Modular, Cascade-like Transcriptional Program of Regeneration in *Stentor*

Pranidhi Sood[1], Athena Lin[1], Rebecca McGillivary[1], and Wallace F. Marshall[1,2]

1. Department of Biochemistry & Biophysics
University of California, San Francisco
San Francisco, CA 94158

2. Chan Zuckerberg Biohub, San Francisco, CA 94158

Abstract
The giant ciliate *Stentor coeruleus* is a classical model system for studying regeneration and morphogenesis at the level of a single cell. The anterior of the cell is marked by an array of cilia, known as the oral apparatus, which can be induced to shed and regenerate in a series of reproducible morphological steps, previously shown to require transcription. If a cell is cut in half, each half will regenerate an intact cell, including a new oral apparatus in the posterior half. We used RNAseq to assay the dynamic changes in *Stentor*'s transcriptome during regeneration, after both oral apparatus shedding and bisection, allowing us to identify distinct temporal waves of gene expression including kinases, RNA binding proteins, centriole biogenesis factors, and orthologs of human ciliopathy genes implicated in Meckel and Joubert syndromes. By comparing transcriptional profiles of different regeneration events in the same species, we were able to identify distinct modules of gene expression corresponding to oral apparatus regeneration, posterior holdfast regeneration, and recovery after wounding. By measuring gene expression in cells in which translation is blocked, we show that the sequential waves of gene expression involve a cascade mechanism in which later waves of expression are triggered by translation products of early-expressed genes. Among the early-expressed genes, we identified an E2F transcription factor and the conserved RNA binding protein Pumilio as potential regulators of regeneration based on the expression pattern of their predicted target genes. This work allows us to classify regeneration genes into groups based on their potential role for regeneration in distinct cell regeneration paradigms, and provides new insight into how a single cell can coordinate complex morphogenetic pathways to regenerate missing structures.

Keywords: ciliate morphogenesis, single-cell regeneration, cellular wound healing, regeneration, transcriptomics

Introduction

Regeneration and wound healing are processes that are typically studied at the tissue level in multicellular organisms. A cell scale response to injury is a crucial feature of repair even in multicellular organisms, as individual cells also must be able to repair wounds following mechanical disruption. Injured cells must be able to not only patch over the site of injury to prevent leakage of cytoplasm, they also need to re-establish polarity, rebuild organelles and reorganize the cytoskeleton (Tang & Marshall, 2017). Deficiencies in cell repair have become implicated in disease, for example in diseases of the heart, lung, and nervous system (Oeckler 2008; Wang 2010; Angelo 2015), including ARDS (Cong 2017). Yet there is still much to be learned about how an individual cell responds to a wound.

A second reason to study regeneration in single cells is to study the mechanisms of how cells build and maintain their shape and organization. The study of regeneration has played a key role in revealing mechanisms of animal development, because regeneration allows developmental process to be induced experimentally (Morgan 1901a). Similarly, understanding how cells are able to rebuild cellular components and re-establish global patterning holds the promise of shedding new light on the largely unanswered fundamental question of how cells perform morphogenesis and pattern formation (Kirschner 2000; Shulman 1999; Harold 2005; Marshall 2020). Ciliates, in particular, provide an excellent opportunity for investigating the mechanisms of cellular patterning and regeneration due to the fact that they have highly stereotyped cell shapes, with easily visible surface structures that serve as landmarks to assess the progress of regeneration (Aufderheide 1980; Frankel 1989).

The giant ciliate *Stentor coeruleus* (Slabodnick 2014a) is a single cell that can fully regenerate its complex subcellular structure after injury. In this classical system, virtually any portion of the cell, when excised, will give rise to a normally proportioned cell with intact subcellular organization (Morgan 1901b; Tartar 1961). *Stentor* provides a unique opportunity to study regeneration and patterning at the cell scale. Its large size (up to 1-2 millimeters in length), clear anterior/posterior axis, detailed cortical patterning, and remarkable ability to heal even large wounds in the cell membrane make it especially amenable to surgical manipulation and imaging approaches. Importantly, in *Stentor*, principles of single-cell regeneration can be studied without confounding effects of surrounding cells that may non-autonomously influence an intracellular injury response in the context of tissues. Studies in *Stentor* thus are expected to reveal key general features of wound healing, regeneration, and morphogenesis at the scale of an individual cell.

Transcriptome studies of various multicellular animal species including zebrafish, axolotl, and planaria, all of which are capable of regeneration throughout life, have begun to reveal key regulators of multicellular regeneration. Many of these studies have delineated the molecular players in regeneration by identifying genes that are expressed when stem cells differentiate into various cell types required to rebuild lost tissue or organs. This transcriptomic approach has thus proven its utility in revealing cell-specific requirements for regeneration in the context of tissues. We have sought to take a similar approach to the problem of single-cell regeneration in *Stentor*, focusing on two regenerative processes: regeneration of the oral apparatus (OA) after its removal by sucrose shock, and regeneration of cells that have been bisected into two half-cells.

One of the most dramatic and tractable regeneration paradigms in *Stentor* is the regeneration of the oral apparatus (OA). The oral apparatus is a prominent structure on the anterior side of the cell that contains thousands of basal bodies and cilia (Paulin and Bussey 1971) organized into a ciliated ring known as a membranellar band. At one end of the ring is an invagination of the

plasma membrane, where food particles are ingested. This invagination together with its associated cytoskeletal structures is known as the mouth. The OA can be induced to shed using sucrose shock (Tartar 1957) after which a new OA regenerates over the course of 8 hours, progressing through a series of well-characterized morphological stages (**Figure 1A**; Tartar 1961). Removal of the macronucleus, at any stage, causes OA regeneration to halt at the next stage, suggesting that several waves of gene expression may be required to drive different processes at different stages (Tartar 1961). Chemical inhibitor studies showed that regeneration of the oral apparatus requires transcription (Whitson 1965; James 1967; Burchill 1968; Younger 1972) and it is also known that overall levels of RNA synthesis increase several fold during the OA regeneration process (Ellwood 1966; Burchill 1968; Younger 1972).

The second regeneration paradigm we consider is bisection (**Figure 1B**). When a *Stentor* cell is cut in half along its longitudinal axis, the posterior half needs to regenerate a new OA, while the anterior half has to regenerate a new tail. Thus, the list of genes expressed during regeneration in posterior halves of bisected cells presumably includes genes involved in building a new OA, but is also expected to include genes involved in building other anterior structures such as the contractile vacuole, while the list of genes expressed in the anterior half would include genes involved in rebuilding lost posterior structures. Both halves are expected to express genes involved in recovering the normal metabolic state of the cell following closure of the large wound created in the plasma membrane during surgery.

Given that transcription is required for regeneration, there is likely to be a set of genes whose products drive the regenerative process. By learning the identity of these genes, we can determine the molecular pathways involved in building new cellular structures. At the same time, knowledge of the transcriptional program of regeneration would provide a molecular foothold to identify the upstream signals that trigger the process. This type of approach has previously been used successfully to identify genes involved in ciliogenesis, by identifying genes expressed in cells as they regenerate flagella (Schloss 1984; Stolc 2005; Albee 2013). These flagellar regeneration studies showed that transcriptomics is a viable approach to studying organelle biogenesis and showed that in addition to identifying structural components of the organelle itself, transcriptomics also reveals genes whose products are not incorporated into the final structure but are required for building it. In the case of a complex regenerative process such as cellular regeneration in *Stentor*, this latter strength is of particular importance.

Although *Stentor* regeneration was the focus of study for over 100 years, the lack of genetic tools in the organism prevented detailed molecular analysis of its processes. Now that the *Stentor coeruleus* genome has been determined (Slabodnick, 2017) and RNA interference methods developed (Slabodnick, 2014b), molecular and genomic approaches enable new ways to explore this classic model system. It has been possible to use RNA sequencing to show that a portion of the genome becomes expressed during regeneration in *Stentor* (Sood 2017; Onsbring 2018; Wei 2020). These studies focused on two different regeneration paradigms. In Sood (2017) and Wei (2020), genes were identified that were upregulated during oral apparatus regeneration in *Stentor coeruleus*, induced using sucrose shock or urea shock, respectively. The list of genes identified was presumably a mixture of those involved in building the new oral apparatus and genes involved in recovery from the stress of sucrose or urea shock that was used to remove the OA. In Onsbring 2018, regeneration of bisected *S. polymorphus* cells were analyzed. Each of these three prior studies focused on a different regeneration paradigm in different species. However, by comparing the transcriptional programs of regeneration in sucrose shocked versus bisected

cells within the same species, it is possible to distinguish shared modules of gene expression, common to all regenerative processes, from structure-specific regeneration modules.

Here we report a comparative transcriptomic analysis in which RNA sequencing was used to analyze gene expression during OA regeneration following sucrose shock as well as regeneration in posterior and anterior cell fragments following surgical bisection. By comparing gene sets expressed in these different cell fragments, we can identify genes specific to regeneration of the OA and the posterior tail, as well as additional sets of genes involved in the recovery from surgical wounding. The timing with which specific groups of genes are expressed correlates with the formation of specific cellular structures. These results suggest a modular organization of the regeneration program. By focusing on earlier stages of regeneration, we identified conserved transcriptional regulators as well as RNA binding proteins that are differentially expressed during regeneration. Inhibition of protein translation allows the transcriptional program to be divided into an early set of genes whose expression does not require protein synthesis, and a later set of genes whose expression is apparently dependent on production of proteins encoded by the early genes, suggesting a cascade-like logic. This work opens a new window into the molecular details underlying the century-old question of regeneration in this extraordinary single celled organism.

Materials and Methods

Stentor culture and regeneration

Cells were obtained from Carolina Biological Supply and cultured as previously described (Slabodnick 2013). Cells were maintained in Pasteurized Spring Water (Carolina Biological Supply) and fed with *Chlamydomonas* and wheat seeds. Cells were collected from the same culture for each RNAseq experimental replicate.

To induce regeneration of the oral apparatus by sucrose shock (Lin 2018), cells were placed in a 15% sucrose solution for 2 minutes (Tartar 1957), and then washed in Carolina Spring Water thoroughly. Samples of ~20 cells were collected before shock, then at 30 minutes post shock, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours and 8 hours. At each time point, a sample of cells was lysed into RNA-stabilizing buffers specified by the extraction kit, and then stored on ice until the end of the experiment when the RNA purification was performed in parallel on all samples (see below). Four replicates were analyzed for each time-point.

For analysis of regeneration in bisected cells, cells were cut in half as previously described (Slabodnick 2013). A 50µl drop of methylcellulose was placed onto a slide, and 40-50 *Stentor* cells were collected in a volume of 50µl and added to the drop of methylcellulose. Cells were individually cut with a glass needle (prepared as described in Slabodnick 2013), making sure to complete all cutting within 10 min from the time the first one was started. As a result, all samples are synchronous to within 10 minutes of each other. The anterior and posterior half-cells were manually separated into two tubes and washed once with spring water to remove the methylcellulose. Samples were then incubated at room temperature for the specified time period (60, 90, 120, 180, 360, or 420 min). Samples designated as t=0 were collected within two minutes of cutting. After the desired time had elapsed, the media was removed from the cells to produce a final volume of less than 20µl. 350µl of RLT buffer from Qiagen micro easy kit was added to the sample and mixed by pipetting 20 times. Lysate for each sample was stored at 4 ºC while additional samples were prepared.

Total RNA extraction

RNA was extracted at each time point using the Nucleospin RNA XS kit from Clontech (cat. num. 740902.250). RNA quality was assessed using a NanoDrop and then Bioanalyzer was used to quantify RNA amount. ERCC spike ins (ThermoFisher cat. num. 4456739) were added to each sample in a dilution ranging from 1:1000 to 1:10000 depending on the initial amount of RNA extracted.

RNA-Seq library preparation and sequencing

RNA-seq libraries were prepared with Ovation RNA-seq system v2 kit (NuGEN). In this method, the total RNA (50 ng or less) is reverse transcribed to synthesize the first-strand cDNA using a combination of random hexamers and a poly-T chimeric primer. The RNA template is then partially degraded by heating and the second strand cDNA is synthesized using DNA polymerase. The double-stranded DNA is then amplified using single primer isothermal amplification (SPIA). SPIA is a linear cDNA amplification process in which RNase H degrades RNA in DNA/RNA heteroduplex at the 5′-end of the double-stranded DNA, after which the SPIA primer binds to the cDNA and the polymerase starts replication at the 3′-end of the primer by displacement of the existing forward strand. Random hexamers are then used to amplify the second-strand cDNA linearly. Finally, libraries from the SPIA amplified cDNA were made using the Ultralow V2 library kit (NuGEN). The RNA-seq libraries were analyzed by Bioanalyzer and

quantified by qPCR (KAPA). High-throughput sequencing was done using a HiSeq 2500 instrument (Illumina). Libraries were paired-end sequenced with 100 base reads.

RNAseq data preparation -- trimmed and filtered reads

We used trimmomatic (Bolger 2014) to trim RNAseq reads with the following flags: ILLUMINACLIP:$adapterfile:2:30:10 HEADCROP:6 MINLEN:22 AVGQUAL:20 The settings ensured that we kept reads of at least 22 bases, an average quality score of 20 and trimmed any remaining Illumina adapter sequences.

Transcriptome generation

To generate a transcriptome, we combined all the reads from all RNAseq samples and timepoints. We ran Tophat2 (Kim 2013) to align the reads to the genome (StentorDB http://stentor.ciliate.org; Slabodnick 2017) We used the following flags to ensure proper mapping in spite of Stentor's unusually small introns (Slabodnick 2017): -i 9 -I 101 --min-segment-intron 9 --min-coverage-intron 9 --max-segment-intron 101 --max-coverage-intron 101 -p 20

We then ran Trinity (Haas 2013) using a genome guided approach. We used the following flags: --genome_guided_max_intron 1000.

In total, we assembled 34,501 genes aligned with the existing *Stentor* genome models in StentorDB, plus 143 novel models. Our prior gene prediction for the *Stentor* genome indicated approximately 35,000 genes were present (Slabodnick 2017). Such a large number of genes is typical of ciliates.

