A microRNA-mRNA expression network during oral siphon regeneration in Ciona

Elijah J. Spina¹, Elmer Guzman¹, Hongjun Zhou², Kenneth S. Kosik¹,² and William C. Smith¹,²,*

ABSTRACT

Here we present a parallel study of mRNA and microRNA expression during oral siphon (OS) regeneration in Ciona robusta, and the derived network of their interactions. In the process of identifying 248 mRNAs and 15 microRNAs as differentially expressed, we also identified 57 novel microRNAs, several of which are among the most highly differentially expressed. Analysis of functional categories identified enriched transcripts related to stress responses and apoptosis at the wound healing stage, signaling pathways including Wnt and TGFβ during early regrowth, and negative regulation of extracellular proteases in late stage regeneration. Consistent with the expression results, we found that inhibition of TGFβ signaling blocked OS regeneration. A correlation network was subsequently inferred for all predicted microRNA-mRNA target pairs expressed during regeneration. Network-based clustering associated transcripts into 22 non-overlapping groups, the functional analysis of which showed enrichment of stress response, signaling pathway, and extracellular proteases in late stage regeneration. Consistent with the expression results, we found that inhibition of TGFβ signaling blocked OS regeneration. A correlation network was subsequently inferred for all predicted microRNA-mRNA target pairs expressed during regeneration. Network-based clustering associated transcripts into 22 non-overlapping groups, the functional analysis of which showed enrichment of stress response, signaling pathway, and extracellular protease categories that could be related to specific microRNAs. Predicted targets of the miR-9 cluster suggest a role in regulating differentiation and the proliferative state of neural progenitors through regulation of the cytoskeleton and cell cycle.

KEY WORDS: Regeneration, RNAseq, MicroRNA, Ciona, Correlation network

INTRODUCTION

The ability of metazoan animals to regenerate varies widely. This ability can be interpreted as a spectrum extending from extreme cases of whole-body regeneration, such as is observed in planarians (Reddien and Sánchez Alvarado, 2004), to cases of limited cell replacement during tissue homeostasis, such as is observed in the mammalian digestive tract (Barker, 2014; Barker et al., 2008). In addition, many species display regenerative capacities that vary with life stage and injury severity (Rinkevich and Rinkevich, 2012). Most vertebrates, including humans, heal wounds using fibrotic mechanisms that inhibit subsequent tissue regrowth by leaving dense connective tissue that is never remodeled to the original functional state (Harty et al., 2003). By contrast, there are several well-characterized examples of vertebrates that are able to replace whole organs and/or appendages composed of multiple tissue types (King and Newmark, 2012; Stoick-Cooper et al., 2007). Furthermore, differences in cellular responses appear to directly underlie the differences in regenerative capacity of various organisms, as evidenced by differences in scar formation between regenerating vertebrates (e.g. salamanders) and non-regenerating vertebrates (Seifert et al., 2012). Regeneration is likely to be an ancestral animal trait that utilizes many of the molecular mechanisms of development, thus implying evolutionary loss of this trait in some lineages (Bely and Nyberg, 2010; Morgan, 1901; Tiozzo and Copley, 2015). By identifying points of evolutionary conservation and divergence in the molecular mechanisms underlying regeneration, progress can be made toward illuminating how different species can accomplish the same task in vastly different ways.

Whereas regenerative ability is highly variable among the vertebrates, and whole-organ regeneration is limited to very few species, the tunicates, which are primitive chordates and the closest extant relatives of the vertebrates (Delsuc et al., 2006), show robust regenerative ability. Tunicates appear to have diverged before two whole-genome duplications that are speculated to have occurred at the origin of the vertebrate lineage (Dehal and Boore, 2005), and are thought to have subsequently evolved towards reduced morphological and genomic complexity (Crow and Wagner, 2006). These traits, in combination with a sequenced and well-annotated genome, make C. robusta a powerful model for identifying the components and interactions that comprise gene regulatory networks that govern biological processes, including regeneration (Dehal et al., 2002; Satou et al., 2008). The colonial tunicates, such as Botryllus schlosseri, show whole-body regeneration (Kürm et al., 2011; Rinkevich et al., 1995, 2010; Voskoboynik et al., 2007). Ciona intestinalis and Ciona robusta, two species that until recently were both called Ciona intestinalis (Brunetti et al., 2015), are well-established models for embryology, and as adults they can rapidly and robustly regenerate their oral siphons as well as their central nervous systems (Jeffery, 2015a). The oral siphon (OS) is a cylindrical appendage composed primarily of muscular tissue, vasculature, nerves, epidermis, eight oral pigment organs (OPOs) located at the distal tip, and an outer coating of tunic (Chiba et al., 2004).

Although many studies have identified molecular mechanisms governing cellular responses during regeneration, little is known about how different molecular events are coordinated with each other (Endo et al., 2004; Jhamb et al., 2011; King and Newmark, 2012). This deficiency is evident in the current lack of effective molecular therapies available to stimulate tissue regeneration (Stoick-Cooper et al., 2007). The ability to detect transcript level changes in a comprehensive and unbiased manner can identify previously unknown coordination between signaling pathways, transcription factors and effector genes required for regeneration. There have been several transcriptome-wide studies of gene expression of both mRNA (Campbell et al., 2011; Hamada et al., 2015; Knapp et al., 2013; Love et al., 2011; Monaghan et al., 2009; Schebesta et al., 2006; Stewart et al., 2013; Wu et al., 2013) and
microRNA (miRNA) (Gearhart et al., 2015; Holman et al., 2012; Hutchins et al., 2016; Thatcher et al., 2008) during appendage regeneration. However, none of these studies connected the expression profiles of miRNA to those of mRNA.

