

# SCIENTIFIC REPORTS



OPEN

## Transcriptomic investigation of wound healing and regeneration in the cnidarian *Calliactis polypus*

Zachary K. Stewart<sup>1</sup>, Ana Pavasovic<sup>2,3</sup>, Daniella H. Hock<sup>4</sup> & Peter J. Prentis<sup>1,5</sup>

Received: 26 October 2016  
 Accepted: 19 December 2016  
 Published: 02 February 2017

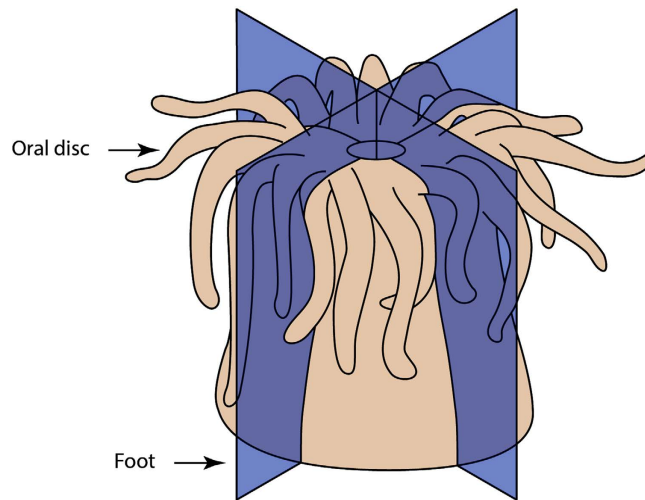
Wound healing and regeneration in cnidarian species, a group that forms the sister phylum to Bilateria, remains poorly characterised despite the ability of many cnidarians to rapidly repair injuries, regenerate lost structures, or re-form whole organisms from small populations of somatic cells. Here we present results from a fully replicated RNA-Seq experiment to identify genes that are differentially expressed in the sea anemone *Calliactis polypus* following catastrophic injury. We find that a large-scale transcriptomic response is established in *C. polypus*, comprising an abundance of genes involved in tissue patterning, energy dynamics, immunity, cellular communication, and extracellular matrix remodelling. We also identified a substantial proportion of uncharacterised genes that were differentially expressed during regeneration, that appear to be restricted to cnidarians. Overall, our study serves to both identify the role that conserved genes play in eumetazoan wound healing and regeneration, as well as to highlight the lack of information regarding many genes involved in this process. We suggest that functional analysis of the large group of uncharacterised genes found in our study may contribute to better understanding of the regenerative capacity of cnidarians, as well as provide insight into how wound healing and regeneration has evolved in different lineages.

The extent of an organism's ability to heal, repair and regenerate damaged tissues varies substantially across eumetazoan lineages. Understanding the molecular mechanisms that underpin these processes has been of long standing interest, particularly as most vertebrate species, upon maturation, adopt a wound healing and regeneration strategy that often results in the loss of original tissue function through scar formation<sup>1</sup>. The mechanism by which wounds heal is distinct from that which controls tissue regeneration; wound healing involves the interplay of factors that attempt to close a wound and prevent infection, often resulting in scar formation<sup>2</sup>, whereas regeneration restores tissue back to its pre-injury state without scar formation<sup>3</sup>. Understanding the processes that determine whether wound healing will progress into regeneration or end as scar formation is not only an important area of clinical research, but is also key to our understanding of the evolutionary loss or gain of regeneration in different eumetazoan lineages<sup>3</sup>.

To date, a number of studies have investigated regeneration in species with high regenerative capacity. Examples include those in the cnidarian *Hydra*<sup>4</sup>, the planarian flatworms<sup>5</sup>, and the urodele salamanders<sup>6</sup>. These species display significant morphological differences, yet certain key molecular mechanisms appear to be present in some or all of these regeneration competent organisms<sup>7</sup>. Brookes and Kumar<sup>7</sup> summarise many of these mechanisms, which include bioelectrical currents generated by the nervous system, the formation of wound epithelium at the site of injury, the immune system and inflammation, as well as a number of tissue patterning pathways key to embryological development. While studying these similarities contributes greatly to our knowledge of wound healing and regeneration, outlining the differences between these organisms is perhaps of equal importance, in order to determine how this process has evolved in different lineages.