Calculating transcript abundance and differential expression analysis

We used Kallisto to quantify transcript abundance (Bray 2016) using the following flags: -t 15 -b 30. We then used Sleuth (Pimentel 2017) to identify genes which are differentially expressed genes through the regeneration time course. We use an approach similar to that used by Ballgown (Frazee 2016). We fit the expression data to time using natural splines (R function "ns") with 3 degrees of freedom. Then, using Sleuth, we compared this model to a null model where change in expression is only dependent upon noise. To decide if transcripts were differentially expressed, we defined the minimum significance value (qval in the Sleuth model) to be ten times the minimum significance value of all the ERCC spike-in transcripts. We found that nearly 5583 transcripts are differentially expressed during oral apparatus regeneration. Of these, 485 had no clear homology to proteins in NCBI databases nor PFAM. We identified 234 that did not map to existing gene models. We averaged the expression of all transcripts that mapped to gene models as well as those which were part of a Trinity transcript cluster. All subsequent analysis was performed on these averaged values. Clustering analysis was performed as follows. First, genes whose maximum expression among the post-shock timepoints was found 30 minutes after sucrose shock were put into one cluster manually. Gene expression profiles before sucrose shock and thirty minutes after are highly correlated (correlation coefficient from Pearson's correlation = 0.99). Then, the remaining genes were clustered into 4 groups using "clara".

Annotation of transcriptome

Following the approach of trinotate (https://trinotate.github.io), we annotated the transcriptome. First we used transdecoder (http://transdecoder.github.io/) to find the longest

ORFs (minimum protein length is 100AA and uses the standard genetic code). We used blastx and blastp (Altschul 1990) to search the Uniprot database (Uniprot consortium 2017). Then Hmmscan (hmmer.org, HMMER 3.1b1) was used to search the pfam-a database (Finn 2016). Alignments of genes of interest were further manually inspected using a blastp search against the "Model Organism" or "Uniprot-KB/Swiss-Prot" databases.

Mapping Transcripts to gene models and to genome

We used Gmap (Wu 2005) to map transcripts to gene models following the approach outlined here:
https://github.com/trinityrnaseq/RagonInst_Sept2017_Workshop/wiki/genome_guided_trinity.

We used a built in script from Trinity to utilize gmap to align transcripts to a repeat-masked (rmblastn 2.2.27+) Stentor genome. We used bedtools (Quinlan 2014) on the resulting bam file to identify overlaps between the aligned transcripts and existing gene models (Slabodnick 2017). Annotations from StentorDB were collected, most of which refer to other ciliate genomes. For any gene with a predicted domain or a ciliate homolog, BLAST search was performed against the Chlamydomonas genome versions 5.5 on the JGI Phytozome database.

Annotation of subsets of genes

Because centriole and cilia genes have often been poorly annotated in existing databases, we manually curated "ancestral centriole genes" (Azimzadeh 2012) and other genes involved with ciliogenesis and centriole biogenesis (Ishikawa 2011; Bettencourt-Dias 2007). We used a reciprocal best BLAST search approach to identify genes in the *Stentor* genome with homologs to these manually curated sets of genes.

Results

Identifying gene expression modules by comparative transcriptomics

Stentor coeruleus cells were subjected to sucrose shock to remove the OA (**Figure 1A**), or else surgically bisected to produce anterior and posterior half-cells (**Figure 1B**). These experiments produced three regenerating samples: intact cells from which the OA had been removed, which then regenerated the OA; posterior half-cells which regenerated a new anterior half, including the OA and other anterior structures; and anterior half-cells which retained the pre-existing OA and regenerated a new posterior half including contractile tail and holdfast.

For each sample, cells were collected at regular intervals and analyzed by RNA sequencing (see Materials and Methods). The process of building a new OA takes ~8 hours as detailed in **Figure 1**. RNA samples from ~20 cells were collected prior to sucrose shock, at 30 min after shock, and at 1, 2, 3, 4, 5, 6, 7, and 8 hours after sucrose shock. The number of time points was based on the time required to complete regeneration. At each stage, RNA was extracted, and RNA-seq libraries were sequenced (see Materials and Methods). We combined reads from all samples and replicates and used these to assemble a genome-guided *de novo* transcriptome (Grabherr 2011) using TRINITY. To identify genes with dynamic expression patterns during regeneration, we first mapped reads from each sample to this transcriptome and then identified differentially expressed genes over the time course in each sample. To identify genes that were differentially expressed we compared two models – a generalized additive model where changes in expression over time are modeled by natural splines, and one in which there is no dependence on time.

Overall, we identified 4,323 genes that exhibited dynamic expression patterns through regeneration in sucrose shocked cells, 1,020 in posterior half-cells following bisection, and 229 in anterior half-cells following bisection. As indicated in **Figure 1C**, these three samples showed partially overlapping expression patterns, but each also expressed its own unique set of genes. The anterior half-cells showed by far the smallest number of differentially expressed genes, while the sucrose shocked cells showed the largest. In total we detected 4,811 differentially expressed genes, which constitutes roughly 10% of the *Stentor* genome. Based on the Venn diagram in **Figure 1C**, we defined six sets of differentially expressed genes: genes expressed in both sucrose shock and regenerating posterior halves, which we take to indicate genes required for OA regeneration; sucrose-shock specific, which we interpret as reflecting aspects of OA regeneration in sucrose shocked but not bisected cells, possibly including osmotic response to the sucrose shock itself; genes expressed only in regenerating posterior halves, which we interpret as relating to regeneration of anterior structures other than the OA; genes expressed only in regenerating anterior halves, which we interpret as being involved in regenerating the tail of the cell; genes expressed in both regenerating half cells but not sucrose shock, which we term bisection-specific; and, finally, genes expressed in all three samples, which we take to indicate general regeneration genes. Each of these sets of genes were clustered using clara (Kaufman 1990).

Differentially expressed genes specific for OA regeneration

OA regeneration can be stimulated in a number of ways. If *Stentor* cells are treated with sucrose, the OA detaches via an autotomy process (Tartar 1957). If *Stentor* cells are cut in half, the posterior half will regenerate a new OA. Analyzing genes expressed in either situation alone will reveal OA regeneration-specific genes but also genes that may be induced by the stresses of

sucrose shocking or bisection, respectively. Thus, in order to obtain a list of highest confidence OA-specific genes, we compared sucrose shocked cells and regenerating posterior half-cells in order to identify a set of genes that are differentially expressed in both samples and thus likely to be specific for OA regeneration. These genes fall into five clusters (**Figure 2A**). Cluster 1 consists of genes whose expression decreases during regeneration, possibly suggesting that these genes encode proteins whose function is dispensable for, or possibly even inhibitory of, the regeneration process. The other four clusters correspond to upregulated genes whose peak of expression takes place at successively later times during regeneration. The identity of OA-specific genes corresponding to gene models in the published *Stentor* genome (StentorDB) are listed in **Supplemental Table S1**. Classes of genes present in the OA-specific module are categorized in Figure 2B,which includes genes encoding centriole-related proteins, kinases, and RNA binding proteins.

Among the annotated genes in the OA-specific set (**Supplemental Table S1**) many were found that are related to centriole biogenesis, in keeping with the fact that a large number of new basal bodies form during OA regeneration in order to act as basal bodies for the cilia of the membranellar band of the OA. **Table 1** lists the centriole-related genes found to be upregulated in the OA-specific module. It is notable that most of these are expressed in cluster 2 and 3, consisting of genes whose expression peaks at 1.5-2 hours after sucrose shock, during the period of time at which the first evidence of an oral primordium becomes visible in scanning electron microscopy (Paulin and Bussey 1971), and at which transmission electron microscopy reveals that thousands of new basal bodies are being formed de novo to form a so-called "anarchic field" (Bernard 1981). These basal bodies will ultimately organize themselves into arrays and become the basal bodies that nucleate the ciliature of the OA. Among these are SAS6, SAS4, POC1,CETN3, CETN2, CEP76, CEP135, CEP120. SAS6 is notable as one of the earliest known factors involved in assembling the ninefold symmetric structure of the centriole (Leidel 2005). Out of the 29 most conserved ancestral centriole genes, we find that 10 are expressed in either cluster 2 or 3. This expression of the core centriole gene set at the exact stage when basal bodies are forming thus provides a biological confirmation of our analysis. One centriole-related gene, LRC45, is expressed later than the others in cluster 4. Implications of this delayed expression of a centriole-related gene will be discussed below. A handful of genes involved in ciliary assembly are expressed in later clusters, including several IFT proteins, but almost no genes encoding components of the motile ciliary machinery were differentially expressed in any cluster.

Sucrose shock-specific genes

In contrast to the OA-specific module, which was defined by looking for genes shared in common between the expression program during OA regeneration in bisected cells and sucrose shocked cells, the sucrose shock specific module groups genes differentially expressed only in sucrose shocked cells. **Figure 2C** depicts genes showing differential expression in sucrose shocked cells but not in regenerating posterior or anterior half-cells. These genes may be involved in stress response to the sucrose shock, or remodeling the remnant of the previous OA left behind after the shock. They may also represent redundant additional OA biogenesis factors that happen to be expressed during sucrose shock only.

Similar to the OA-specific module (**Figure 2A**), we observe five clusters of gene expression among the sucrose shock specific genes (**Figure 2C**). The timing of these clusters matches the that seen in the OA specific module. The identity of sucrose shock-specific genes corresponding

to gene models in the published Stentor genome (StentorDB) are listed in **Supplemental Table S2**. Genes upregulated during sucrose shock response included 12 of the 29 ancestral centriole genes. As with the OA specific gene set, these genes were enriched in cluster 2 of the sucrose shock response, for example POC1, OFD1, CEP164, and CP131/AZI. In many cases, the sucrose shock module includes paralogs of these genes that were not expressed in the OA specific module, consistent with the idea of redundancy between bisection and sucrose shock induced OA formation. Cluster 3 contains genes relating to ciliary assembly such as the intraflagellar transport proteins. Within the sucrose-specific module, the expressed paralog of the key centriole biogenesis initiating protein SAS6 is not expressed highly until late in regeneration (Cluster 5).

In stark contrast to the lack of cilia-related genes in the OA specific gene set presented above, we found that clusters 3 and 4 of the sucrose shock specific gene set contain a large number of genes that encode proteins components of motile cilia. Expression of genes encoding components of the inner and outer dynein arms, which power ciliary motility, are see across clusters 2, 3, and 4. In addition to the dynein arms themselves, two other multi-protein complexes are required to coordinate dynein activity, radial spokes (Smith and Yang 2004) and the dynein-regulatory complex (Viswanadha et al., 2017). Genes encoding the dynein regulatory complex are expressed exclusively during cluster 3 (DRC 1, 2, 3, 4, 5, 7, 9, 10 and 11). Expression of genes encoding radial spoke components begins in cluster 3 and is most apparent in cluster 4, where eight radial spoke proteins are expressed (RSP1, 3, 4, 7, 9, 10, 14, and 16). The radial spokes interact with the central pair microtubule complex, and several central pair specific proteins are upregulated during clusters 2-4 (PF6, PF20, CPC1, Hydin). The axoneme of motile cilia contains structural proteins that are located at the junction between the A and B tubules. These junctional proteins are expressed in clusters 3 and 4 (Rib43a, Rib72, PACRG, FAP20/BUG22). We thus find that clusters 3 and 4, ranging in expression timing from 120-300 min, contain many genes involved in supporting ciliary motility. None of these genes is required for the assembly of cilia, but instead are involved in coordinating the activity of axonemal dyneins to generate motility (Zhu 2017). In *Chlamydomonas*, radial spoke protein synthesis reaches its maximum rate 30-60 minutes after the flagella have begun assembling onto pre-existing basal bodies (Remillard 1982). This timing roughly matches the delay of one hour seen in our data between the peak expression of genes involved in ciliary assembly (cluster 2) and genes encoding radial spokes during *Stentor* regeneration (Cluster 4). The timing of group 4 also correlates with the time period during which the oral cilia transition from their initial random beating motility to their characteristic coordinated beating motility, forming metachronal waves (Paulin and Bussey 1971; Wan 2020).

Overall, we observe that the peak expression of genes encoding centriole proteins occurs earlier than the peak expression of genes encoding motile cilia proteins (**Figure 2E**), which is consistent with the fact that ciliogenesis takes place later than basal body biogenesis in *Stentor*.

Regeneration of non-oral anterior structures

In bisected cells, the posterior half-cell regenerates anterior structures, including the oral apparatus but also other structures such as the contractile vacuole and cytopyge. The contractile vacuole can still regenerate in enucleated posterior half-cells (Stevens 1903; Tartar 1956), but this does not necessarily mean that contractile vacuole related genes are not upregulated during normal regeneration. Our data show three clusters of gene expression specific for regeneration of anterior structures in posterior half-cells (**Figure 2F,G**). Cluster 1 represents down-regulated

genes. Cluster 2 spans the 60-180 minute time-points and therefore matches the expression of cluster 3 of the OA specific program (**Figure 2A**). In Cluster 3, expression drops midway through regeneration (120 min) and then rises and peaks around 360 min [240-420 minutes]. This expression thus resembles the timing of cluster 5 of the OA specific gene set. Not only does the timing resemble cluster 5 of the OA specific genes, the types of genes that are in this cluster are also similar, specifically EF hand, shippo-rpt proteins, and glutathione s-transferase. The identity of anterior regeneration-specific genes corresponding to gene models in the published Stentor genome (StentorDB) are listed in **Supplemental Table S3**.

The same lack of cilia-specific gene expression that is seen in the shared OA regeneration (**Figure 2A**) is also seen in posterior halves regenerating anterior structures. If we consider all clearly annotated genes encoding centriole structural proteins or assembly factors (**Figure 2H**) we see that the vast majority were observed in the sucrose shock experiment. The same trend is seen to an even greater extent for genes encoding protein components of motile cilia (**Figure 2I**).

Posterior tail regeneration

In a bisected cell, the anterior half cell regenerates a new posterior half, including the contractile tail and holdfast. We refer to the gene expression module seen in these cells as the "tail regeneration" program (**Figure 3A**). Two clusters of expression were found during tail regeneration, a cluster of genes that are down-regulated during tail regeneration, and a single cluster of genes that are upregulated during regeneration. The tail regeneration-specific transcriptional program is dramatically different from that seen in OA regeneration, in terms of both the number and type of genes that show differential expression. The number of genes upregulated during tail regeneration (13) is far smaller than the number upregulated during OA regeneration (564; **Figure 2A clusters 2-5**), consistent with the fact that the posterior holdfast can still regenerate in enucleated cells and therefore does not rely on transcription for its regeneration (Tartar 1956). Increased expression for many of the tail-regeneration genes does not occur until at least 120 minutes, by which time it has been reported that the holdfast has already completely regenerated (Morgan 1901; Weisz 1951), suggesting that the upregulation of these genes is taking place after assembly, perhaps to regenerate depleted pools of precursor protein.