In the current study, we present a transcriptome-wide investigation of both mRNA and miRNA expression profiles from stage-matched samples during appendage regeneration of *C. robusta*. We constructed a network of correlations between these two RNA types and inferred the function of miRNAs within a network based on the functions of their predicted target mRNAs. Our approach expands upon previous work by analyzing tissues both distal and proximal to the amputation plane using high-throughput RNA sequencing (RNAseq) of miRNAs and mRNAs, then comparing relative expression levels between distinct stages of regeneration. We identified the largest changes in gene expression during OS regeneration and conducted a systematic characterization of functional categories of all genes as well as those predicted to be targeted by specific miRNAs. This study provides a resource that will facilitate future investigation into the genetic requirements of appendage regeneration in chordates. To make our results comparable with those of a previous microarray study (Hamada et al., 2015) we also sequenced RNA from non-regenerating OSs and identified several of the same genes that they identified as the most significantly differentially expressed at 3 days post-amputation.

**RESULTS**

**Differential expression during OS regeneration**

High-throughput RNAseq was used to estimate the relative abundance of mRNAs and miRNAs in OSs of 6-month-old *C. robusta* adults. As an initial validation of our approach, mRNA samples were collected from non-regenerating (NR) OSs and used to estimate the efficiency of read alignment and power to detect differentially expressed (DE) genes based on sequencing depth and number of replicates. In this preliminary analysis, we successfully aligned an average of 92.6% (94.8% across all stages) of sequencing reads to *C. robusta*, with ~63.6% (58.4% across all stages) of reads aligning to unique genomic locations (Table S1). Next, we estimated the power of our chosen experimental design to accurately quantify differential expression using the Scotty web tool (Busby et al., 2013). Given the sequencing depth and variation within the NR OS samples, we could expect to recover nearly half of the genes that are DE greater than 2-fold at any stage as significantly DE [false discovery rate (FDR) ≤0.05] using our experimental design (Fig. S1).

For collection of regenerating tissue samples, animals were amputated distal to the buccal tentacle band (Chiba et al., 2004) (Fig. 1). Either immediately following amputation [day (D) 0] or 3 or 8 days following amputation (D1, D3 or D8) tissue was collected by making a second cut proximal to the first, but distal to the peripharyngeal (transverse) muscle band (Fig. 1, yellow dashed lines). These time points were chosen to represent three stages of regeneration observed in vertebrates: wound healing, transition and redevelopment, respectively). The red dashed line indicates the original amputation plane, and the yellow dashed lines indicate the proximal limit of tissue collected for expression profiling. Scale bars: 5 mm.

Analysis was performed using two programs in parallel: EdgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014a). Only transcripts identified as DE by both programs were used in order to reduce the likelihood of false positives, as has been described previously (Robles et al., 2012). The EdgeR and DESeq2 programs identified 337 and 472 of the mRNA transcript models, respectively, to be DE (FDR≤0.05) on at least one of the days of regeneration (i.e. D1, D3 or D8) when compared with D0. Of the DE genes on these two lists, 248 overlapped (Table S2). A roughly equivalent number of mRNAs were upregulated or downregulated (Table 1). We assessed the expression of 15 DE mRNA transcripts using quantitative real-time PCR (qRT-PCR), with particular attention to transcripts linked to morphogenetic processes and signaling pathways (Fig. 2A). Of these, nine were validated using qRT-PCR in at least one stage during regeneration (Fig. 2A). For example, transcripts for the tumor necrosis factor receptor (TNFR) and an auxiliary protein (TNFR-associated) were found to be significantly (P≤0.05) downregulated by both RNAseq and qRT-

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Method</th>
<th>LRT</th>
<th>D1 (+/-)</th>
<th>D3 (+/-)</th>
<th>D8 (+/-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA</td>
<td>EdgeR</td>
<td>15</td>
<td>4/11</td>
<td>4/11</td>
<td>6/9</td>
</tr>
<tr>
<td></td>
<td>DESeq2</td>
<td>23</td>
<td>8/15</td>
<td>7/16</td>
<td>8/15</td>
</tr>
<tr>
<td></td>
<td>Concurrence</td>
<td>14</td>
<td>4/10</td>
<td>4/10</td>
<td>5/9</td>
</tr>
<tr>
<td>mRNA</td>
<td>EdgeR</td>
<td>337</td>
<td>178/159</td>
<td>204/133</td>
<td>186/151</td>
</tr>
<tr>
<td></td>
<td>DESeq2</td>
<td>472</td>
<td>244/228</td>
<td>254/216</td>
<td>252/220</td>
</tr>
<tr>
<td></td>
<td>Concurrence</td>
<td>248</td>
<td>132/116</td>
<td>150/98</td>
<td>138/110</td>
</tr>
</tbody>
</table>

Transcripts included in pairwise comparisons (D1, D3 or D8 versus D0) were selected using likelihood ratio tests (LRT) (FDR≤0.05) by either EdgeR or DESeq2 or the concurrence of both programs. The total number of DE transcripts across all comparisons determined by LRT is indicated as well as the number of transcripts upregulated/downregulated (+/-) for each pairwise comparison.

