A. et al. DNA methylation atlas and machinery in the developing and regenerating annelid *Platynereis dumerilii. BMC Biol.* 19, 148 (2021).
25. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* 19, 562–578 (2017).
26. García-Castro, H. et al. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. Genome Biol. 22, 89 (2021).
27. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* (2018) doi:10.48550/arxiv.1802.03426.
28. Brunet, T. et al. The evolutionary origin of bilaterian smooth and striated myocytes. *eLife* 5, e19607 (2016).
29. Zakrzewski, A.-C. et al. Early Divergence, Broad Distribution, and High Diversity of Animal Chitin Synthases. Genome Biol. Evol. 6, 316–325 (2014).
30. Zakrzewski, A. Molecular characterization of chaetae formation in annelida and other Lophotrochozoa. (Freie Universität Berlin, Germany, 2011).
31. Song, S. et al. Globins in the marine annelid *Platynereis dumerilii* shed new light on hemoglobin evolution in bilaterians. *BMC Evol. Biol.* 20, 165 (2020).
32. Kostyuchenko, R. P. & Kozin, V. V. Morphallaxis versus Epimorphosis? Cellular and Molecular Aspects of Regeneration and Asexual Reproduction in Annelids. *Biol. Bull.* 47, 237–246 (2020).
33. Kostyuchenko, R. P., Kozin, V. V., Filippova, N. A. & Sorokina, E. V. FoxA expression pattern in two polychaete species, Alitta virens and *Platynereis dumerilii:* Examination of the conserved key regulator of the gut development from cleavage through larval life, postlarval growth, and regeneration. *Dev. Dyn.* 248, 728–743 (2019).
34. Pfeifer, K., Dorresteijn, A. W. C. & Fröbius, A. C. Activation of Hox genes during caudal regeneration of the polychaete annelid *Platynereis dumerilii. Dev. Genes Evol.* 222, 165–179 (2012).
35. Juliano, C. E., Swartz, S. Z. & Wessel, G. M. A conserved germline multipotency program. Development 137, 4113–4126 (2010).
36. Balavoine, G. Segment formation in Annelids: patterns, processes and evolution. Int. J. Dev. Biol. 58, 469–483 (2015).
37. Percharde, M., Bulut-Karslioglu, A. & Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Dev. Cell* 40, 9–21 (2017).
38. Kim, Y.-K. et al. Absolute scaling of single-cell transcriptomes identifies pervasive hypertranscription in adult stem and progenitor cells. *Cell Rep.* 42, 111978 (2023).
39. Grimmel, J., Dorresteijn, A. W. C. & Fröbius, A. C. Formation of body appendages during caudal regeneration in *Platynereis dumerilii:* adaptation of conserved molecular toolsets. EvoDevo 7, 10 (2016).
40. Demilly, A., Steinmetz, P., Gazave, E., Marchand, L. & Vervoort, M. Involvement of the Wnt/β-catenin pathway in neurectoderm architecture in *Platynereis dumerilii. Nat. Commun.* 4, 1915 (2013).
41. Raible, F. et al. Vertebrate-Type Intron-Rich Genes in the Marine Annelid Platynereis dumerilii. *Science* 310, 1325–1326 (2005).
42. Saudemont, A. et al. Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid *Platynereis.* Dev. Biol. 317, 430–443 (2008).
43. Ackermann, C., Dorresteijn, A. & Fischer, A. Clonal domains in postlarval *Platynereis dumerilii* (Annelida: Polychaeta). *J. Morphol.* 266, 258–280 (2005).
44. Fischer, A. H. L. & Arendt, D. Mesoteloblast-Like Mesodermal Stem Cells in the Polychaete

863 Annelid *Platynereis dumerilii* (Nereididae). *J. Exp. Zool. (Mol. Dev. Evol.)* 320, 94–104 (2013).
864 45. Özpolat, B. D., Handberg-Thorsager, M., Vervoort, M. & Balavoine, G. Cell lineage and cell
865 cycling analyses of the 4d micromere using live imaging in the marine annelid *Platynereis*
866 *dumerilii*. eLife 6, e30463 (2017).