Unlike OA regeneration, tail regeneration does not involve expression of any known centriole or cilia-related genes. Instead, the expressed gene set is dominated by EF-hand proteins, which constitute four of the nine upregulated genes for which annotation data exists. Studies of cellular structure in Stentor showed that the posterior half of the cell contains long contractile fibers composed of centrin-like EF hand proteins (Huang and Pitelka, 1973; Maloney 2005), and studies of cell movement showed that the contractile behavior of the cell, which is driven by these EF hand protein fibers, occurs primarily in the posterior half of the cell (Newman 1972). The identity of tail regeneration-specific genes corresponding to gene models in the published Stentor genome (StentorDB) are listed in **Supplemental Table S4**.

Bisection-specific genes

Sucrose shock removes the OA cleanly, without creating a wound. In contrast, surgical bisection disrupts the membrane with visible loss of cytoplasm. In order to investigate the molecular response to this wounding, we looked for genes that showed similar patterns of differential expression in both halves of bisected cells, but not in sucrose shocked cells. This analysis revealed 98 differentially expressed genes which grouped into three clusters based on

temporal pattern (**Figure 3B**). The identity of bisection-specific genes corresponding to gene models in the published *Stentor* genome (StentorDB) are listed in **Supplemental Table S5**.

Cluster 1 consists of genes that are turned off rapidly during regeneration of bisected cells, and consists predominantly of metabolic enzymes and chaperones. Cluster 2 are genes that turn off more gradually during bisected regeneration, with expression levels reducing during the first hour of regeneration. This cluster includes a number of proteases. We noted that cluster 2 of the bisection response (slow down-regulation) contains a number of genes encoding protein classes similar to those seen in cluster 1 (rapid down-regulation) of the OA specific gene set. Specifically, both gene sets include orthologs of von Willebrand factor domain protein, serine carboxypeptidase, papain family cysteine protease, glycosyl hydrolase, and aldo/keto reductase. The genes are different in the two datasets but encoding similar proteins. This similarity suggests that similar genes are inactivated during regeneration in bisected and sucrose shocked cells, but with slower kinetics of repression in the bisected cells.

Cluster 3 contains genes whose expression increases during bisected regeneration, with the peak of expression generally in the range of 120-240 min. This cluster includes membrane transporters as well as carbonic anhydrase.

Genes shared by all regeneration processes

Is regeneration a single process, or a collection of distinct processes that depend on which part is missing? By considering the overlap of expression patterns among all samples analyzed, we identified candidates for general regeneration genes expressed during all forms of regeneration (**Figure 3C**). The identity of the general regeneration genes corresponding to gene models in the published *Stentor* genome (StentorDB) are listed in **Supplemental Table S6**. Far fewer genes were contained in this group than in any other regeneration module, suggesting that most regeneration genes are specific to distinct aspects of regeneration. The small number of general genes could potentially play a role in building replacement cortical structures shared by all parts of the cell or in recognition by the cell that regeneration is taking place.

Cluster 1 consists of genes whose expression is reduced in all forms of regeneration. This cluster consists mostly of genes encoding metabolic proteins, and suggests a general trend, also seen in other modules, for the cell to down-regulate metabolic activity during regeneration. Cluster 2 and 3 are genes upregulated in all forms of regeneration, while cluster 4 consists of genes that show differential expression in all three cases, but unlike clusters 2 and 3, cluster 4 genes are upregulated following sucrose shock but down-regulated in both halves of bisected cells. The genes in these clusters do not fall into any discernable functional families.

Regeneration in the absence of translation

A central question in regeneration is the nature of the stimulus that triggers the appropriate transcriptional response. As an initial step towards addressing that question, we ask whether the triggering signal, whatever it is, directly drives expression of all the genes upregulated during regeneration (**Figure 4A**) or whether, instead, the triggering signal may drive a sub-set of genes which then, in turn, drive subsequent rounds of gene expression, thus leading to a cascade-like mechanism (**Figure 4B**). In order to distinguish between these direct and cascade schemes, we repeated the RNAseq analysis of OA regeneration in sucrose shocked cells treated with cycloheximide to block translation. This treatment would not be expected to affect pre-existing proteins that constitute the triggering stimulus pathway, but would prevent any of the direct targets of the triggering stimulus from being produced to trigger subsequent waves of expression.

Out of the 431 genes that show differential expression in the OA-specific regeneration module (**Figure 2A**), we found that 21 genes, constituting only 5% of the total gene set, were completely unaffected by cycloheximide treatment as judged by a high correlation coefficient in expression profiles with and without cycloheximide (**Figure 4C**). The remaining 411 genes (95% of the total) had their expression affected to varying degrees as judged by a reduction in the correlation with their untreated expression pattern (**Figure 4D**). The largest effects were seen in cluster 1, consisting of genes that normally are repressed during regeneration, and in clusters 4 and 5, consisting of genes that normally are upregulated late in regeneration. In the case of cluster 1, there was a general loss of repression when translation was blocked. In the case of clusters 4 and 5, there was a loss of upregulation when translation was blocked.

The results suggest a cascade model in which an "immediate early" set of genes are directly triggered by a pathway that relies entirely on existing proteins, and then one or more of these gene products trigger the rest of the program, possibly by acting as transcription factors. Here we use the term "immediate early" to denote position in a regulatory cascade (**Figure 4B**), but we would also expect that these genes, whose products are required to trigger further changes, would be expressed early in the overall program, since they would not be able to cause changes in expression that take place before they, themselves, are expressed. Consistent with this view, the majority of the genes unaffected by cycloheximide treatment are contained in cluster 2, the earliest expressing cluster, while the loss of expression in cycloheximide treated cells is most dramatic in clusters 4 and 5, the latest-expressing clusters.

Transcriptional regulators expressed during regeneration

The cascade-like organization of regeneration described in **Figure 4** suggests that transcription factors expressed early in the regeneration program (cluster 2) might trigger genes at subsequent stages of the process. A search for transcription factors in the OA-specific cluster 2, which consists of the earliest genes to be upregulated during OA regeneration, revealed the transcription factor E2F, as well as its dimerization partner E2FDP1, and several alleles of Rb, a regulator of the E2F-DP1 interaction. The Rb-E2F-DP1 module plays important roles in regulating cell cycle dependent processes in many species (Bertoli 2014; Nair 2009), including *Tetrahymena* and *Chlamydomonas* (Zhang 2018. Cross 2020). The simultaneous presence of all three members of the Rb-E2F-DP1 module in cluster 2 of the OA specific program make this module a potential candidate regulatory factor for later events in the program.

The temporal pattern of E2F expression is shown in **Figure 5A**, which indicates that E2F is expressed both in sucrose shock responding cells and posterior halves that are regenerating anterior halves, but not in anterior half cells that are regenerating posterior tails. Importantly, E2F is upregulated in OA-regenerating cells even if they are treated with cycloheximide, consistent with a possible role as an immediate early gene that may serve as a regulator of later expression waves. To explore this possibility, we identified putative E2F targets based on promoter motif analysis (Rabinovich 2008), and asked whether these predicted targets exhibited specific expression patterns during regeneration. As shown in **Figure 5B**, there is indeed a tight pattern of E2F targets expressed during a one-hour window that corresponds roughly to the time in regeneration at which centriole related genes are expressed (**Figure 2E**). This pattern closely matches the pattern of expression of the E2F ortholog, designated E2F-1. In cells treated with cycloheximide, 12/13 of these targets fail to be expressed. The majority (10/13) of the E2F targets are in cluster 2, two are in cluster 3, and one is in cluster 5. None of the predicted E2F targets are found in cluster 1 consisting of genes that are repressed during regeneration,

suggesting that E2F likely plays an activating function in this process. The identities of predicted E2F targets among the OA-specific genes are annotated in **Supplemental Table S1**, which indicates that predicted E2F targets do not fall into any single characteristic functional families. However, consistent with the role of E2F in regulating cyclin transcription in other organisms, the target list in *Stentor* does include cyclin and cyclin associated protein A.

A Pumilio ortholog and its targets expressed during regeneration

An important question is to what extent do patterning mechanisms in *Stentor* involve similar molecules or pathways compared to developmental mechanisms in animals, particularly developmental mechanisms that may occur at the one-cell stage. One of the most highly conserved developmental regulators in early animal development is the RNA binding protein Pumilio, which mediates mRNA localization and translation control during pattern formation in animal embryos such as *Drosophila* (Wreden 1997; Gamberi 2002; Sonoda 1999). Our differential expression analysis identified five Pumilio orthologs showing differential gene expression during regeneration, one specific to OA regeneration (cluster 2; SteCoe_27339; **Figure 6A**), one specific to tail regeneration (cluster 2; SteCoe_37495; **Figure 6B**), and 3 specific to the sucrose shock response (cluster 1 SteCoe_9692; cluster 2 SteCoe_16166; cluster 5 SteCoe_22534). The tail-specific Pumilio SteCoe_37495 also shows differential expression during OA regeneration, but with different timing (**Figure 6B**). No Pumilio orthologs were identified as having differential expression in the anterior regeneration, bisection-specific, or general regeneration datasets.

We hypothesized that these Pumilio orthologs may play a role in regulating the localization or translation of other regeneration-specific messages. If this were true, then we would expect the gene expression program of regeneration to include genes whose messages contain Pumilio binding sites. Analysis of Pumilio recognition motifs (Ray 2013) among the set of differentially expressed genes showed that indeed there were differentially expressed genes that contained pumilio recognition sites, and that these were located in the 3' UTR region (**Figure 6C**). Among the OA regeneration specific genes, 34 genes were found to be putative Pumilio targets (see **Supplemental Table S1**), and these distinctly cluster into two groups with peaks of expression at 120 minutes and 300 minutes respectively after the start of regeneration (**Figure 6D**). The timing of these peaks coincides with the peak expression of Pumilio orthologs. For example, the OA-specific Pumilio SteCo_27339 shows peak expression at 120 min, while SteCoe_37495 peaks at 120 min during tail regeneration but 300 min during OA regeneration. Notably, out of the 34 Pumilio target genes showing differential expression, none of them were contained in cluster 1 of the OA regeneration gene set, the cluster that contains genes whose expression decreases during regeneration. Thus, 100%of the Pumilio targets showed increased rather than decreased expression during OA regeneration. All of the predicted Pumilio targets showed reduced induction in cycloheximide treated cells. There was no overlap between the set of predicted Pumilio targets and the set of predicted E2F targets.

Sixteen of the 34 predicted targets showed recognizable homology, and out of these, six were either kinases or phosphatases, potentially suggesting a role for Pumilio in regulating signaling pathways. Another four targets corresponded to basal body associated proteins.

Discussion

Modularity in the regeneration program

A fundamental question has been whether regeneration represents a single process that restores proper morphology following the loss of any part, or instead a collection of distinct, part-specific processes. Our comparative transcriptomic approach in *Stentor* clearly supports the latter view, in that we found regeneration genes could be decomposed into modules specific for different aspects of regeneration. Each module contains temporally distinct waves of expression. Most of the gene expression modules can be interpreted as representing genes encoding proteins required in large quantities for regenerating various cellular structures such as the oral apparatus or the posterior holdfast. However, it is important to recognize that some of the upregulated genes may be involved not in rebuilding lost structures, but in helping the cell to recover from the stress of the perturbation that was used to drive regeneration in the first place. In the case of the bisection-specific module, we interpret these genes as involved in recovery from the surgical wounding (see below). In the case of the sucrose shock response, we expect that some of the genes upregulated during this response but not during bisection may be specific for osmotic stress incurred during sucrose treatment.

Given that the different regeneration paradigms (OA regeneration, anterior regeneration, and posterior regeneration) involve the rebuilding of different structures, and are induced by different stimuli (cutting or sucrose), the function of the general regeneration module genes (**Figure 3C**) is not clear. Since all regeneration events studied here involve rebuilding cortical patterns, we may expect that some of the general regeneration genes have to do with cortical organization. Consistent with this idea, one of the genes upregulated during general regeneration is NPHP4, a ciliary disease gene whose product is involved in linking cilia with cell polarity pathways (Yasunaga 2015). The fact that the general regeneration module contains the fewest genes out of all the modules identified, suggests that regeneration does not represent a single "master" program of expression, but rather a composite of distinct expression modules or subroutines specific for regenerating individual parts of the cell. Somehow, the cell must recognize which part is missing and trigger the appropriate module to restore that part.

Cascade versus production schedule

All regeneration modules showed distinct temporal waves of expression. Several features of the timing of these waves match our *a priori* expectations about the gene expression program of OA regeneration. First, the duration of each gene expression group is roughly one to two hours (**Figure 2A,C**), corresponding to the length of time that regeneration is known to proceed following surgical removal of the nucleus (Tartar 1961). The persistence of regeneration over this time scale likely reflects the lifetime of the mRNAs that drive each stage. This time scale also matched the period of time during which visibly distinct morphological processes occur, for example ciliogenesis initiates in different regions of the oral primordium at slightly different times, with early-stage events of ciliogenesis taking place over a roughly 1-2 hour period (Paulin 1971). Therefore, we expected that groups of related genes would show peaks of expression lasting on the order of 1-2 hours, as we observed. Second, the number and timing of the five clusters correspond with the number and timing of known morphological events, consistent with our *a priori* expectation that different morphological events in regeneration may be coordinated by distinct modules of genes. Finally, we observed a strong correlation between the types of genes expressed at a given stage, and the cell biological events taking place at that stage. Based on this correlation, we believe that examination of other genes with correlated expression

patterns will reveal previously unknown molecular players in organelle regeneration, centriole biogenesis and ciliogenesis.