*Fig. 1. Stages of oral siphon regeneration.* (Top) The two reference stages used to quantify relative expression levels at subsequent stages. Left image shows an unamputated adult oral siphon (OS) of *C. robusta*; the right image is immediately post-amputation. (Bottom) Three time points (1, 3 and 8 D) selected to represent three stages of appendage regeneration (wound healing, transition and redevelopment, respectively). The red dashed line indicates the original amputation plane, and the yellow dashed lines indicate the proximal limit of tissue collected for expression profiling. Scale bars: 5 mm.
Overall, out of the 550 known miRNAs in miRBase found to be strongly and significantly upregulated during expressed novel miRNA. Another novel miRNA, >50-fold greater number of reads than the second most highly identical seed sequence to expressed of these novel miRNAs, miRNAs (probability 95±3%; Table S7). The most highly however, miRDeep2 also predicted 52 previously undescribed Norden-Krichmar et al., 2007; Shi et al., 2009; Terai et al., 2012); detected during OS regeneration have been described previously align and quantify miRNA reads. Most of the miRNAs that we detected during OS regeneration (Fig. 2A). We also confirmed that the expression of two potential activators of transforming growth factor β (TGFβ) signaling were significantly upregulated during regeneration (Fig. 2A). These two pathways have previously been implicated in the regulation of programmed cell death and proliferation during regeneration (Gilbert et al., 2013; Godwin et al., 2013; Ho and Whitman, 2008; Lévesque et al., 2007; Rao et al., 2009; Stoick-Cooper et al., 2007; Wu et al., 2013).

The miRDeep2 program (Friedländer et al., 2012) was used to align and quantify miRNA reads. Most of the miRNAs that we detected during OS regeneration have been described previously (Hendrix et al., 2010; Keshavan et al., 2010; Missal et al., 2005; Norden-Krichmar et al., 2007; Shi et al., 2009; Terai et al., 2012); however, miRDeep2 also predicted 52 previously undescribed miRNAs (probability 95±3%; Table S7). The most highly expressed of these novel miRNAs, I_2011, was found to have an identical seed sequence to hsa-miR-4709-5p and was detected at >50-fold greater number of reads than the second most highly expressed novel miRNA. Another novel miRNA, I2_9033, was found to be strongly and significantly upregulated during regeneration. Overall, out of the 550 known miRNAs in miRBase (Griffiths-Jones, 2004; Kozomara and Griffiths-Jones, 2014) plus the novel miRNAs predicted by miRDeep2, we detected 279 miRNAs in the miRNA library reads at a threshold of ≥1 counts per million (CPM) in three or more samples. Comparisons were made between miRNA transcripts in the newly amputated siphon (D0) and each of the days of regeneration (D1, D3 and D8). Of the 279 miRNAs expressed, 15 and 23 were found to be DE on at least one of the days of regeneration by EdgeR and DESeq2, respectively. Of the DE miRNAs identified by EdgeR, 14 were also identified by DESeq2 (Table 1 and Table S3). In contrast to the miRNAs, there were approximately twice as many miRNAs significantly downregulated versus upregulated, suggesting an active role for miRNAs in adult OS homeostasis. The total number of DE transcripts for each day of regeneration is shown in Table 1.

Several DE miRNAs were also selected for validation based on their large estimated fold-change, significance of relative change, or previous reports indicating roles for these miRNAs in morphogenetic and regenerative processes (Fig. 2B). In regenerating OSs miR-9 is highly upregulated at all stages of regeneration analyzed, having no RNAseq reads and being nearly undetectable by qRT-PCR at D0. Additionally, the bicistronic mir-1/133 (Kusakabe et al., 2013), which has been implicated in muscle development and regeneration (Li et al., 2013; Mitchelson and Qin, 2015; Yin et al., 2008), was also upregulated during OS regeneration. In summary, the relative expression changes estimated by RNAseq for 9/15 mRNA transcripts (60%) were validated using qRT-PCR in at least one stage during regeneration. Likewise, we were able to validate that the levels of 10/17 miRNAs (58.8%) were significantly changed during at least one stage.

Functional comparison between three stages of regeneration

In addition to identifying DE transcripts following amputation, enrichment for functional categories of genes was performed using classifications curated by the Gene Ontology (GO) Consortium (Ashburner et al., 2000) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database (Kanehisa and Goto, 2000) (Table S8). Enrichment of functional categories based on the mean expression value (RPKM) of genes within each category in NR OS is shown in Fig. S2. Enrichment of functional categories based on the mean significance (−log2 FDR) of genes within each category was performed by comparing the three days of regeneration (D1, D3 and D8) with either D0 (Fig. 3) or NR OSs (Fig. S3). The complete list of z-scores for all categories relative to either reference stage is listed in Table S9.

Several GO Biological Process (BP) categories showed a pattern of enrichment across all stages, including ‘negative regulation of peptidase activity’, ‘xenobiotic metabolic processes’, ‘protein folding’, ‘cellular calcium ion homeostasis’ and ‘multicellular organismal development’ (Fig. 3A). Similarly enriched GO Molecular Function (MF) categories included ‘peptidase inhibitor’, ‘heme binding’, ‘unfolded protein binding’ and ‘growth factor activity’ (Fig. 3B). For the most part, categories modified across the 8 days of regeneration studied are consistent with a pattern of activated cellular growth and metabolism.