867 46. Chou, H.-C., Pruitt, M. M., Bastin, B. R. & Schneider, S. Q. A transcriptional blueprint for a
868 spiral-cleaving embryo. BMC Genom. 17, 552 (2016).
869 47. Dorresteijn, A. W. C. Quantitative analysis of cellular differentiation during early
870 embryo*genesis* of Platynereis dumerilii. *Roux's Arch. Dev. Biol.* 199, 14–30 (1990).
871 48. Elgin, S. C. R. & Reuter, G. Position-Effect Variegation, Heterochromatin Formation, and
872 Gene Silencing in Drosophila. *Cold Spring Harb. Perspect. Biol.* 5, a017780 (2013).
873 49. Zhulyn, O. et al. Evolutionarily divergent mTOR remodels translatome for tissue regeneration.
874 *Nature* 620, 163–171 (2023).
875 50. Johnson, K., Bateman, J., DiTommaso, T., Wong, A. Y. & Whited, J. L. Systemic cell cycle
876 activation is induced following complex tissue injury in axolotl. Dev. Biol. 433, 461–472 (2018).
877 51. Hirose, K., Shiomi, T., Hozumi, S. & Kikuchi, Y. Mechanistic target of rapamycin complex 1
878 signaling regulates cell proliferation, cell survival, and differentiation in regenerating zebrafish
879 fins. *BMC Dev. Biol.* 14, 42 (2014).
880 52. Molinaro, A. M., Lindsay-Mosher, N. & Pearson, B. J. Identification of TOR-responsive slow-
881 cycling neoblasts in planarians. *EMBO Rep.* 22, e50292 (2021).
882 53. Peiris, T. H. et al. TOR signaling regulates planarian stem cells and controls localized and
883 organismal growth. *J. Cell Sci.* 125, 1657–1665 (2012).
884 54. Lund-Ricard, Y., Cormier, P., Morales, J. & Boutet, A. mTOR Signaling at the Crossroad
885 between Metazoan Regeneration and Human Diseases. *Int. J. Mol. Sci.* 21, 2718 (2020).
886 55. Shao, Y. et al. Genome and single-cell RNA-sequencing of the earthworm *Eisenia andrei*
887 identifies cellular mechanisms underlying regeneration. Nat. Commun. 11, 2656 (2020).
888 56. Wagner, D. E., Wang, I. E. & Reddien, P. W. Clonogenic Neoblasts Are Pluripotent Adult
889 Stem Cells That Underlie Planarian Regeneration. Science 332, 811–816 (2011).
890 57. Reddien, P. W. The Cellular and Molecular Basis for Planarian Regeneration. Cell 175, 327–
891 345 (2018).
892 58. Varley, Á., Horkan, H. R., McMahon, E. T., Krasovec, G. & Frank, U. Pluripotent, germ cell
893 competent adult stem cells underlie cnidarian regenerative ability and clonal growth. *Curr. Biol.*
894 33, 1883-1892.e3 (2023).
895 59. Holstein, T. W. The Hydra stem cell system – Revisited. *Cells Dev.* 174, 203846 (2023).
896 60. Álvarez-Campos, P. et al. Annelid adult cell type diversity and their pluripotent cellular
897 origins. *bioRxiv* 2023.04.25.537979 (2023) doi:10.1101/2023.04.25.537979.
898 61. Denes, A. S. et al. Molecular Architecture of Annelid Nerve Cord Supports Common Origin of
899 Nervous System Centralization in Bilateria. Cell 129, 277–288 (2007).
900 62. Bely, A. E. Early Events in Annelid Regeneration: A Cellular Perspective. Integr. Comp. Biol.
901 54, 688–699 (2014).
902 63. Smith, K. N., Singh, A. M. & Dalton, S. Myc Represses Primitive Endoderm Differentiation in
903 Pluripotent Stem Cells. *Cell Stem Cell* 7, 343–354 (2010).
904 64. Bernabé-Rubio, M. et al. Myc-dependent dedifferentiation of Gata6+ epidermal cells
905 resembles reversal of terminal differentiation. *Nat. Cell Biol.* 25, 1426–1438 (2023).