We can imagine two general schemes by which the timing of these waves could be determined. One model is a "production schedule", in which a master clock triggers successive waves at different times after the initiation of the process. The alternative is a "cascade" model in which the products of early waves of gene expression are required to trigger expression of later waves. In the production schedule type of model, later waves of expression do not depend on the products of earlier genes, and would thus occur normally even in the absence of protein synthesis. The fact that cycloheximide treatment had larger effects on later expressed groups of genes (**Figure 4C,D**) is thus more consistent with a cascade mechanism. Gene cascades are well known in many different systems, such as bacterial sporulation, bacteriophage infection, and insect morphogenesis. Our results are particularly reminiscent of events seen during the response of serum-starved mammalian cells to the re-addition of serum as well as the response of mammalian cells to viral infections. In both cases, a set of co-called immediate early genes are expressed even in the absence of protein synthesis (Lau and Nathans, 1985), and these genes encode factors required for driving further waves of gene expression. The organization of a regeneration response into temporally distinct modules of gene expression is similar that seen in animal regeneration (Monaghan 2007), and these expression patterns appear to be controlled by gene regulatory networks (Smith 2011).

In addition to a genetic cascade, regeneration involves expression of a large number of signaling molecules, in particular kinases. Indeed, kinases are expressed at all stages of regeneration, consistent with the massive expansion of the kinome in *Stentor* (Reiff 2017) although specific kinase families tend to be expressed at specific stages. One abundant class of kinases observed among the upregulated genes was the dual-specificity DYRK kinases, with 13 different DYRK family members expressed (5 in the OA regeneration gene set, 8 in the sucrose shock response gene set). We note, however, that the kinome of *Stentor coeruleus* has been predicted to contain 142 DYRK family members, making them one of the most highly expanded kinase families in the genome (Reiff 2017). Given that the *Stentor* genome contains 35,000 genes, we would expect that roughly 14 DYRK genes would be present in any randomly chosen set of 3,000 genes. The regeneration program does not, therefore, show any particular enrichment for DYRK kinases.

Comparison to other studies

Three previous studies analyzed changes in gene expression during different individual forms of regeneration in *Stentor* (Sood 2017; Onsbring 2018; Wei 2020). Each of these studies addressed just one regeneration paradigm – either OA regeneration (Sood 2017, Wei 2020) or bisection (Onsbring 2018), and therefore did not permit the decomposition of expression patterns into gene modules as was done in the present study.

Onsbring et al. (2018) analyzed transcription in bisected cells of *Stentor polymorphus*, a different *Stentor* species that is somewhat smaller than *S. coeruleus* and contains a green algal endosymbiont. Major findings reported in that study were that the posterior half-cell expressed approximately 9-fold more genes during regeneration than the anterior half-cell. In our results on bisected *S. coeruleus*, we also see a much larger number of genes expressed in the posterior compared to the anterior, although in our case the difference is closer to 5-fold. GO-term analysis of the *S. polymorphus* data showed enrichment for several classes of genes (signaling,

microtubule-based movement, replication, DNA repair, and cell cycle) among the upregulated genes, while a different class of genes were downregulated during regeneration (cellular metabolism and processes related to translation, biogenesis, and the assembly of ribosomes). In broad terms, we see a similar trend in these gene families in bisected *S. coeruleus*. More specific groups of genes found to be highly represented in the *S. polymorphus* data were the DYRK family of kinases and MORN domain proteins. We also find genes in these families to be represented among the upregulated genes during regeneration in *S. coeruleus*, both following bisection and following sucrose shock. In regenerating *S. polymorphus* anterior halves, the most highly expressed gene was reported to encode an ortholog of Lin-54, a DNA-binding regulator of cell cycle related genes (Schmit 2009). In our data, we identified a LIN54 ortholog in the sucrose shock response, but not in regenerating anterior half cells. In regenerating posterior halves, Onsbring (2018) reported a putative E2F/DP family member as highly upregulated, which is consistent with our findings.

Wei et al (2020) analyzed OA regeneration in a strain of *Stentor coeruleus*, designed WHEL, that is distinct from the strain used in our study, and using the chaotropic agent urea to induce OA shedding rather than sucrose shock as was done in our experiments.

Following OA shedding by urea, they observed three clusters of genes that differed in their temporal pattern of gene expression. One cluster consisted of genes that were down-regulated, thus resembling our cluster 1. This group consisted of genes involved in metabolism, mitochondria, and proteolysis. A second cluster consisted of whose expression initially decreased but then peaked at 3-6 hours, similar to our cluster 4. This group contained many heat shock proteins and chaperones (HSP70, HSP90, Cpn60/TCP-1 family) which is different from our finding, in that we picked up these classes of genes as being repressed and not upregulated until the end of regeneration, thus falling into our Cluster 1. Proteomic analysis of the shed OA in their experiments indicated a large component of mitochondrial and ribosomal proteins, suggesting that the urea shock, perhaps due to the chaotropic nature of the chemical, was removing additional cellular contents besides the OA itself. As a result, their transcriptional response may involve not only genes involved in OA regeneration, but also genes involved in recovery from the cellular disruption by urea.

The third cluster reported by Wei et al (2020) consisted of genes peaking at a range of times from 30 min – 9 hours, thus covering the time-periods described by our clusters 2-5. As in our study, they identified several EF hand proteins as well as GAS2. However, in contrast to our results, they only found a single gene annotated as encoding a basal body or cilia-related protein, namely WDR90. The lack of cilia-related genes is consistent with what we observed in our OA-specific module, but differs from what we observed in our sucrose shock response data.

Cell cycle genes expressed during regeneration

One particularly notable class of genes expressed during regeneration are genes encoding cell cycle related proteins. This has also been observed in previous reports (Sood 2017; Onsbring 2018; Wei 2020). Examples include not only general cell cycle regulatory machinery such as cyclins and CDKs, but also mitotic kinetochore proteins.

One hypothesis to explain expression of cell cycle genes during regeneration is that the cell cycle machinery might help regulate the timing of regeneration. The morphological steps of OA regeneration visible on the cell surface (**Figure 1**) are virtually identical to the steps by which a new OA forms during normal cell division (Tartar 1961). Likewise, the macronucleus of *Stentor* undergoes a similar set of morphological changes during both division and regeneration. Like

other ciliates, *Stentor* contains a single large polyploid macronucleus that contains approximately 50,000 copies of the expressed genome as well as several smaller diploid micronuclei. During division, the macronucleus is simply pinched in half by the cleavage furrow. Prior to this pinching, the elongated macronucleus shortens and compacts into a more spheroidal shape, which then re-elongates just before cytokinesis. These same shape changes occur during regeneration, even though the cell is not going to divide (Paulin and Brooks 1975). The strong morphological similarity between regeneration and division, at both the cortical and nuclear level, together with the expression of cell cycle and mitosis related genes during regeneration, suggests the possibility that OA regeneration might involve co-option of parts of the cell cycle machinery to regulate the timing of events. This potential connection between regeneration and cell division in *Stentor* highlights a classical question in the biology of regeneration: is regeneration a distinct process in its own right, or instead does it reflect a re-activation of development? In the case of the unicellular *Stentor*, for which development is equivalent to cell division, the use of cell cycle machinery in regeneration would support the latter view.

However, an alternative explanation for the upregulation of cell cycle genes during regeneration could be the fact that in *Stentor*, the micronuclei undergo mitosis during regeneration (Guttes 1959). Expression of cell cycle and mitosis genes during regeneration might thus simply happen in order to allow for mitosis of the micronuclei, and have nothing to do with the events of regeneration itself. Consistent with this alternative hypothesis, we find that cell cycle genes are not seen among the genes upregulated during tail regeneration, during which micronuclear mitosis does not occur (Guttes 1959).

The two explanations need not be mutually exclusive – it is possible that the program of OA regeneration is under control of cell cycle timing machinery that also directs the events of micronuclear mitosis. In keeping with this idea, it has been shown that inhibition of the cell-cycle related Aurora kinases, several orthologs of which show differential expression during OA regeneration (clusters 2, 4, and 5 of the sucrose shock response) can advance or delay the later stages of regeneration (Lin 2020).

We note that nuclear shape change only happens in the posterior half and not the anterior half cells following bisection (Weisz 1949). Similarly, mitosis only happens during regeneration in the posterior half and not the anterior half (Guttes 1959). It is therefore notable that there appear to be far fewer nucleus and cell cycle related genes in the tail-regeneration, bisection, or general regeneration modules.

Implications for OA assembly

The OA is a complex cellular structure consisting of a tightly packed and highly ordered array of basal bodies and their associated cilia in the form of a membranellar band. We therefore would expect *a priori* that among the list of OA specific regeneration genes will be genes encoding protein components of centrioles/basal bodies, and that the expression of such genes should coincide with the time period in which the basal bodies are forming. Indeed, the OA specific module does contain many known centriole biogenesis factors, and by far the majority of centriole-biogenesis genes are expressed in cluster 2 of the OA-specific pathway, which coincides with the period of basal body synthesis during OA regeneration. In the sucrose shock response pathway, we see the same trend of centriole related genes expressed during cluster 2, and we also see cilia genes expressed in later clusters, coinciding with the timing of ciliogenesis and establishment of ciliary motility (**Figure 2E**). We thus find strong temporal correlation between the expression of centriole and cilia related genes and the corresponding events of basal

body biogenesis and ciliary assembly, respectively. Based on this positive correlation with known genes involved in centrioles and cilia, we predict that at least some of the genes in these clusters with no or poor homology to known genes may encode undiscovered factors involved in centriole biogenesis and ciliogenesis. In particular, while the proteome of the centriole/basal body has become increasingly well characterized (Keller 2005; Jakobsen 2011; Lauwaet 2011; Firat-Karalar 2014; Hamel 2017), we hypothesize that cluster 2 may contain genes whose products are needed for basal body assembly, or for positioning the OA in relation to other cellular structures, but may not encode structural components of the basal body itself. Such proteins would have been missed in proteomic analyses of the final structure.

Interestingly, while many centriole / basal body related genes are expressed in the early clusters (2 and 3), the only clear centriole-related genes we find in the late-expressed cluster 4 is an ortholog of LRRC45 which is a linker component required for centriole cohesion (He 2013). In this regard, we note that when basal bodies first assemble during oral regeneration, they do so with random orientations relative to each other, creating a so-called "anarchic field" (Bernard 1981). It is only later in the process that the basal bodies associate into pairs and then larger groups to form the membranelles that are the dominant ultrastructural motif of the oral apparatus. The expression of LRRC45 at exactly this stage suggests that this linker may be a key element for assembling the membranellar band from the initially randomly oriented basal bodies.

While we have noted that the OA-specific module has relatively few genes encoding ciliary proteins compared to the sucrose shock response dataset, it does contain a number of orthologs of genes implicated in the ciliopathies Meckel Syndrome and Joubert Syndrome. Both of these syndromes involve defects in non-motile cilia, and the proteins encoded by the Meckel and Joubert syndrome genes are involved in gating the import of proteins into the cilium (Takao 2016). The sucrose-shock specific program also included two Meckel/Joubert syndrome genes (B9D1 in cluster 2 and NPHP3 in cluster 4). We hypothesize that these genes may be expressed during regeneration in order to equip the newly formed basal bodies with appropriate protein machinery to generate OA-specific cilia.

In addition to basal bodies and cilia, the OA is also known to contain a set of protein fibers made of centrin-like EF hand proteins (Huang and Pitelka, 1973). The OA-specific module contains several predicted EF-hand proteins including at least one ortholog of conventional centrin (see **Supplemental table S1**).

Finally, we note that the late expressing clusters contain orthologs of GAS2, a protein involved in coupling actin and microtubule cytoskeleton in other organisms. GAS2 has been found in the membranellar band proteome (Wei 2020). Our results further support involvement of GAS2 in OA assembly, and suggest it functions at a late stage in the regeneration process. We hypothesize that GAS2 may play a role in aligning the membranelles with the longitudinal microtubule bundles that define the ciliary rows on the cell body.

Re-use versus new synthesis of ciliary proteins

The OA is an assemblage of motile cilia. The cilia of the OA are substantially longer and more densely packed than the body wall cilia. Upregulation of cilia-related genes during ciliogenesis has been reported in other ciliates, including *Tetrahymena* (Soares 1991) and *Paramecium* (Kandl 1995), but it also seen in green algae (Schloss 1984; Stolc 2005), lower plants (Tomei 2016), and vertebrate multiciliated epithelial cells (Ross 2007; Hoh 2012). Given the fact that cilia represent the most visibly obvious structure within the OA, it was expected that regeneration of the OA would be accompanied by upregulation of genes encoding ciliary

proteins. Consistent with this expectation, the sucrose-shock specific genes (**Figure 2C**) include a large number of genes encoding protein components of motile cilia, including radial spokes and the dynein regulatory complex. These genes are present mainly in clusters 3 and 4, and reach peak expression at 200 min (**Figure 2E**). It was therefore somewhat surprising to observe that genes encoding proteins specific to motile cilia, such as dynein arms or radial spoke proteins, are for the most part not found among the OA-specific genes (**Figure 2A**), indicating that these genes are expressed when the OA forms in sucrose shocked cells, but not when the OA forms in posterior half-cells that are regenerating a new anterior half.

Given that the regenerating OA in posterior half-cells needs to be equipped with cilia, how is it possible that the genes encoding ciliary proteins are not, by in large, upregulated during regeneration? One possibility is that in these cells, assembly of motile cilia onto the basal bodies is carried out using protein obtained from either a pre-existing cytoplasmic pool or else from the pre-existing cilia on the cell body. Schmähl (1926) has reported that in the giant ciliate *Bursaria*, some body cilia shorten while others are growing, suggesting an ability to redistribute protein between old and new structures. In the green alga *Chlamydomonas*, severed flagella are able to regenerate using protein from a cytoplasmic pool, but, importantly, they can also "borrow" protein from other flagella present on the same cell, with those other flagella shortening as a result (Coyne 1970). As to why OA assembly in sucrose-shock cells involves expression of cilia-related genes while OA assembly in regenerating posterior half-cells does not, we speculate that scaling of organelle size may be involved. As Morgan (1901) has pointed out, when a cell is bisected and the posterior half forms an OA, the membranellar band of the new OA is half the size of that in an intact cell. In contrast, a sucrose-shocked cell has to build a full-sized OA. It is therefore possible that the larger size of the OA being formed after sucrose shock requires synthesis of new protein, existing pools being insufficient. On the other hand, a bisected cell would start out with only half as much protein as an intact cell, so it is unclear if this type of scaling argument can really explain the differences that are seen in the transcriptional program. Clearly, direct assays for protein re-utilization will be needed to answer this question in the future.