Other categories were enriched during specific stages of regeneration. As expected, categories enriched early in regeneration (D1 and D3) included those related to wound healing, stress response, activation of morphogenetic processes and signaling. For example, the KEGG pathway ‘AGE-RAGE signaling in diabetic complications’ (which includes upregulated immune/stress-response genes like JAK1, STAT5A and Jun) was
enriched only at D1 (Fig. 3D). GO categories enriched exclusively at D1 include ‘response to endoplasmic reticulum stress’, ‘cilium organization’, ‘steroid biosynthetic processes’ and ‘cysteine type endopeptidase activity involved in apoptosis’. Other categories enriched at both D1 and D3 include ‘microtubule-based movement and motor activity’, ‘hormone-mediated signaling’, ‘fatty acid biosynthesis’, ‘cellular homeostasis’ and ‘bicellular tight junction’.

Likewise, a number of categories with the potential to directly regulate morphogenetic processes during regeneration were enriched exclusively at D3. KEGG pathways enriched at this stage included ‘glutathione metabolism’, ‘DNA replication’ and ‘Wnt signaling’ (Fig. 3D). Corresponding GO categories enriched at this stage included ‘steroid hormone-mediated signaling’, ‘steroid hormone receptor activity’, ‘G-protein-coupled receptor activity’, ‘metalloendopeptidase activity’ and ‘voltage-gated potassium channel complex’.

Finally, categories enriched in later stages are expected to be involved in a return to homeostasis of the regenerating tissue, but could also represent downstream effects of changes observed at earlier stages and thus might not be exclusively enriched at late stage expression. Categories enriched only at D8 appeared to be less associated with signaling pathways but instead more related to tissue growth, such as ‘glucose import’, ‘negative regulation of protein kinase activity’ and ‘amino acid transmembrane transporter activity’.

**Requirement of TGFβ signaling during early stages of regeneration**

The increased expression of TGFβ activators observed post-amputation suggests a role for this signaling pathway in OS regeneration (Fig. 2). This was further supported by enrichment of functional categories related to TGFβ signaling at 3 days post-amputation (D) (Fig. 3A,B). It has been previously reported that TGFβ signaling is required for multiple events during regeneration of axolotl limbs, Xenopus tadpole tails and Eublepharis tails (Gilbert et al., 2013; Ho and Whitman, 2008; Lévesque et al., 2007).

To investigate whether TGFβ is also required during OS regeneration, we treated amputated C. robusta juveniles with 10 μM SB431542, a potent and specific inhibitor of activin receptor-like kinase receptors that mediate TGFβ signal transduction (Mita and Fujiwara, 2007).
Soaking *C. robusta* in SB431542 for 4 days following amputation completely blocked regeneration (Fig. 4). Whereas 100% of control-treated animals regenerated OPOs by 8 D, none of the SB431542-treated animals showed any visible signs of regeneration (Table 2). To further refine the temporal requirement for TGF\(\beta\) signaling, we treated separate cohorts of animals with SB431542 over 24 h periods for each of the 4 days following amputation (Fig. 4A). Surprisingly, only the cohort treated over the first 24 h following amputation failed to regenerate completely, whereas the other three cohorts showed partial OPO regeneration (Table 2). The temporal requirement was further studied using three cohorts of animals treated with SB431542 for 48 h periods starting either (1) immediately after amputation, (2) 24 h after amputation or (3) 48 h after amputation (Fig. 4A). These 48 h treatments completely blocked regeneration in all cases (Fig. 4B), indicating that TGF\(\beta\) signaling is required on at least two separate occasions during OS regeneration (once during a short window in the first 24 h, then again during a second >24 h window that includes 3 D).

### A network of miRNA target interactions

Here we report a non-biased correlation-based network to infer a comprehensive set of miRNA-target interactions over the course of 8 days of OS regeneration. Several transcriptome-wide miRNA profiling studies have been performed during limb or appendage regeneration (Gearhart et al., 2015; Holman et al., 2012; Thatcher et al., 2008), although not in parallel with mRNA transcriptome analysis. miRNAs primarily exert their effects on gene products through degradation or translational repression of target RNAs (Carthew and Sontheimer, 2009). In the absence of proteomic data, our analysis is limited to those cases in which miRNA-mRNA interactions result in degradation of the target mRNA (Carthew and Sontheimer, 2009). Assuming that an miRNA does result in degradation of a target transcript, the strength of miRNA-mRNA interactions can be estimated by identifying miRNA-target pairs that display inverse (negative) correlations between different stages of a process.

A preliminary step in this analysis required prediction of complementary sequence pairing between the full set of miRNAs (including novel miRNAs identified in this study) and the 3′-UTRs of potential target transcripts. We accomplished this using TargetScanS (Agarwal et al., 2015; Lewis et al., 2005), which was first employed to detect conserved binding sites between *C. robusta* and *C. savignyi*, before identifying additional non-conserved sites in non-orthologous genes. Our complete set of target predictions for *C. robusta* (conserved and non-conserved) contained ~521,000 pairwise interactions and is available for download (https://labs.mcb.ucsb.edu/smith/william). Notably, within the conserved set of predictions between *C. robusta* and *C. savignyi* we identified the same 824 targets for miR-124 as reported by another study (Chen et al., 2011), which identified miRNA seed pairing as a strong predictor of target downregulation in *Ciona*. Thus, we expect a large proportion of the predictions to be relevant *in vivo*. Further support for the relevance of these predictions was found by identifying a large proportion of the predicted miRNA-target pairs that are conserved with other species and have been validated experimentally (either by expression profiling, reporter assay or western blot) by comparing with information listed in miRTarBase (Chou et al., 2016) (data not shown).