The phylum Cnidaria consists of a diverse range of aquatic species with varied regenerative potential, including anthozoans (corals and sea anemones), scyphozoans (jellyfish), cubozoans (box jellyfish) and hydrozoans. While freshwater *Hydra* may be considered as the model organism for studies of regeneration in Cnidaria,

<sup>1</sup>School of Earth, Environmental and Biological Sciences, Queensland University of Technology, 2 George Street, Brisbane, Australia. <sup>2</sup>School of Biomedical Sciences, Queensland University of Technology, 2 George Street, Brisbane, Australia. <sup>3</sup>Institute of Health and Biomedical Innovation, Queensland University of Technology, 2 George Street, Brisbane, Australia. <sup>4</sup>Biological Sciences Centre, Federal University of Santa Catarina, Florianópolis, Brazil. <sup>5</sup>Institute for Future Environments, Queensland University of Technology, 2 George Street, Brisbane, Australia. Correspondence and requests for materials should be addressed to Z.K.S. (email: zkstewart1@gmail.com)



**Figure 1. Diagrammatic depiction of the planes through which anemones were cut.** Blue coloured longitudinal planes indicate where each *C. polypus* individual ( $n = 3$ ) was precisely cut from oral disc (top) to foot (bottom), resulting in four equally sized quarters with the same representation of tissues and cell types. Each fragment was sampled to represent a separate time point (0, 3, 20, and 96 hours post sectioning).

comparative analyses with other cnidarians need to be conducted as more evidence emerges to indicate that the patterns observed in *Hydra* may not be representative of other cnidarian species. For example, the actiniarians *Calliactis parasitica* and *Nematostella vectensis* appear to regenerate despite the absence of stem cells that contribute to regeneration observed in *Hydra*<sup>8,9</sup>. Furthermore, the ability to regenerate is not seen in all actiniarians, providing a suitable model to investigate the differences between organisms that have retained this trait and those that have lost it. Given the evolutionary position of Cnidaria as sister group to Bilateria<sup>10,11</sup>, it is likely that discoveries made in this group of organisms will affect our perspective on bilaterian regeneration as a whole.