906 65. Koh, F. M. et al. Emergence of hematopoietic stem and progenitor cells involves a Chd1-
907 dependent increase in total nascent transcription. *Proc. Natl. Acad. Sci.* 112, E1734–E1743
908 (2015).
909 66. Bideau, L. et al. Antero-posterior gradients of cell plasticity and proliferation modulate
910 posterior regeneration in the annelid *Platynereis*. bioRxiv 2023.06.12.544593 (2023)
911 doi:10.1101/2023.06.12.544593.
912 67. Satoh, A., Cummings, G. M. C., Bryant, S. V. & Gardiner, D. M. Neurotrophic regulation of

913 fibroblast dedifferentiation during limb skeletal regeneration in the axolotl (*Ambystoma*
914 *mexicanum*). Dev. Biol. 337, 444–457 (2010).
915 68. Satoh, A., Gardiner, D. M., Bryant, S. V. & Endo, T. Nerve-induced ectopic limb blastemas in
916 the axolotl are equivalent to amputation-induced blastemas. Dev. Biol. 312, 231–244 (2007).
917 69. Nohno, T. et al. A Chicken Homeobox Gene Related to Drosophila paired Is Predominantly
918 Expressed in the Developing Limb. Dev. Biol. 158, 254–264 (1993).
919 70. Logan, M. et al. Expression of Cre recombinase in the developing mouse limb bud driven by
920 a Prxl enhancer. genesis 33, 77–80 (2002).
921 71. Gerber, T. et al. Single-cell analysis uncovers convergence of cell identities during axolotl
922 limb regeneration. Science 362, eaaq0681 (2018).
923 72. Lin, T.-Y. et al. Fibroblast dedifferentiation as a determinant of successful regeneration. Dev.
924 Cell 56, 1541-1551.e6 (2021).
925 73. Zantke, J. et al. Circadian and Circalunar Clock Interactions in a Marine Annelid. Cell Rep. 5,
926 99–113 (2013).
927 74. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
928 *EMBnetJ.* 17, 10–12 (2011).
929 75. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data.
930 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
931 76. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for
932 multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016).
933 77. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2012).
934 78. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
935 assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
936 79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
937 RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
938 80. Kumar, L. & Futschik, M. E. Mfuzz: A software package for soft clustering of microarray data.
939 *Bioinformation* 2, 5–7 (2007).
940 81. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
941 82. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272 (2019).
942 83. Steger, J. et al. Single-cell transcriptomics identifies conserved regulators of neuroglandular
943 lineages. Cell Rep. 40, 111370 (2022).
944 84. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A Greedy Algorithm for Aligning DNA
945 Sequences. *J. Comput. Biol. : a J. Comput. Mol. cell Biol.* 7, 203–214 (2000).
946 85. Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in
947 2020. *Nucleic Acids Res.* 49, gkaa937- (2020).
948 86. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von & Minh, B. Q. IQ-TREE: A Fast and Effective
949 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–
950 274 (2015).
951 87. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online
952 phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44, W232–W235 (2016).
953 88. Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q. & Vinh, L. S. UFBoot2:
954 Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35, 518–522 (2018).
955 89. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
956 display and annotation. Nucleic Acids Res. 49, gkab301- (2021).
957 90. Blanco-Carmona, E. Generating publication ready visualizations for Single Cell
958 transcriptomics using SCpubr. bioRxiv 2022.02.28.482303 (2022)
959 doi:10.1101/2022.02.28.482303.
960 91. Kuehn, E. et al. Segment number threshold determines juvenile onset of germline cluster
961 expansion in *Platynereis dumerilii*. *J. Exp. Zoöl. Part B: Mol. Dev. Evol.* 338, 225–240 (2022).
962 92. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods*

963 9, 676–682 (2012).
964 93. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases.
965 Nat. Methods 6, 343–345 (2009).
966 94. Backfisch, B. et al. Stable transgenesis in the marine annelid *Platynereis dumerilii* sheds new
967 light on photoreceptor evolution. Proc. Natl. Acad. Sci. 110, 193–198 (2013).