The average log₂ fold-change estimated by DESeq2 and EdgeR at all stages (D0 versus D1, D3 or D8) was used to calculate Pearson correlation coefficients (\(\rho\)) for all predicted miRNA-target pairs. Subsequent analysis was limited to miRNAs with observed changes \(\geq |1 \log_2\text{fold}|\) at any stage and \(\rho \geq -0.9\) with their predicted targets. The resulting directed network contained 2033 nodes (129 miRNA, 1904 mRNA) and 2854 edges, with an average of 2.808 neighbors per node (Table S10). Evolution of biological networks is assumed to have proceeded by preferential attachment of new nodes to an existing core network, resulting in a ‘scale-free’ topology that can be modeled using a power law function (Albert, 2005; Barabási and Oltvai, 2004; Wolf et al., 2002). The in-degree distribution of edges/nodes was fitted to a power law curve with the equation \(y=2919.4x^{-1.362}\) (\(\rho=0.983, \ R^2=0.917\)), suggesting the
topology of our predicted network resembles that of a true biological network. Plotting these interactions shows a dense cluster of downregulated miRNAs associated with a dense cluster of upregulated mRNAs, joined by a few sparse connections (Fig. 5).

To assign putative functions to miRNAs within the correlation network, we implemented a clustering algorithm on all the nodes and then examined whether the list of targets in each cluster were enriched for particular GO categories and KEGG pathways. Non-overlapping clusters of miRNAs and targets were identified by analyzing the correlation network using the LinkLand algorithm (Kovács et al., 2010) provided by the ModuLand plug-in (Szalay-Beko et al., 2012) for Cytoscape (Smoot et al., 2011). Briefly, each node is assigned a set of influence scores that relate it to all other nodes in the network based on their relative position and connections. Next, clusters are determined by identifying local maxima (which indicate the centers) and local minima (which define the boundaries) of influence scores throughout the network. Table 2. Number of animals with regenerated OPOs after SB431542 treatment

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Duration</th>
<th>4 D</th>
<th>8 D</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMSO</td>
<td>4 days</td>
<td>56/67</td>
<td>61/67</td>
</tr>
<tr>
<td>0-4</td>
<td>4 days</td>
<td>0/28</td>
<td>0/20</td>
</tr>
<tr>
<td>0-1</td>
<td>24 h</td>
<td>0/11</td>
<td>0/11</td>
</tr>
<tr>
<td>1-2</td>
<td>24 h</td>
<td>2/13</td>
<td>8/13</td>
</tr>
<tr>
<td>2-3</td>
<td>24 h</td>
<td>0/11</td>
<td>3/11</td>
</tr>
<tr>
<td>3-4</td>
<td>24 h</td>
<td>4/9</td>
<td>5/9</td>
</tr>
<tr>
<td>0-2</td>
<td>48 h</td>
<td>0/27</td>
<td>0/20</td>
</tr>
<tr>
<td>1-3</td>
<td>48 h</td>
<td>0/28</td>
<td>0/20</td>
</tr>
<tr>
<td>2-4</td>
<td>48 h</td>
<td>0/28</td>
<td>0/20</td>
</tr>
</tbody>
</table>

Cohort names correspond to conditions illustrated in Fig. 4A.

The miRNAs and targets assigned to each cluster are listed in Table S11. All of the targets within each resulting cluster were grouped and then the significance of overlap between each cluster and GO/KEGG functional categories was assessed using a hypergeometric test for enrichment (Table S12). Interpretations summarizing the significantly enriched functional categories are shown alongside their respective color-coded miRNA-target clusters in Fig. 5.

DISCUSSION

Molecular signatures indicate conserved features of regeneration

Appendage regeneration is thought to occur in three morphologically distinct stages; wound healing, transition and redevelopment (Knapp et al., 2013). The wound healing stage is primarily defined by a localized immune response, closure of the epithelium and initiation of blastema formation (Murawala et al., 2012). We observed several transcriptional changes indicating conserved features of wound healing during OS regeneration. GO categories supporting this hypothesis that are enriched at D1 include ‘response to ER stress’, ‘response to oxidative stress’, ‘protein folding’, ‘unfolded protein binding’, ‘cysteine-type endopeptidase activity involved in apoptosis’ and the KEGG pathway ‘AGE-RAGE signaling pathway in diabetic complications’ (Fig. 3). Following healing, a transition occurs in which the existing appendage begins to redevelop the lost tissue. During this transition phase, positional identity is regained and signaling to activate progenitor cells that are required to initiate growth occurs. Insulin growth factor (IGF) signaling was first implicated in limb regeneration in newts over three decades ago (Vethamany-Globus, 1987; Vethamany-Globus et al., 1984). More recently, it was...
discovered that IGF secreted from wounded zebrafish epithelia stimulates underlying mesenchymal blastema cells to proliferate (Chablais and Jazwinska, 2010). Further, it has been hypothesized that a Warburg effect (Vander Heiden et al., 2009) can be promoted in regenerating vertebrate tissues to favor structural biosynthesis over generation of ATP (Love et al., 2014b). We observed high levels of IGF and IGF binding protein (IGFbp) transcript levels throughout regeneration (Fig. 2). IGF binding proteins have been reported to act as carriers for IGF to promote increased persistence throughout regeneration (Hwa et al., 1999). Further investigation into the role of IGF signaling during OS regeneration could help determine whether the effects of this signaling pathway on cell proliferation and energy metabolism in different models of vertebrate regeneration are likely to be derived from a common evolutionary origin.