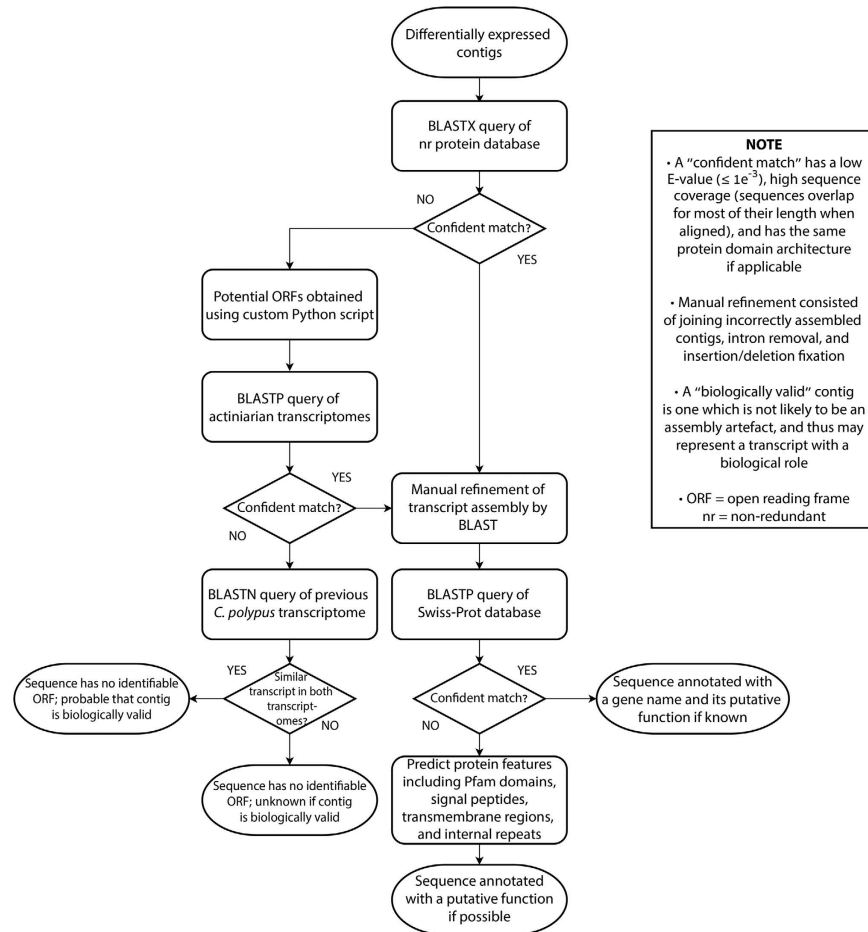
*Calliactis polypus*, in addition to being regeneration competent, is well positioned as a distant relative of the model actiniarian *N. vectensis* and as a closer relative to *Exaiptasia pallida*<sup>12</sup>. *Nematostella vectensis* has been the subject of recent investigations into the transcriptomic response elicited during regeneration<sup>9,13</sup>. Additionally, *E. pallida* will serve as an important source of information regarding actiniarian regenerative processes in the future due to its recently published genome<sup>14</sup>. In this study, we assessed the regenerative response of *C. polypus* after catastrophic injury, using a differential gene expression analysis over a course of 96 hours. Our results highlight the role of conserved eumetazoan genes in coordinating wound healing and regeneration in actiniarians, while we also identify the potentially important role of previously uncharacterised lineage-specific genes involved in *C. polypus* regeneration.

## Materials and Methods

**Experimental set-up.** Genetically distinct *C. polypus* were obtained from rock pools at Port Cartwright, Sunshine Coast, Australia (refer to Supplementary Data 1 [1. genetic\_relatedness]). Animals were housed in glass aquaria with ambient temperatures ranging between 20–22 °C and water conductivity maintained between 33.5 to 35.5 mS/m until experimentation. Four days prior to experimentation, three *C. polypus* individuals were isolated in a separate glass tank and feeding was halted. During the experiment, water quality was maintained at the same conditions as that of the glass aquaria noted previously.

**Animal sectioning.** To stimulate regeneration, animals were sectioned along their longitudinal axes at right angles, resulting in four equally sized fragments (Fig. 1). It was observed from preliminary experiments that all fragments resulting from this process were able to survive and regenerate into whole organisms. The four fragments originating from each of the three individuals were collected at 0, 3, 20, and 96 hours post sectioning (hps), immediately frozen in liquid nitrogen, and stored at –80 °C for later RNA extraction. These four time points were chosen to reflect physiological and behavioural changes (i.e., at 3 hps the secretion of a mucus layer was evident, and at 96 hps sea anemones began to pulsate their entire bodies) or to reflect stages of regeneration noted in previous studies (i.e., 20 hps marks the early regenerative phase of *N. vectensis*)<sup>9</sup>.

**Transcriptome sequencing and assembly.** Total RNA was extracted from each sea anemone fragment after homogenisation in liquid Nitrogen using an RNeasy Midi Kit (Qiagen), with resulting RNA visually inspected on a 1.5% agarose gel stained with GelRed (Biotium). Further quality assessment was performed using an RNA nano chip on a Bioanalyzer 2100 (Agilent) to confirm RNA integrity and concentration. Seventy-five base pair (bp) single-end libraries for each sample were prepared using the Illumina TruSeq Stranded mRNA Library Preparation Kit as per the manufacturer's instructions, with resulting cDNA quality and concentration being inspected using a high sensitivity DNA chip on a Bioanalyzer 2100 (Agilent). Sequencing was performed using an Illumina NextSeq 500.