Implications for cellular wound recovery

The bisection-specific genes (**Figure 3B**) are shared between half-cells that are regenerating distinct structures. What the two half-cells have in common is that they were mechanically wounded during the surgical bisection, unlike the sucrose shocked cells. We therefore interpret the bisection-specific differential gene expression pattern as reflecting a response to physical wounding of the cell. Direct measurements of membrane integrity in *Stentor* indicate that the plasma membrane seals itself on a time-scale of 100-1000 sec after wounding (Zhang 2021). In comparison, the upregulated genes in the bisection response peak at 120-240 minutes after wounding, long after the wound itself has been closed. Thus, the bisection-specific transcriptional module is more likely to reflect recovery of cell state after wound closure, rather than the wound closure process itself.

Two of the upregulated genes in the bisection-specific module encode ammonium transporters, and a third encodes carbonic anhydrase. Both ammonium transporters and carbonic anhydrase are predicted to alter intracellular pH. In animal cells, intracellular pH is generally acidic, and becomes more basic when cells are wounded (Chambers 1961). Since Stentor grows in pond water that is relatively acidic, carbonic anhydrase may act to increase the pH of

cytoplasm after wounding by reducing bicarbonate to CO2. In any case we infer that a key function of the bisection-specific transcriptional module is restoration of proper intracellular pH once the membrane rupture has been healed. Similarly, the expression of both an ABC transporter and a Major Facilitator superfamily member, both of which are involved in transporting a wide range of small molecules out of cells (Wong 2014; Quistgaard 2016) may indicate a role in pumping contaminants out of the cell that may have entered through the open wound.

Role of conserved developmental regulators in single-cell regeneration

Giant, complex cells like *Stentor* face many of the same morphogenetic challenges as animal embryos in the need for establishing body axes, creating and maintaining patterns, and ensuring that anatomical features are present in the correct positions (Marshall 2020). It is usually assumed that the similarity of such processes as axiation, regeneration, or induction, between unicellular protists and animal embryos must reflect analogous processes that are implemented using completely different, non-homologous mechanisms. The identification of Pumilio, a highly conserved developmental regulator (Wreden 1997; Gamberi 2002; Sonoda 1999) in the *Stentor* regeneration program, suggests that there may in fact be conserved molecular mechanisms at work during morphogenesis in both single celled and multi-celled organisms.

Although multiple pumilio orthologs were among the sucrose-shock specific gene list, this list did not contain any additional Pumilio targets besides those identified in the OA-specific list. Likewise, although the tail regeneration program contains pumilio orthologs, it does not contain any predicted Pumilio target genes. One potential explanation is that some of the relevant Pumilio targets in these other programs may be genes whose transcripts are already present prior to initiation of regeneration. In such cases, expression of Pumilio orthologs may alter the localization or translation of those targets, even if they show no change at the transcriptional level.

Conclusion

The transcriptional analysis *of Stentor* regeneration described here begins to reveal key molecular details of intracellular patterning and regeneration mechanisms, such as evidence for modularity and a cascade organization. We find that *Stentor* regeneration involves expression of regulatory genes conserved across eukaryotes, suggesting a deep conservation of developmental mechanisms.

Acknowledgments

This work was supported by an American Cancer Society postdoctoral fellowship (PS), an NSF predoctoral fellowship (AL), and NIH grant R35 GM130327 (WFM). WFM is a Chan Zuckerberg Biohub investigator. Library preparation and QC for sequencing was conducted by the Gladstone Institute Genomics Core. The authors thank Robert Stroud, Hiten Madhani, and members of the Marshall lab for many helpful discussions, and Ulises Diaz, Ambika Nadkarni, Vincent Boudreau, Sindy Tang, and Mark Slabodnick for helpful comments on the manuscript.

Table 1. List of centriole/Basal body genes in the OA-specific module (Figure 2A)

	Cluster of peak expression				
	1	2	3	4	5
SAS6		+			
STIL		+			
POC5		+			
POC11/ CCD77		+			
POC18 / WDR67 / Tbc31		+			
VFL3		+			
GCP2		+			
GCP4		+			
Jouberin		+			
MKS1		+			
MKS6		+			
XRP2/TBCC		+			
RTTN		+			
CCD61		+			
Bld10/CEP135		+	+		
CEP350		+	+		
POC16		+	+		
Centrin 2		+	+		
CEP44			+		
POC12			+		
Centrin 3			+		
MKS3			+		
LRRC45				+	

References

Albee AJ, Kwan AL, Lin H, Granas D, Stormo GD, Dutcher SK. 2013. Identification of cilia genes that affect cell-cycle progression using whole-genome transcriptome analysis in Chlamydomonas reinhardtii. *G3 (Bethesda)* 3, 979–91

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215, 403–10.

Angelo S Mao DJM. 2015. Regenerative medicine: Current therapies and future directions. *Proc Natl Acad Sci USA* 112, 14452–9.

Aufderheide, K.J., Frankel, J., Williams, N.E. 1980. Formation and positioning of surface-related structures in protozoa. *Microbiol. Rev.* 44, 252-302

Azimzadeh J, Wong ML, Downhour DM, Sánchez Alvarado A, Marshall WF. 2012. Centrosome loss in the evolution of planarians. *Science* 335, 461–3.

Balestra FR, Strnad P, Flückiger I, Gönczy P. 2013. Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev Cell* 25, 555–71.

Bell SP, Labib K. 2016. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* 203, 1027-67.

Bernard F, Bohatier J. 1981. Ultrastructure et mise en place des organelles buccaux au cours de la régénération orale chez Stentor coeruleus (Cilié Hétérotriche). *Can. J. Zool.* 59, 2306-2318.

Bertoli C, Skotheim JM, de Bruin RAM. 2013. Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology* 14, 518–28.

Bettencourt-Dias M, Glover DM. 2007. Centrosome biogenesis and function: centrosomics brings new understanding. *Nature Reviews Molecular Cell Biology* 8, 451–63.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–20.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34, 525–7.

Burchill BR. 1968. Synthesis of RNA and protein in relation to oral regeneration in the ciliate Stentor coeruleus. *J. Exp. Zool.* 167, 427-438.

Chambers R, Chambers EL. 1961. *Explorations into the nature of the living cell.* Harvard University Press, Cambridge MA. 352 pp.

Cong X, Hubmayr RD, Li C, Zhao X. 2017. Plasma membrane wounding and repair in pulmonary disease. *Am. J. Physiol. Lung Cell Mol. Physiol.* 312, L371-L391.

Coyne B, Rosenbaum JL. 1970. Flagellar elongation and shortening in Chlamydomonas. II. Re-utilization of flagellar protein. *J. Cell Bio*l. 47, 777-781.

Cross FR. 2020. Regulation of multiple fission and cell-cycle-dependent gene expression by CDKA1 and the Rb-E2F pathway in Chlamydomonas. *Curr. Biol.* 30, 1855-1865.

Ellwood LC, Cowden RR. 1966. RNA Metabolism during Regeneration in Stentor coeruleus. *Cytologia* 31,80–8.

Frankel J. 1989. *Pattern formation. Ciliate studies and models.* Oxford University Press, New York. 314 pp

Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol* 33,243–6.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44, D279–85.

Firat-Karalar EN, Sante J, Elliott S, Stearns T. 2014. Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J Cell Sci.* 127, 4128-33

Furrer DI, Swart EC, Kraft MF, Sandoval PY, Nowacki M. 2017. Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep* 20, 505–20.

Galeote VA, Alexandre H, Bach B, Delobel P, Dequin S, Blondin B. 2007. Sfl1p acts as an activator of the HSP30 gene in Saccharomyces cerevisiae. *Curr Genet* 52, 55–63.

Gamberi C, Peterson DS, He L, Gottlieb E. 2002. An anterior function for the Drosophila posterior determinant Pumilio. *Development* 129, 2699–710.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–52.

Guertin MJ, Lis JT. 2010. Chromatin Landscape Dictates HSF Binding to Target DNA Elements. *PLoS Genet* 6, e1001114.

Guttes E, Guttes S. 1959 Regulation of mitosis in Stentor coeruleus. *Science* 129, 1483

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494–512.

Hamel V, Steib E, Hamelin R, Armand F, Borgers S, Flückiger I, Busso C, Olieric N, Sorzano COS, Steinmetz MO, Guichard P, Gönczy P. 2017. Identification of Chlamydomonas central core centriolar proteins reveals a role for human WDR90 in ciliogenesis. *Curr Biol*. 27, 2486-2498

Harold FM. 2005. Molecules into cells: specifying spatial architecture. *Microbiol. Mol. Biol.* Rev. 69, 544-564.

He R, Huang N, Bao Y, Zhou H, Teng J, Chen J. 2013. LRRC45 is a centrosome linker component required for centrosome cohesion. *Cell Rep* 4,1100–7.

Hoh RA, Stowe TR, Turk E, Stearns T. 2012. Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. *PLoS One* 7, e52166

Huang B, Pitelka Dr. 1973. The contractile process in the ciliate, Stentor coeruleus. I. The role of microtubules and filaments. *J. Cell Biol*. 57, 704-28.

Ishikawa H, Marshall WF. 2011. Ciliogenesis: building the cell's antenna. *Nature Reviews Molecular Cell Biology* 12, 222–34.

Jakobsen L, Vanselow K, Skogs M, Toyoda Y, Lundberg E, Poser I, Falkenby LG, Bennetzen M, Westendorf J, Nigg EA, Uhlen M, Hyman AA, Andersen JS. 2011. Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J*. 30, 1520-35.

James EA. 1967. Regeneration and division in Stentor coeruleus: the effects of microinjected and externally applied actinomycin D and puromycin. *Dev Biol* 16,577–93.

Kandl KA, Forney JD, Asai DJ. 1995. The dynein genes of Paramecium tetraurelia: the structure and expression of the ciliary beta and cytoplasmic heavy chains. *Mol. Biol. Cell* 6, 1549-62.

Kaufman L, Rousseeuw PJ, editors. *Finding Groups in Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc; 1990.

Keller, L.C., Romijn, E.P., Zamora, I., Yates, J.R. 3rd., and Marshall, W.F. 2005. Proteomic analysis of isolated Chlamydomonas centrioles reveals orthologs of ciliary disease genes. *Curr. Biol*. 15,1090-8.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14, R36.

Kim JJ, Lipatova Z, Majumdar U, Segev N. 2016. Regulation of Golgi Cisternal Progression by Ypt/Rab GTPases. *Dev Cell* 36, 440–52.

Kirschner M, Gerhart J, Mitchison T. 2000. Molecular "vitalism". *Cell* 100, 79–88.

Lau LF, Nathans D. 1985. Identification of a set of genes expressed during the G0/G1 transition of cultured mouse cells. EMBO J. 4, 3145-51.

Lauwaet T, Smith AJ, Reiner DS, Romijn EP, Wong CC, Davids BJ, Shah SA, Yates JR 3rd, Gillin FD. 2011. Mining the giardia genome and proteome for conserved and unique basal body proteins. *Int J Parasitol*. 41, 1079-92

Leidel S, Delattre M, Cerutti L, Baumer K, Gönczy P. 2005. SAS-6 defines a protein family required for centrosome duplication in *C.elegans* and in human cells. *Nat Cell Biol* 7, 115-125.

Lin A, Makushok T, Diaz U, Marshall WF. 2018. Methods for the study of regeneration in Stentor. *J. Vis. Exp*. 136, e57759

Lin A, Summers D, Reiff S, Tipton AR, Tang SKY, Marshall WF. 2020. Aurora kinase inhibitors delay regeneration in Stentor coeruleus at an intermediate step. *Science Matters* https://sciencematters.io/articles/202003000006

Liu Y-C, Chen H-C, Wu N-Y, Cheng S-C. 2007. A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol Cell Biol* 27, 5403–13.

Maloney M, McDaniel W, Locknar S, Torlina H. 2005. Identification and localization of a protein immunologically related to caltractin (centrin) in the myonemes and membranelles of the heterotrich ciliate Stentor coeruleus. *Journal of Eukaryotic Microbiology, 52*, 328-338.

Marshall WF. 2020. Pattern formation and complexity in single cells. *Curr. Biol*. 30, R544-R552.

Matthew L, Bochman AS. 2009. The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. *Microbiol Mol Biol Rev* 73, 652–83.

Monaghan JR, Walker JA, Page RB, Putta S, Beachy CK, Voss SR. 2007. Early gene expression during natural spinal cord regeneration in the salamander Ambystoma mexicanum. *J. Neurochem*. 101, 27-40.

Morgan TH. 1901a. *Regeneration*. Macmillan 316 pp.

Morgan TH. 1901b. Regeneration of proportionate structures in Stentor. *Biol. Bull*. 2, 311-328.

Nair JS, Ho AL, Tse AN, Coward J, Cheema H, Ambrosini G, et al. 2009. Aurora B Kinase Regulates the Postmitotic Endoreduplication Checkpoint via Phosphorylation of the Retinoblastoma Protein at Serine 780. *Mol Biol Cell* 20, 2218–28.

Newman E. 1972. Contraction in Stentor coeruleus: a cinematic analysis. *Science* 177, 447-449.

Oeckler RA, Hubmayr RD. 2008. Cell wounding and repair in ventilator injured lungs. *Respir Physiol Neurobiol* 163, 44–53.

Onsbring H, Jamy M, Ettema T. 2018. RNA sequencing of Stentor cell fragments reveals transcriptional changes during cellular regeneration. *Curr. Biol.* 28, 1281-1288.

Paulin JJ, Brooks AS. 1975. Macronuclear differentiation during oral regeneration in Stentor coeruleus. *J Cell Sci* 19, 531–41.

Paulin JJ, Bussey J. 1971. Oral regeneration in the ciliate Stentor coeruleus: a scanning and transmission electron optical study. *J Protozool* 18,201–13.

Pellegrini L. 2012. The Pol-Primase Complex. In *The Eukaryotic Replisome: a Guide to Protein Structure and Function*, vol. 62, Dordrecht: Springer, Dordrecht; pp. 157–69.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* 14, 687–90.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47,11.12.1–34.

Quistgaard EM, Low C, Guettou F, Nordlund P. 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Cell Mol. Biol.* 17, 123-32.

Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ. 2008. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res* 18, 1763–77.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–7.

Reiff SB, Marshall WF. 2017. A large kinome in a large cell: Stentor coeruleus possesses highly expanded kinase families and novel domain architectures. *bioRxiv* 168187

Remillard SP, Witman GB. 1982. Synthesis, transport, and utilization of specific flagellar proteins during flagellar regeneration in Chlamydomonas. *J Cell Biol* 93, 615–31.

Ren Y, Seo HS, Blobel G, Hoelz A. 2010. Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. *Proc Natl Acad Sci USA* 107, 10406–11.

Ross AJ, Dailey LA, Brighton LE, Devlin RB. 2007. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am. J. Resp. Cell Mol. Biol.* 37, 169-185.

Ross RJ, Weiner MM, Lin H. 2014. PIWI proteins and PIWI-interacting RNAs in the soma. Nature 505, 353–9.

Schloss JA, Silflow CD, Rosenbaum JL. 1984. mRNA abundance changes during flagellar regeneration in Chlamydomonas reinhardtii. *Mol. Cell Biol.* 4, 424-434.

Schmähl O. 1926. Die Neubildung des Peristoms bei der Teilung von Bursaria truncatella. *Arch. Protistenkunde* 54, 359-430

Schmit F, Cremer S, Gaubatz S. 2009. LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdcd2 promoter in a sequence-specific manner. FEBS J. 276, 5703-5716.

Shulman JM, St Johnston D. 1999. Pattern formation in single cells. *Trends Cell Biol.* 9, M60-64.

Slabodnick MM, Prevo B, Gross P, Sheung J, Marshall W. 2013. Visualizing cytoplasmic flow during single-cell wound healing in Stentor coeruleus. *J Vis Exp* 82, e50848

Slabodnick MM, Marshall WF. 2014a. Stentor coeruleus. *Curr. Biol.* 24, R783-784.

Slabodnick MM, Ruby JG, Dunn JG, Feldman JL, DeRisi JL, Marshall WF. 2014b. The kinase regulator mob1 acts as a patterning protein for Stentor morphogenesis. *PLoS Biol* 12, e1001861.

Slabodnick MM, Ruby JG, Reiff SB, Swart EC, Gosai S, Prabakaran S, Witkowska E, Larue GE, Fisher S, Freeman RM, Gunawardena J, Chu W, Stover NA, Gregory BD, Nowacki M, Derisi J, Roy SW, Marshall WF, Sood P. 2017. The macronuclear genome of Stentor coeruleus reveals tiny introns in a giant cell. *Curr. Biol.* 27, 569-75.

Smith EF, Yang P. 2004. The radial spokes and central apparatus: mechano-chemical transducers that regulate flagellar motility. *Cell Motil. Cytoskel.* 57, 8-17.

Smith J, Morgan JR, Zottoli SJ, Smith PJ, Buxbaum JD, Bloom OE. 2011. Regeneration in the era of functional genomes and gene network analysis. *Biol. Bull.* 221, 18-34.

Soares H, Cyrne L, Barahona I, Rodrigues-Pousada C. 1991. Different patterns of expression of beta-tubulin genes in Tetrahymena pyriformis during reciliation. *Eur. J. Biochem.* 197, 291-9.

Sonoda J, Wharton RP. 1999. Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes Dev* 13, 2704–12.

Sood P, McGillivary R, Marshall WF. 2017. The transcriptional program of regeneration in the giant single cell, Stentor coeruleus. *bioRxiv*. preprint 240788 doi: 10.1101/240788

Stevens NM. 1903. Notes on regeneration in Stentor coeruleus. *Archiv f. Entwickelungsmechanik* 16, 461-475.

Stolc V, Samanta MP, Tongprasit W, Marshall WF. 2005. Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. *Proc Natl Acad Sci USA* 102, 3703–7.

Takao D, Verhey KJ. 2016. Gated entry into the ciliary compartment. *Cell Mol. Life Sci*. 73, 119-127.

Tang SKY, Marshall WF. 2017. Self-repairing cells: How single cells heal membrane ruptures and restore lost structures. *Science* 356, 1022–5.

Tartar V. *The biology of Stentor*. 1961. Pergammon Press, Oxford.

Tartar V 1956. Pattern and substance in Stentor. In *Cellular Mechanisms in Differentiation and Growth*, D. Rudnick ed. Princeton University Press, Princeton NJ pgs 73-100.

Tartar V. 1957, Reactions of Stentor coeruleus to certain substances added to the medium. *Experimental Cell Research* 13,317–32.

Tomei EJ, Wolniak SM. 2016. Transcription analysis reveals a diverse family of kinesins essential for spermatogenesis in the fern Marsilea. *Cytoskeleton* 73, 145-59.

UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45, D158–69.

Viswanadha R, Sale WS, Porter ME. 2017. Ciliary motility: regulation of axonemal dynein motors. *Cold Spring Harbor Perspect. Biol*. 9, a018325

Wan KY, Hurlimann SK, Fenix AM, McGillivary RM, Makushok T, Burns E, Sheung JY, Marshall WF. 2020. Reorganization of complex ciliary flows around regenerating Stentor coeruleus. *Phil. Trans. Roy Soc. London B* 375, 20190167.

Wang X, Xie W, Zhang Y, Lin P, Han L, Han P, et al. 2010. Cardioprotection of ischemia/reperfusion injury by cholesterol-dependent MG53-mediated membrane repair. *Circ Re*s 107, 76–83.

Wei W, Jiang C, Yang W, Miao W, Xiong J. 2020. Proteomic identification and expression of oral apparatus constituents in cell regeneration of giant ciliate Stentor coeruleus (strain WHEL). *Gene* 743, 144624

Weisz PB. 1949. A cytochemical and cytological study of differentiation in normal and

reorganizational stages of Stentor coeruleus. *J. Morph.* 84, 335-363

Weisz PB. 1951. An experimental analysis of morphogenesis in Stentor coeruleus. *J. Exp. Zool.* 116, 231-257

Whitson GL. 1965. The effects of actinomycin D and ribonuclease on oral regeneration in Stentor coeruleus. *J Exp Zool* 160, 207–14.

Wong K, Ma J, Rothnie A, Biggin PC, Kerr ID. 2014. Towards understanding promiscuity in multidrug efflux pumps. *Trends Biochem. Sci.* 39, 8-16.

Wreden C, Verrotti AC, Schisa JA, Lieberfarb ME, Strickland S. 1997. Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. *Development* 124, 3015–23.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–75.

Yang P, Diener DR, Yang C, Kohno T, Pazour GJ, Dienes JM, et al. 23006. Radial spoke proteins of Chlamydomonas flagella. *J Cell Sci* 119, 1165–74.

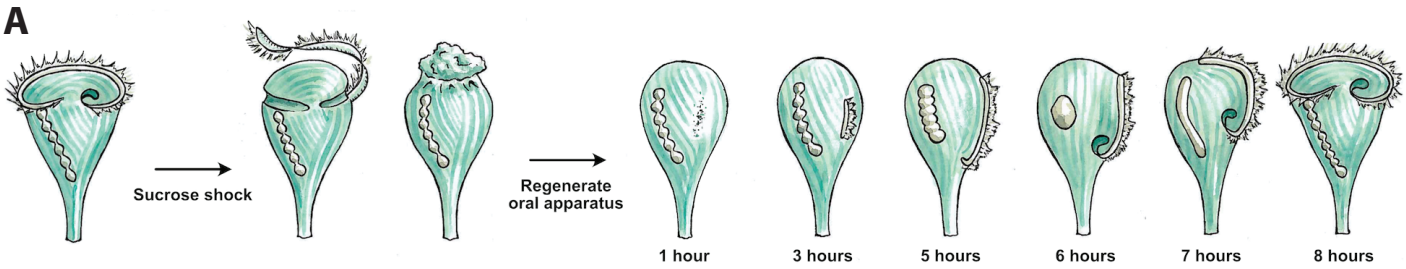
Yasunaga T, Hoff S, Schell C, Helmstädter M, Kretz O, Kuechlin S, Yakulov TA, Engel C, Müller B, Bensch R, Ronneberger O, Huber TB, Lienkamp SS, Walz G. 2015. The polarity protein inturned links NPHP4 to Daam1 to control the subapical actin network in multiciliated cells. *J. Cell Biol.* 211, 963-73.

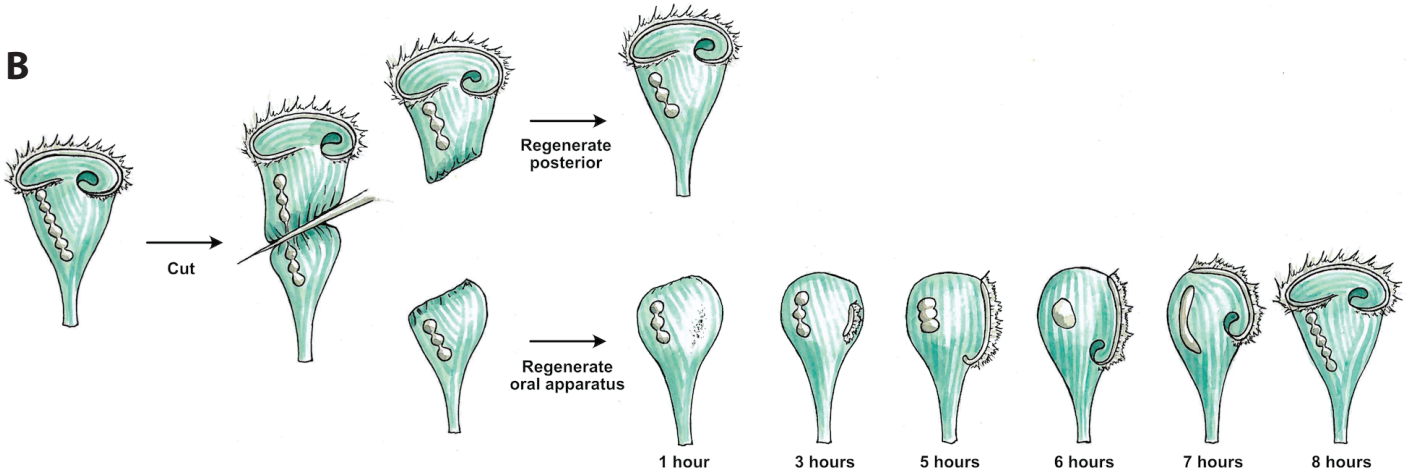
Younger KB, Banerjee S, Kelleher JK, Winston M, Margulis L. 1972. Evidence that the synchronized production of new basal bodies is not associated with DNA synthesis in Stentor coeruleus. *J. Cell Sci.* 11,621–37.

Zhang J, Yan G, Tian M, Ma Y, Xiong J, Miao W. 2018. A DP-like transcription factor protein interacts with E2fl1 to regulate meiosis in Tetrahymena thermophila. *Cell Cycle* 17, 634-642

Zhang KS, Blauch LR, Huang W, Marshall WF, Tang SKY. 2021. Microfluidic guillotine reveals multiple timescales and mechanical modes of wound response in Stentor coeruleus. *BMC Biology* 19, 63.

Zhu X, Liu Y, Yang P. 2017. Radial spokes - a snapshot of the motility regulation, assembly, and evolution of cilia and flagella. *Cold Spring Harbor Perspect. Biol.* 9, a028126


A
Sucrose shock
Regenerate
oral apparatus
1 hour
3 hours
5 hours
6 hours
7 hours
8 hours


B
Cut
Regenerate
posterior
Regenerate
oral apparatus
1 hour
3 hours
5 hours
6 hours
7 hours
8 hours

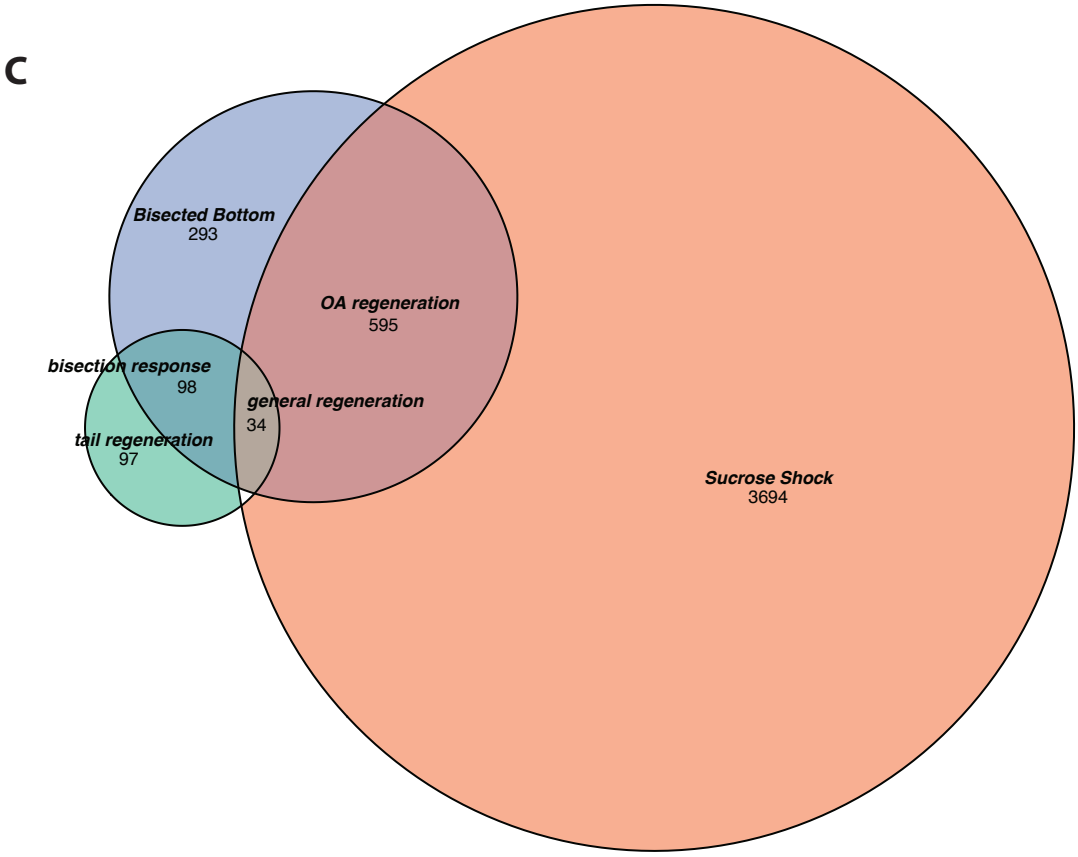

C
Bisected Bottom
293
OA regeneration
595
bisection response
98
general regeneration
34
tail regeneration
97
Sucrose Shock
3694