The transverse vessels of the branchial sac in C. robusta are thought to contain progenitor cells required for OS regeneration that are activated to proliferate and migrate into the OS after amputation (Jeffery, 2015b). Although the extent to which these branchial sac stem cells contribute to various tissues has not yet been investigated, it is suggested that they have the potential to differentiate into multiple lineages due to high expression of pluripotent cell marker genes such as Piwi and Alkaline phosphatase (Juliano et al., 2011, 2014; O’Connor et al., 2008; Stefkova et al., 2015). Piwi-positive stem cells in colonial tunicates are essential for whole-body regeneration (Brown et al., 2009) and also originate from a vascular niche (Rinkevich et al., 2010). Cells of the branchial sac divide following amputation (Jeffery, 2015b); however, EdU and phospho-histone H2 labeling of regenerating OSs showed no detectable increase in the number of dividing cells per unit area until ~4 D (Auger et al., 2010). We observed transcriptional changes supporting the proposed timing of proliferation, such as the enrichment at D3 of the KEGG pathways ‘DNA replication’ and ‘Wnt signaling’ along with the GO categories ‘TGFβ receptor binding’, ‘cell fate commitment’ and ‘growth factor activity’ (Fig. 3).

Our identification of the requirement for TGFβ signaling at 3 D, as indicated by the lack of subsequent tissue regrowth and OPO differentiation following SB431542 treatment, as opposed to any potential disruption of patterning or the null hypothesis of no effect, further supports the hypothesis that a progenitor population is receiving inductive signals, putatively TGFβ itself, at this stage in order to stimulate proliferation and regeneration of lost tissue. Further studies comparing localization of TGFβ pathway members and effects on proliferation after SB431542 treatment during OS regeneration could identify the mechanism by which this pathway regulates regenerative regrowth and differentiation.

Once progenitor cells have been specified and positional identity within the regenerating tissue is established, the latter stages of regeneration are thought to proceed in a similar manner to how the original appendage/tissues were formed during development. This process involves extensive remodeling of the extracellular matrix (ECM) to regulate cellular responses such as apoptosis, proliferation, migration and differentiation (Calve et al., 2010; Lu et al., 2011). ECM remodeling is regulated via activation of specific enzymes such as Serpins (Simone et al., 2014) and TIMPs (Arpino et al., 2015), which inhibit peptidases that degrade structural ECM proteins. We observed strong enrichment of the GO categories ‘peptidase inhibitor’ and ‘negative regulation of peptidase activity’ at later stages of regeneration, particularly at D8 (Fig. 3). This supports the hypothesis that OS regeneration involves substantial ECM remodeling at early and mid stages, which is actively attenuated by expression of peptidase inhibitors such as serpins and TIMPs at later stages.

A microarray-based study of gene expression during OS regeneration was recently published (Hamada et al., 2015). This study reported the 30 genes with the largest changes relative to NR OS at 3, 6 and 9 D. We compared the list of genes reported as DE at day 3 of regeneration with an analogous list derived in this study (Table S4). When converting from the gene identifiers used on the microarray, we were able to identify Ensembl gene models for 26 out of the 30 microarray probe sets listed. Of these Ensembl transcript models, 18/26 transcripts were expressed above the threshold of 5 RPKM in at least three or more samples, but only 6 of these 18 were identified as DE (FDR ≤0.05) by DESeq2, 5 of which were confirmed by EdgeR, notably including Notch and EphA4.

**Stage specificity of functional category overlap with miRNA-target clusters**

We observed significant overlap between transcripts in certain functional categories and particular miRNA-target clusters. For example, categories related to ‘immune response’, ‘stress responses’ and ‘apoptosis’ were enriched at D1 relative to D0. These categories also significantly overlapped with miRNA-target clusters miR-4178b-5p and 4_20211 (Table S12). Further, attenuation of apoptosis and induction of morphogenetic growth factor signaling are crucial for the transition from wound healing to the activation of redevelopement. Several miRNA clusters were found to target mRNAs annotated in functional categories related to Wnt, TGFβ and MAPK signaling, as well as regulation of apoptosis and regulation of cell cycle including miR-4008c-5p, miR-4123-5p, miR-4178b-5p, miR-4178b-5p, 2_15911, 4_20211 and 11_7539 (Table S12). Finally, during the redevelopment stage we observed strong enrichment of ECM peptidase inhibitors (Fig. 3), which significantly overlapped with the miRNA-target clusters miR-4008c-5p, 10_4533 and 11_6940.

During all stages of regeneration we observed increased levels of IGF and IGFbp transcripts (Fig. 2A). The GO MF categories ‘insulin receptor binding’ and ‘insulin-like growth factor binding’ significantly overlapped with the miRNA-target cluster 10_4533 (Table S12). Regulation of IGF-regulated processes by cluster 10_4533 is further supported by significant overlap of its targets with the GO BP categories ‘regulation of cell fate specification’, ‘regulation of cell growth’ and ‘regulation of mitotic cell cycle’ (Table S12). This miRNA-target cluster might also be involved in regulating multiple unrelated processes at specific stages of regeneration. First, this cluster significantly overlaps with the categories ‘unfolded protein binding’ and ‘regulation of apoptotic processes’, which are enriched at the D1 stage (Fig. 3). Second, as mentioned previously, the miRNA-target cluster 10_4533 significantly overlaps with the GO MF category ‘endopeptidase inhibitor activity’, which is highly enriched at D8 (Fig. 3).