**Figure 2.** Flowchart depicting the main bioinformatic process employed to identify open reading frames and annotate contigs with functions.

Resulting reads were assembled into a single reference transcriptome using the Trinity short-read *de novo* assembler v2.0.6<sup>15</sup>. This process involved the trimming of low quality leading and trailing bases ( $< Q30$ ) from each read, while also trimming Illumina adapter sequences with the inbuilt Trimmomatic tool<sup>16</sup>. The final reference transcriptome was analysed using BUSCO<sup>17</sup> to assess the completeness of our transcriptome assembly by identifying the proportion of genes represented in our transcriptome that are expected to be present in almost all metazoans. Additional quality statistics were also obtained, including the number of contigs, N50, and median contig length.

**Differential gene expression.** Differential gene expression (DGE) analysis was performed following the Trinity short-read *de novo* assembler’s edgeR protocol<sup>18</sup>. The list of contigs with a statistically significant differential expression ( $\log_2 |FC| \geq 2$ ,  $FDR \leq 0.01$ ,  $n = 3$ ) was then curated. Curation involved subjecting the differentially expressed contigs to a process of redundancy removal using CD-HIT<sup>19</sup> to remove duplicated and overlapping sequences with an identity parameter of 95%. After this, the remaining contigs were manually curated using MUSCLE<sup>20</sup> alignment of contigs assembled as isoforms by Trinity to identify any redundant sequences that were not removed by CD-HIT.

**Conserved open reading frame identification and assembly refinement.** The resulting pool of curated contigs was subjected to a process of open reading frame (ORF) prediction and assembly refinement (Fig. 2). Firstly, protein-coding ORFs were identified by BLASTX query to the non-redundant (nr) protein database<sup>21</sup> (queried 16/05/16). Contigs lacking ORFs which obtain confidently predicted homologues in the nr protein database (wherein confident refers to a match with a significant E-value [ $\leq 1e^{-3}$ ], high sequence identity and alignment coverage, and the same protein architecture if applicable) had potential ORFs predicted using a custom Python script (script is available from <https://github.com/zkstewart/orf-finder-py>). The parameters of this script were set to obtain ORFs with a minimum size of 33 amino acids, with lower stringency for selection of non-canonical ORFs (refer to Supplementary Methods for more detail). These ORFs were utilised as BLASTP queries against a combined database of predicted ORFs identified using the same Python script in the transcriptomes of five actiniarian species. These species included *Nemanthus annamensis*, *Telmatactis* sp., *Anthopleura*

*buddemeieri*, *Aulactinia veratra*, and *Actinia tenebrosa* (refer to Supplementary Table S1 for SRA accessions and transcriptome assembly quality scores). Queried *C. polypus* ORFs that obtained confidently predicted homologues (as defined above) in some or all of the other five actinarians were assumed to represent previously uncharacterised protein-coding regions.

Using the aforementioned database of predicted actinarian ORFs in conjunction with the nr protein database and a previously assembled *C. polypus* transcriptome (see Supplementary Table S1 for SRA accession and transcriptome assembly quality score), further detailed assembly refinement was performed. This was primarily concerned with manually assembling sequence fragments that had not been correctly assembled by Trinity, the removal of intron sequences in contigs for which we believed evidence suggested no intron should be present, as well as identifying nucleotide insertions or deletions resulting in ORF interruption. The methodology of this is outlined in Supplementary Data 1 (3. assembly\_refinement; 4. in\_depth\_investigation).

**Time point categorisation.** Contigs with expression levels that were only found to be statistically different when comparing treatment time points (3 hps, 20 hps, and 96 hps) to each other rather than to control (0 hps) were subjected to further assessment to determine when the contig could be said to differ from physiological baseline expression. To visualise this, scatterplots were created which allowed for inference of physiological baseline expression. This process is detailed further in the Supplementary Methods.