Figure 1. Transcriptional profiling of Stentor regeneration A. Morphological events in *Stentor* regeneration. At t=0,the membranellar band is shed during sucrose shock. The body cilia remain on the cell. After sucrose shock the frontal field protrudes, resulting in the anterior end of *Stentor* becoming rounded rather than cone-shaped. One hour after the start of regeneration, basal bodies begin to form at the locus of stripe contrast. After three hours, the first cilia of the new membranellar band are visible. These cilia show uncoordinated beating. After five hours, the new membranellar band elongates and extends along the anterior-posterior axis. A site for the new mouthparts is cleared at the posterior end of the membranellar band. During this stage, the cilia become oriented with respect to each other and their beating begins to become coordinated, forming multiple short metachronal waves. The nodes of the macronucleus begin to condense. By six hours, the mouthparts are completely formed and the macronucleus is fully condensed. At seven hours the membranellar band and mouth migrate to the anterior end of the cell. The macronucleus extends into a sausage-like shape. By eight hours after sucrose shock the *Stentor* is fully regenerated. The membranellar band completely wraps around the anterior of the cell, all of the oral cilia coordinate to form a single metachronal wave, the macronucleus is re-nodulated, and the cell resumes normal feeding activity. **B.** Surgical bisection of *Stentor*. When a *Stentor* cell is cut in half perpendicular to the long axis, two cell fragments are produced, an anterior half-cell and a posterior half-cell. Immediately after cutting, both half-cells heal their wounded edges. The anterior half-cell then regenerates a new posterior body including the hold-fast, and the posterior half-cell regenerates a new anterior body including the Oral Apparatus. Oral regeneration in the posterior half-cell has the same general morphological events and timing as oral regeneration in the sucrose shocked cells. Both fragments are able to regenerate because the elongated macronucleus that is divided into both halves during surgery is highly polyploid, ensuring that each half-cell retains many copies of the genome. **C.** Comparative transcriptional profiling. We performed RNAseq on sucrose shocked cells regenerating as in panel A, as well as on both the anterior and posterior half-cells regenerating after bisection as in panel B. Genes were grouped into modules according to correlated expression patterns shared between two or more fragments. For example, the "general regeneration" module was defined based on genes showing differential expression in all three cases of regeneration (OA, anterior half, and posterior half).

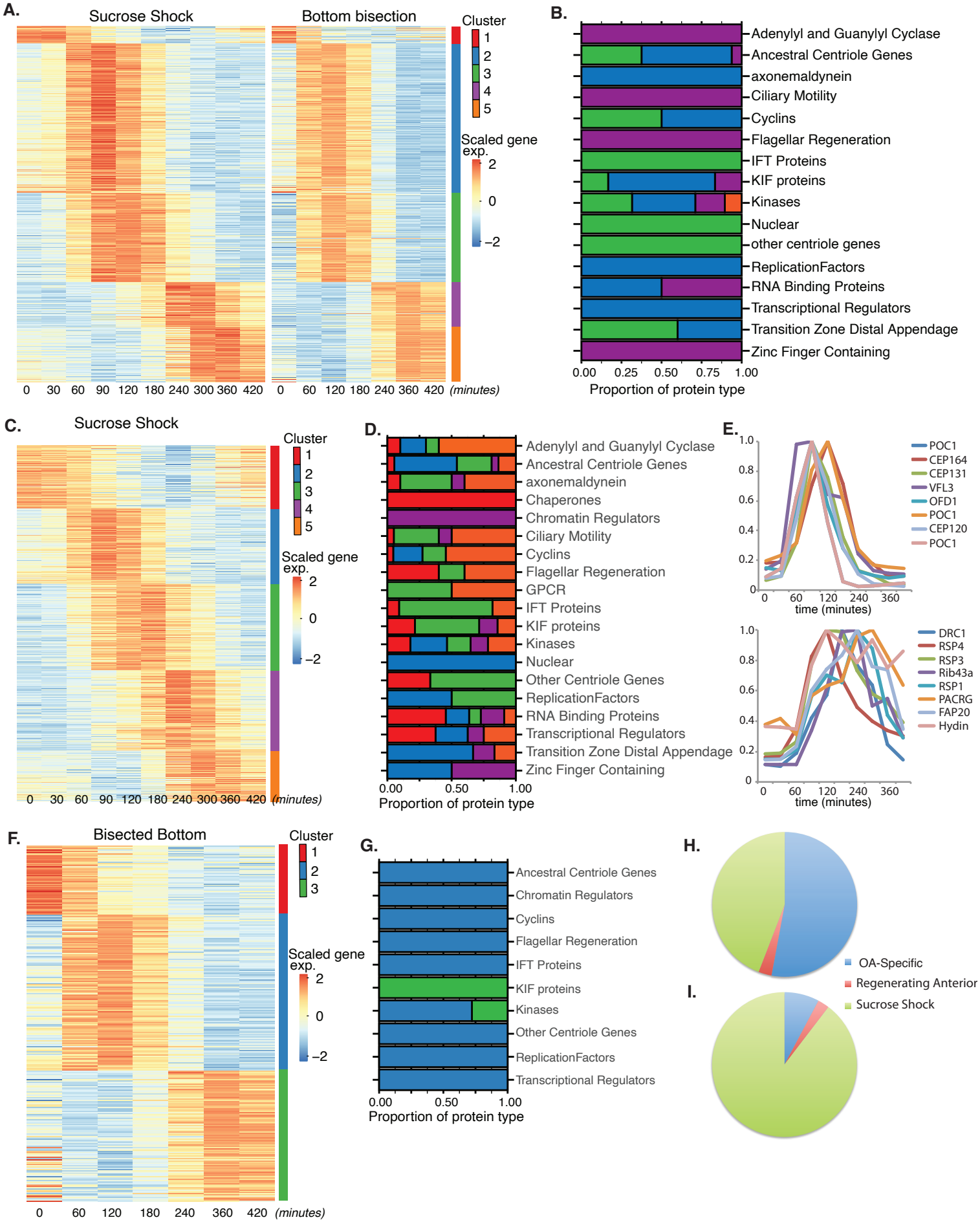

A.
Sucrose Shock
Bottom bisection
Cluster
1
2
3
4
5
Scaled gene exp.
2
0
−2
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420 (minutes)
B.
Adenylyl and Guanylyl Cyclase
Ancestral Centriole Genes
axonemaldynein
Ciliary Motility
Cyclins
Flagellar Regeneration
IFT Proteins
KIF proteins
Kinases
Nuclear
other centriole genes
ReplicationFactors
RNA Binding Proteins
Transcriptional Regulators
Transition Zone Distal Appendage
Zinc Finger Containing
0.00 0.25 0.50 0.75 1.00
Proportion of protein type
C.
Sucrose Shock
Cluster
1
2
3
4
5
Scaled gene exp.
2
0
−2
0 30 60 90 120 180 240 300 360 420 (minutes)
D.
Adenylyl and Guanylyl Cyclase
Ancestral Centriole Genes
axonemaldynein
Chaperones
Chromatin Regulators
Ciliary Motility
Cyclins
Flagellar Regeneration
GPCR
IFT Proteins
KIF proteins
Kinases
Nuclear
Other Centriole Genes
ReplicationFactors
RNA Binding Proteins
Transcriptional Regulators
Transition Zone Distal Appendage
Zinc Finger Containing
0.00 0.50 1.00
Proportion of protein type
E.
1.0
0.8
0.6
0.4
0.2
0
0 60 120 240 360
time (minutes)
POC1
CEP164
CEP131
VFL3
OFD1
POC1
CEP120
POC1
DRC1
RSP4
RSP3
Rib43a
RSP1
PACRG
FAP20
Hydin
F.
Bisected Bottom
Cluster
1
2
3
Scaled gene exp.
2
0
−2
0 60 120 180 240 360 420 (minutes)
G.
Ancestral Centriole Genes
Chromatin Regulators
Cyclins
Flagellar Regeneration
IFT Proteins
KIF proteins
Kinases
Other Centriole Genes
ReplicationFactors
Transcriptional Regulators
0.00 0.50 1.00
Proportion of protein type
H.
I.
OA-Specific
Regenerating Anterior
Sucrose Shock

Figure 2. Oral Apparatus regeneration program. A. OA specific gene expression profile determined using combined sucrose shock and bisection samples. Genes are clustered into 5 groups using "clara" clustering (as indicated by the colored bar on the y axis). Time since sucrose shock or bisection (in minutes) is on the x-axis. Group 1 contains all genes whose expression decreases during regeneration compared to initial levels. The peak expression of each cluster of genes corresponds with major developmental features identified morphologically (as described in Figure 1A). Z-score is calculated per row. **B.** Proportion of gene types in each cluster for a set of reference gene classes. **C.** Expression heatmap of genes expressed in sucrose-shocked cells that are not included in the OA-specific set of panel A. **D.** Proportion of gene types in each cluster for sucrose shock specific genes. **E.** Expression timecourse of selected centriole and cilia genes, illustrating the trend that cilia-related gene expression occurs later than centriole-related gene expression. **F.** Expression heatmap of gene expressed in regenerating bottom (posterior) half-cells that are regenerating new anterior portions, showing genes specific to the regenerating posterior and not included in the OA-specific set of Panel A. **G.** Proportion of gene types in each cluster for regenerating posterior half-cells **H,I.** Pie charts showing fraction of all centriole (**H**) and cilia (**I**) genes showing up-regulation in the two paradigms (sucrose shock and posterior halves), illustrating that most upregulated cilia-related genes are specific to sucrose-shock.

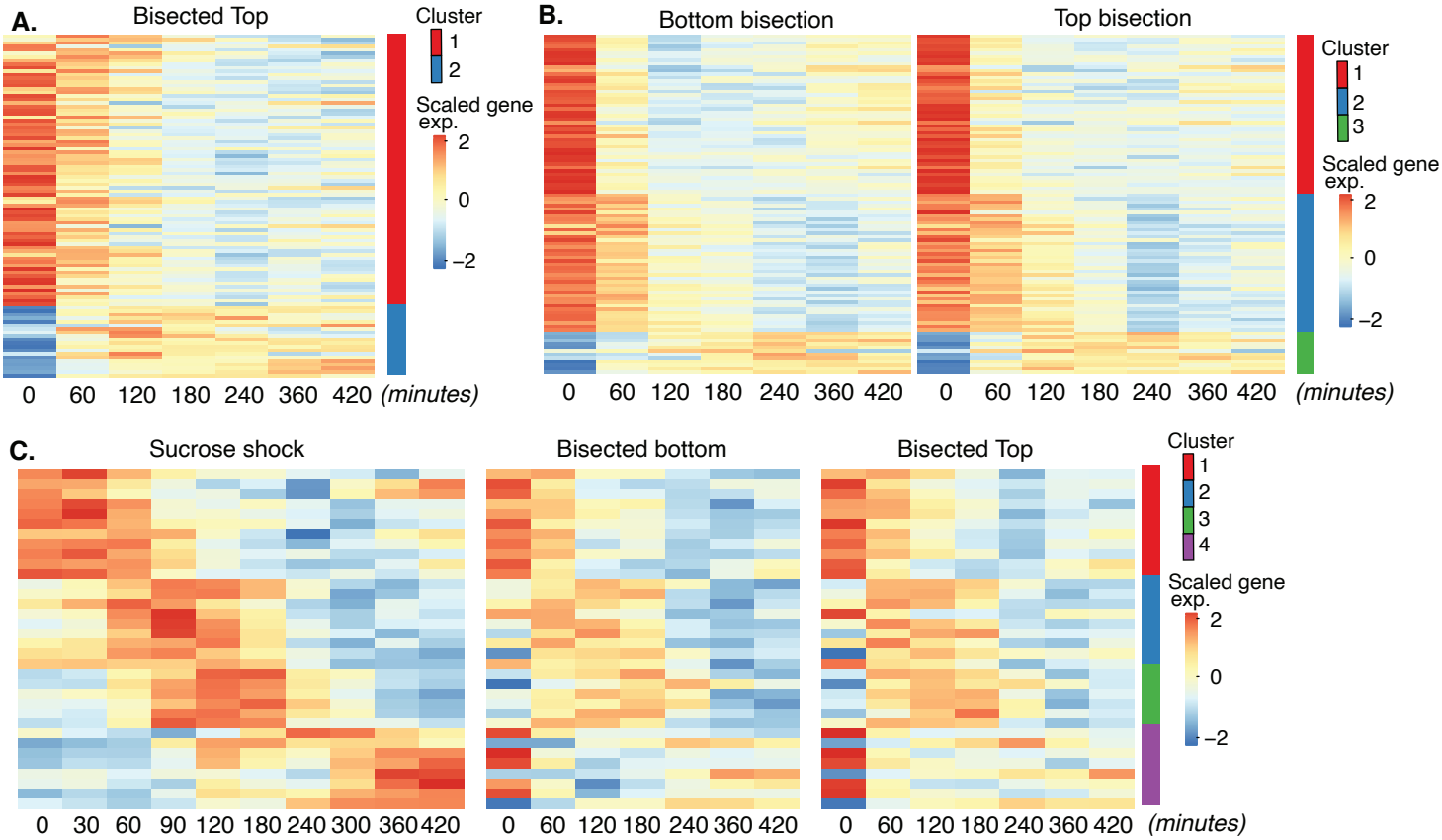

A.
Bisected Top
Cluster
1
2
Scaled gene exp.
2
0
−2
0 60 120 180 240 360 420 (minutes)
B.
Bottom bisection
Top bisection
Cluster
1
2
3
Scaled gene exp.
2
0
−2
0 60 120 180 240 360 420 0 60 120 180 240 360 420 (minutes)
C.
Sucrose shock
Bisected bottom
Bisected Top
Cluster
1
2
3
4
Scaled gene exp.
2
0
−2
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420
0 60 120 180 240 360 420 (minutes)

Figure 3. Expression modules distinct from OA regeneration. **A.** Expression in anterior half-cells that are regenerating posterior tails including holdfast. **B.** Genes showing correlated differential expression in both halves of bisected cells. **C.** Genes showing differential expression in all three types of regenerating cell fragments, sucrose shocked, anterior half cell, and posterior half cell.