We observed miR-9, an ancient and well-conserved miRNA essential for normal function of developing and differentiated neurons (Coolen et al., 2013), to be specifically upregulated during regeneration. This miRNA has previously been identified as expressed during the gastrula and larval stages of C. robusta (Hendrix et al., 2010). In other species miR-9 has been shown to regulate differentiation via targeting of the Notch signaling pathway (Jing et al., 2011; Kuang et al., 2012). The proliferative state of neural progenitors is governed by oscillations in the protein level of Hes1; high levels of miR-9 were shown to dampen oscillations of Hes1 leading to increased proliferation and differentiation (Bonev et al., 2012; Tan et al., 2012). Interestingly, we did not identify...
binding sites for miR-9 in any members of the Notch pathway or predicted downstream targets of Hes1 (data not shown). However, our 17 predicted targets for miR-9 do suggest a possible role in regulating differentiation and the proliferative state of neural progenitors through regulation of the cytoskeleton and cell cycle (Galdersi et al., 2003; McBeath et al., 2004). Functional categories significantly overlapping with miR-9 network cluster targets that are associated with cell cycle regulation are ‘cell division’, ‘DNA replication initiation and DNA replication’, which contain the miR-9 targets Nek7-like, SFI1-like and Myb-binding IA. Categories significantly overlapping with miR-9 network clusters associated with cytoskeletal regulation are ‘intermediate filament’, ‘cytoskeleton organization’, ‘microtubule-based movement’, ‘microtubule motor activity’, ‘kinesin complex’ and ‘microtubule binding’, which include the miR-9 targets villin-1, Kinesin, sideroxin-1-like, LIMK1-like, myosin X and LMNTD1.

Significance and limitations of the predicted miRNA-mRNA network

miRNA regulation is important in several well-studied examples of regeneration (Sen and Ghatak, 2015). miRNAs can either lead to direct degradation or translational repression of their target transcripts (Carthew and Sontheimer, 2009). One obvious caveat of miRNA and mRNA transcriptional profiling is that translational repression cannot be detected. If a given miRNA is predicted to target a given transcript based on seed pairing and the expression levels of these two transcripts are inversely correlated, we infer the predicted interaction is specific and results in target degradation. Predicted target pairs that do not show an inverse correlation could still result in translational repression; however, the relevance and extent of this type of interaction was not considered in this study.

For 8 of the 22 clusters detected by ModuLand (Fig. 5 and Table S11), the most central nodes identified as representing each cluster were novel miRNAs identified in this study. This underscores the complementary nature of co-expression and differential expression analyses. Co-expression analysis was able to identify individual miRNAs(s) with small relative fold changes as important during regeneration by virtue of the position of that miRNA within a correlation network. On the other hand, differential expression analysis did not identify many novel miRNAs to be important during regeneration but was able to identify miRNAs that had the largest fold changes at any stage.

Conclusions

Concurrent expression profiling of mRNA and miRNA has proven to be a useful approach for characterizing transcriptome-wide changes in gene expression during OS regeneration. We successfully identified a variety of transcriptional changes supporting the hypothesis that several major features of vertebrate appendage regeneration are conserved in C. robusta. Examples include categories of genes related to wound healing, proliferation, differentiation and ECM remodeling that were enriched at specific stages of regeneration. Further, we present a high-confidence network of miRNAs and their predicted targets during OS regeneration. This enabled subsequent annotation of putative miRNA functions during regeneration by virtue of identifying discrete clusters of miRNAs and their associated targets within the network. Several clusters of miRNAs and their targets were found to significantly overlap with functional categories that were likely to be involved in specific stages of OS regeneration owing to the preferential enrichment of these functional categories at specific stages.

This work provides a systematic characterization of mRNA and miRNA expression during OS regeneration, a comprehensive set of predicted interactions between these two gene product types and predicts effects of those interactions on cellular processes during regeneration – all of which are necessary to facilitate future investigation into the genetic requirements of appendage regeneration in chordates.

MATERIALS AND METHODS

Animal husbandry and RNA extraction

C. robusta were collected at the Santa Barbara Yacht Harbor. Gametes from two adults were mixed and the resulting offspring were grown to ~6 months of age. Approximately 75 sibling animals ranging in size from 5-10 cm were selected for amputation and anesthetized in 0.04% MS222 for 30 min. Amputated siphon samples were immediately frozen in liquid nitrogen then transferred to 1 ml RNAlater-ICE (Life Technologies) and stored at −20°C for 1 week. Total RNA was extracted from a pool of five tissue samples for each replicate using the miRvana Total RNA Extraction Kit (Life Technologies). During timecourse experiments, batches of animals were housed in five-gallon (~19 liter) buckets of seawater, changed daily and maintained at 15°C with a 12/12 h artificial light/dark cycle.