**Functional annotation of contigs.** Contigs containing putatively conserved ORFs were functionally annotated (Fig. 2) with reference to proteins in the Swiss-Prot database<sup>22</sup> when confidently predicted homologues were found. For all other sequences that could not be annotated by Swiss-Prot comparison, translated ORFs had protein features predicted to enable inference of function. This consisted of Pfam<sup>23</sup> domain prediction (29.0 December 2015 database), as well as features predicted by SMART<sup>24</sup> which included signal peptides (SignalP 4.1)<sup>25</sup>, transmembrane regions (TMHMM 2.0)<sup>26</sup>, and internal repeats (Prospero)<sup>27</sup>. Contigs that lacked putatively conserved ORFs were compared with a previously assembled *C. polypus* transcriptome (as mentioned above) to find similarly assembled contigs using BLAST, in order to determine the likelihood of these contigs representing genuine biological transcripts. In this, we used stricter parameters than that described above for a “confident” match, requiring the contigs to be similar in overall length, with little to no mismatches and no gap openings outside of the extreme 5′ and 3′ ends.

## Results

**Transcriptome assembly and differential gene expression.** A total of 762 million single-end reads originating from 12 samples were generated from this study (SRA accession SRP091618). The *de novo* reference transcriptome consisted of 252, 263 contigs with an N50 of 1,167 bp and median contig length of 385 bp. BUSCO completeness scoring of the *de novo* assembly found 91% of the 843 core metazoan genes (BUSCO summarised benchmark = C: 91% [D: 39%], F: 4.6%, M: 3.6%, n: 843). Individual transcriptomes for each of the 12 samples were also assembled and assessed for completeness using BUSCO, the results of which are presented in Supplementary Table S2.

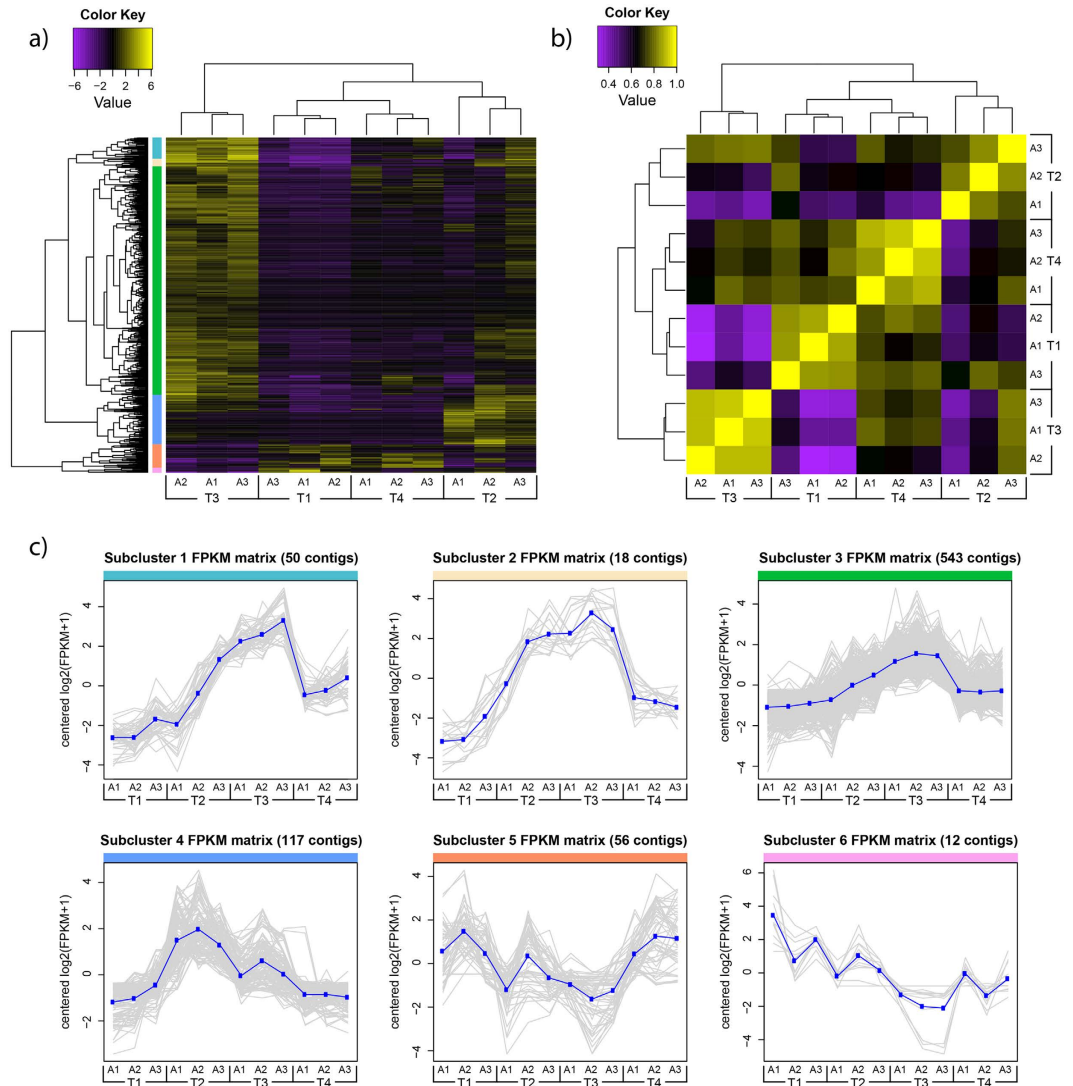
Differential gene expression analysis revealed 796 contigs whose abundance varied over the course of the experiment (Fig. 3a). Sample correlation (Fig. 3b) indicated that time points T1 (0 hps) and T3 (20 hps) were most dissimilar in their contig expression pattern, with T1 and T4 (96 hps) being most similar. Subclustering of differentially expressed contigs with similar RNA-Seq expression patterns was performed by cutting the dendrogram at 50% of its height (Fig. 3a), resulting in 6 subclusters (Fig. 3c).