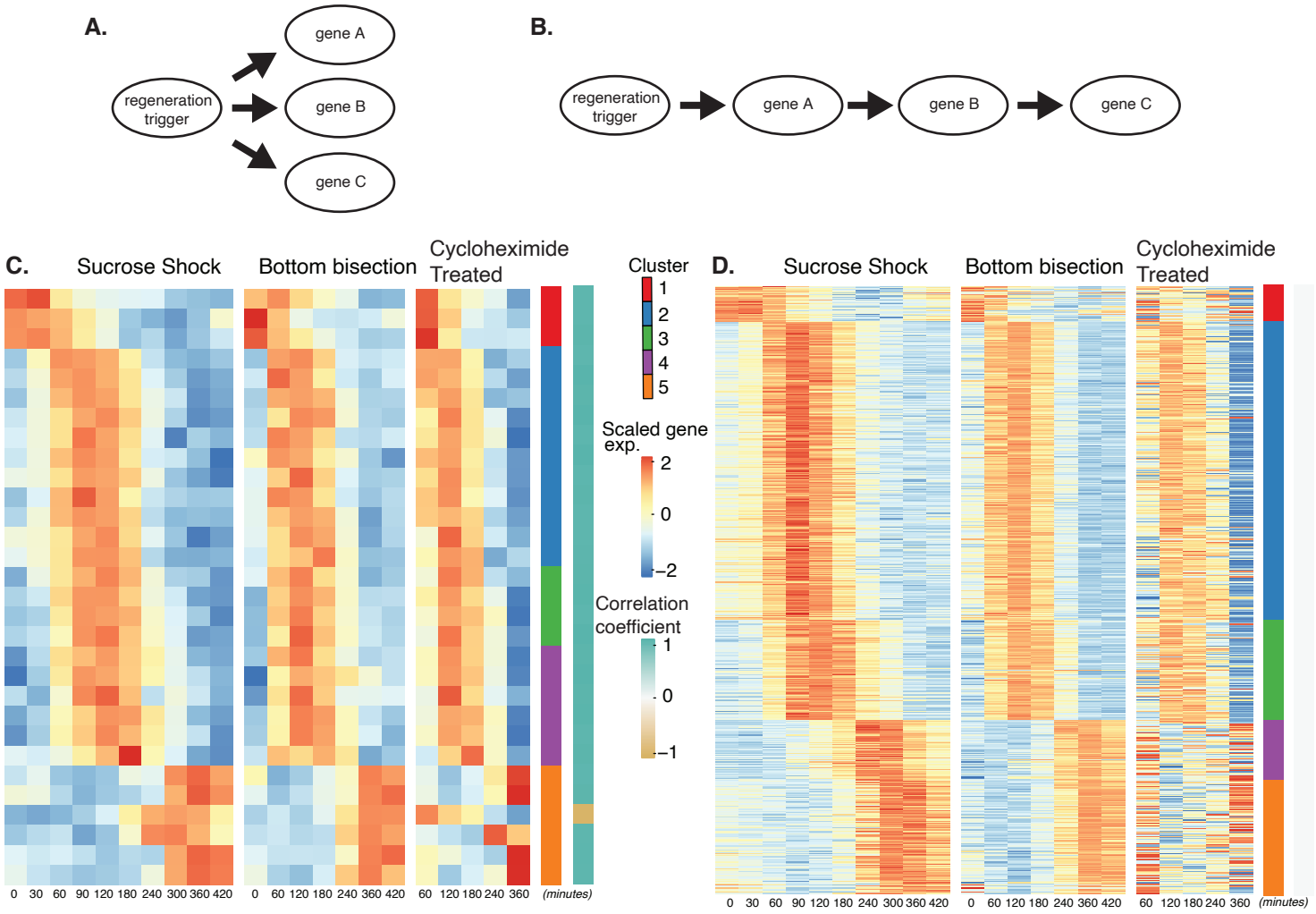

A.
regeneration trigger
gene A
gene B
gene C
B.
regeneration trigger
gene A
gene B
gene C
C.
Sucrose Shock
Bottom bisection
Cycloheximide Treated
Cluster
1
2
3
4
5
Scaled gene exp.
2
0
−2
Correlation coefficient
1
0
−1
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420
60 120 180 240 360
(minutes)
D.
Sucrose Shock
Bottom bisection
Cycloheximide Treated
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420
60 120 180 240 360
(minutes)

Figure 4. Regeneration in the absence of protein synthesis. A. Direct model of coordination, in which the regeneration-triggering stimulus directly triggers each gene. **B.** Cascade model, in which regeneration-triggering stimulus drives one gene, whose product then drives another gene, with potentially multiple steps being required to ultimately trigger all genes in the cascade. **C.** Heatmap showing OA-specific genes whose expression pattern in cycloheximide cells is highly correlated with the expression pattern in untreated cells. **D.** Heatmap showing OA-specific genes whose expression pattern in cycloheximide cells is uncorrelated with the expression pattern in untreated cells.

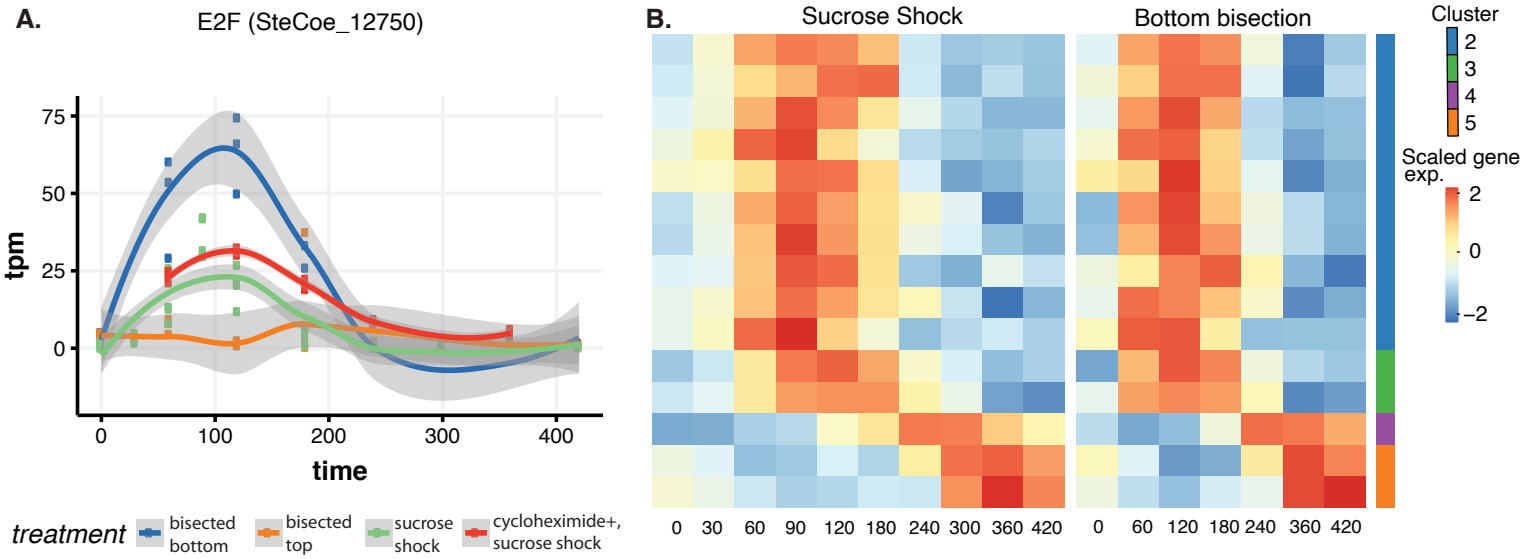

A.
E2F (SteCoe_12750)
tpm
75
50
25
0
0
100
200
300
400
time
treatment
bisected bottom
bisected top
sucrose shock
cycloheximide+, sucrose shock
B.
Sucrose Shock
Bottom bisection
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420
Cluster
2
3
4
5
Scaled gene exp.
2
0
−2

Figure 5. Expression of E2F and its targets during OA regeneration. A. E2F is an early-expressed gene. Plot shows expression of E2F homolog SteCoe_12750 as a function of time including in cycloheximide treated cells. OA (blue curve) indicates expression in sucrose shocked cells. The purple curve shows that E2F is still upregulated in cells in which protein synthesis is inhibited, thus placing it among the early genes in the cascade model. **B.** Expression heatmap of predicted E2F targets in the OA specific expression module. Clusters refer to clusters in the original clustering of OA specific genes (Figure 2A). Cluster 1 is not included in the plot because there were no predicted E2F targets among the cluster 1 genes.

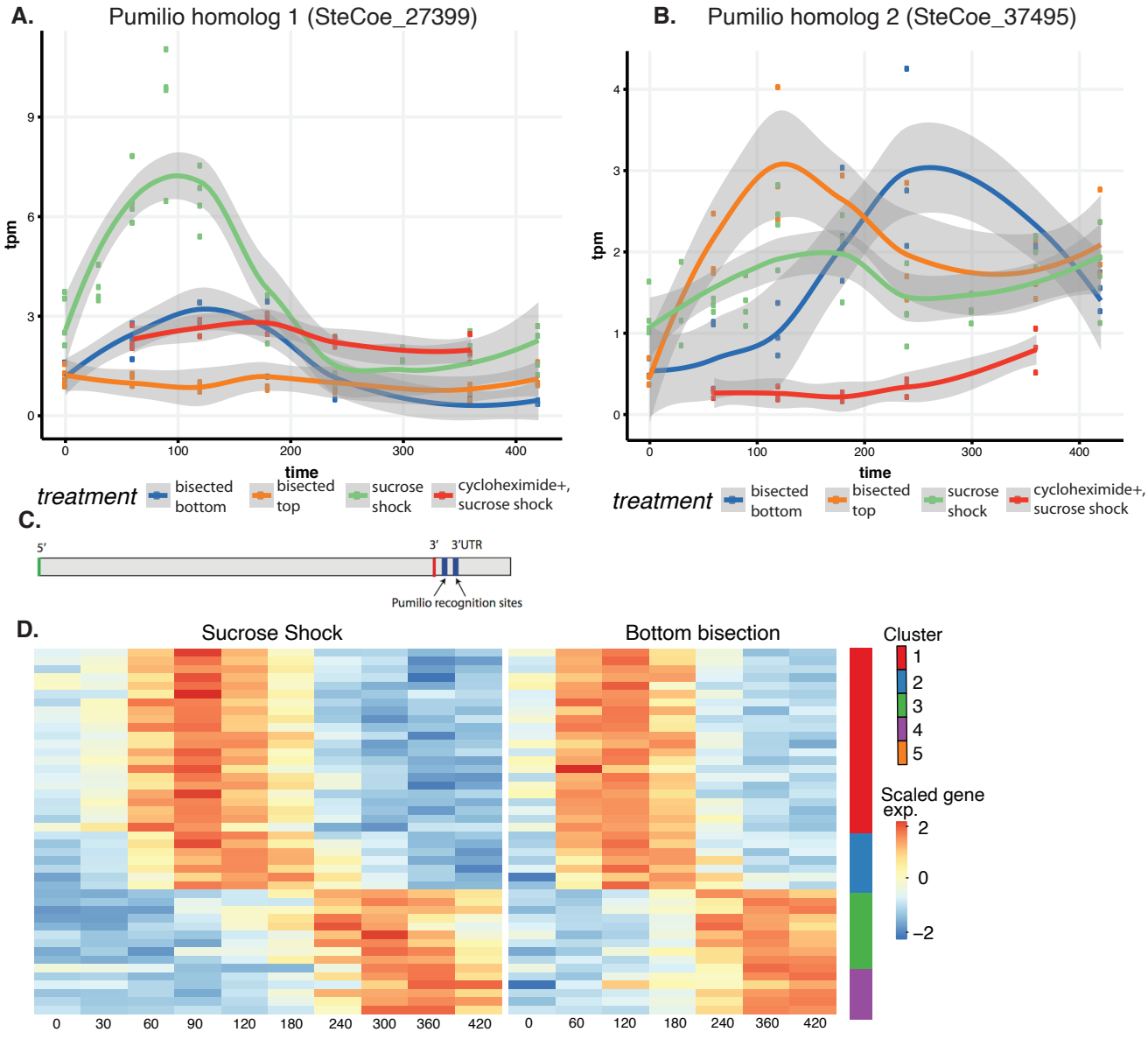

A.
Pumilio homolog 1 (SteCoe_27399)
tpm
time
treatment
bisected bottom
bisected top
sucrose shock
cycloheximide+, sucrose shock
B.
Pumilio homolog 2 (SteCoe_37495)
tpm
time
treatment
bisected bottom
bisected top
sucrose shock
cycloheximide+, sucrose shock
C.
5'
3'
3'UTR
Pumilio recognition sites
D.
Sucrose Shock
Bottom bisection
Cluster
1
2
3
4
5
Scaled gene exp.
2
0
−2
0 30 60 90 120 180 240 300 360 420
0 60 120 180 240 360 420

Figure 6. Pumilio is an early-expressed gene in OA regeneration. **A**. Expression versus time plot of Pumilio ortholog SteCoe_27339 upregulated in both cases of OA regeneration: sucrose shock (blue curve) and regeneration of the anterior half of a posterior half-cell (red curve). Expression is still seen in cycloheximide treated cells, indicating that this Pumilio ortholog is an "early" gene in the cascade. SteCoe_27339 is not expressed in anterior half cells which are not regenerating an OA. **B.** Expression versus time of Pumilio ortholog SteCoe_37495 which is upregulated in all three regeneration cases, but with different timing as indicated by the blue, green, and red curves. Unlike SteCoe_27339, this Pumilio ortholog is not expressed in cells treated with cycloheximide, indicating it is a "late" gene whose expression depends on earlier gene products of the cascade. **C**. Example of predicted Pumilio recognition sequences in a target gene (SteCoe_10652) encoding a CAMK family kinase showing OA-specific expression. **D.** Expression heatmap of predicted Pumilio targets in the OA-specific module.