RNA sequencing and data preprocessing

Poly(A) RNA was isolated from 5 µg total RNA using Dynabeads (Life Technologies). Small RNAs (~18-30 nucleotides), including miRNAs, were isolated from 1 µg total RNA by 1% Tris-borate-EDTA acrylamide gel electrophoresis. Sequencing libraries were prepared with the Total RNAseq Kit v2.0 (Life Technologies). Libraries were sequenced in multiplex using an Ion Proton sequencer (Life Technologies). Raw reads were trimmed of adapter sequences and preprocessed for base quality according to the default software specifications. Further quality control was performed using FastQC and SamTools (Li et al., 2009). The C. intestinalis (now called robusta) JGI genome version 2.0 and Ensembl transcript models (release 83) were used as references for alignment. miRNA libraries were first aligned to the C. robusta genome and set of Ensemble transcript models using TopHat (Trapnell et al., 2009) with strict parameters (‘end-to-end’ mode) to accommodate spliced reads, then the initially unaligned reads were processed using Bowtie (Langmead et al., 2009) with relaxed parameters (‘local’ mode) to increase the overall robustness of alignment to the highly polygenic C. robusta genome. HTseq-count (Anders et al., 2015) was used to count reads for each miRNA transcript. miRNA precursor sequences were downloaded from miRBase (Kozomara and Griffiths-Jones, 2014) and used along with the C. robusta genome by miRDeep2 (Erdeländer et al., 2012) to align and quantify the miRNA libraries. Counts of novel miRNAs were extracted from the miRDeep2 output using a custom Python script.

Differential expression, co-expression and enrichment analyses

miRNA samples were initially normalized for sequencing depth and transcript length (RPKM), whereas miRNA samples were first normalized only by sequencing depth (CPM). To reduce the number of tests performed, low and non-expressed genes were removed using filters of RPKM ≥5 or CPM ≥1 in at least three individual samples for mRNA and miRNA libraries, respectively. Normalization of counts and likelihood ratio tests were performed with the EdgeR and DESeq2 packages in R to determine differential expression relative to D0. Pairwise correlations (ρ≤−0.9) were calculated using R from log2 fold-change values for miRNA and miRNA data sets estimated by DESeq2. To determine GO categories and KEGG pathways that were enriched at each stage relative to a reference stage (NR or D0), the mean −log transformed FDR-adjusted P-value of each gene determined by EdgeR and DESeq2 was used as input for a two-tailed z-test (Maciejewski, 2013; Perez-Llanas and Lopez-Bigas, 2011). z-tests were performed using a 10,000 replicate bootstrap and FDR multiple testing adjustment using Giotools 2.2.3 (Perez-Llanas and Lopez-Bigas, 2011). Significance of overlap between miRNA targets and functional categories was performed using hypergeometric tests in R, with the resulting P-values corrected for multiple testing (FDR). For
overlap tests, lists or categories with fewer than five transcripts were removed to reduce false positives.

qRT-PCR validation of relative expression
Custom TaqMan assays manufactured by Thermo Fisher were used for quantifying relative amounts of miRNA expression. SYBR Green detection was used for quantifying relative amounts of mRNA expression. All experiments were performed using a QuantStudio 1200K-Flex platform and analyzed using ExpressionSuite software (Applied Biosciences). mRNA expression levels were normalized to the geometric mean of GAPDH and RNA polymerase 2. Global normalization was used for miRNA expression levels because a sufficiently stable endogenous reference gene could not be identified despite testing several of the least variable genes from the deep-seq data. Primer sequences for mRNA detection were designed using BatchPrimer3 (You et al., 2008) and are listed in Table S5 along with Ensembl transcript identifiers and all gene names used in this study. The sequences of the miRNA detection probes are proprietary (Applied Biosystems) and not available to the researchers. One-way ANOVA was used to determine transcripts that were significantly DE during at least one stage of regeneration, as assessed by qRT-PCR (Table S6).

miRNA target prediction
3′-UTR sequences and a paired list of orthologous genes for C. robusta and C. savignyi were downloaded from Ensemble Biomat: orthologous 3′-UTR sequences were then grouped according to the pairs of orthologous transcript IDs. Each resulting group of orthologous 3′-UTRs was aligned using Clustal Omega (Sievers et al., 2011) and then the resulting alignments were concatenated into a single file that included the remaining C. robusta 3′-UTRs that did not have orthologs in C. savignyi. The multiple alignment results were used with mature miRNA sequences downloaded from miRBase directly by TargetScanS (Lewis et al., 2005) to identify conserved seed regions for all C. robusta miRNAs (known and predicted in this study).

Acknowledgements
We thank Dr Matthew Kourakis and Dr Erin Newman-Smith for helpful criticisms of this manuscript, and Dr Otto Guedelhoefer and Dr Sarah Abdul Wajid for insight and feedback while designing experiments.

Competing interests
The authors declare no competing or financial interests.

Author contributions

Funding
This work was supported by the National Institutes of Health [HD038701 to W.C.S.]. Deposited in PMC for release after 12 months.

Data availability
Raw and processed data (transcript counts) are available from the NCBI Gene Expression Omnibus repository under accession number GSE84387 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=kwqotcwqzvzap&acc=GSE84387).

Supplementary information
Supplementary information available online at http://dev.biologists.org/lookup/doi/10.1242/dev.144097.supplemental

References


Stoick-Cooper, C. L., Moon, R. T. and Weidinger, G. (2007). Advances in signaling in vertebrate regeneration as a prelude to regenerative medicine. Genes Dev. 21, 1292-1315.