**Assembly refinement and open reading frame identification.** After assembly refinement and redundancy removal, 489 unique differentially expressed contigs remained. Assembly refinement identified 23 incompletely assembled contigs which were manually assembled into eight contiguous transcripts. In addition, we identified four sequences with intron retentions, one misassembled sequence with artefact nucleotide insertion, and one sequence that was full length in an individual transcriptome but not in the combined reference assembly. Through this process, we also removed three putative contaminant contigs. Details of these contigs can be found in Supplementary Data 1 (3. assembly\_refinement; 4. in\_depth\_investigation).

Through the process of ORF identification, 208 contigs contained ORFs with significant identity to sequences in the nr protein database or predicted ORFs from the transcriptomes of the five sea anemone species used for comparison (*N. annamensis*, *Telmatactis* sp., *A. buddemeieri*, *A. veratra*, and *A. tenebrosa*), leaving 281 without a predicted ORF. Of the 208 contigs with predicted ORFs, 98 had been compared to predicted ORFs from the five sea anemone transcriptomes to assist in the prediction of the *C. polypus* ORF (refer to Supplementary Data 1 [5. anemone\_orfs]). Many of these 98 contigs did not return sequences with significant E-values ( $\leq 1e^{-3}$ ) when submitted as BLAST queries to the nr protein database. From this we can conclude that these represent ORFs that are novel and likely to be lineage-specific, and suggest that many of these may have been missed during genome annotations of *N. vectensis* and *E. pallida*.

**Time point categorisation.** The 489 differentially expressed contigs were categorised as being up or down-regulated at 3 hps, 20 hps, and 96 hps when compared to physiological baseline expression. Contigs that did not statistically differ when compared to control expression were assessed to determine if we could categorise them as being up or downregulated when compared to a proposed physiological baseline expression (as detailed in the Supplementary Methods). Four contigs could not be categorised due to their variable expression profile across time points and large variance in expression levels between replicates. These were not considered in downstream analyses, resulting in a final group of 485 contigs.

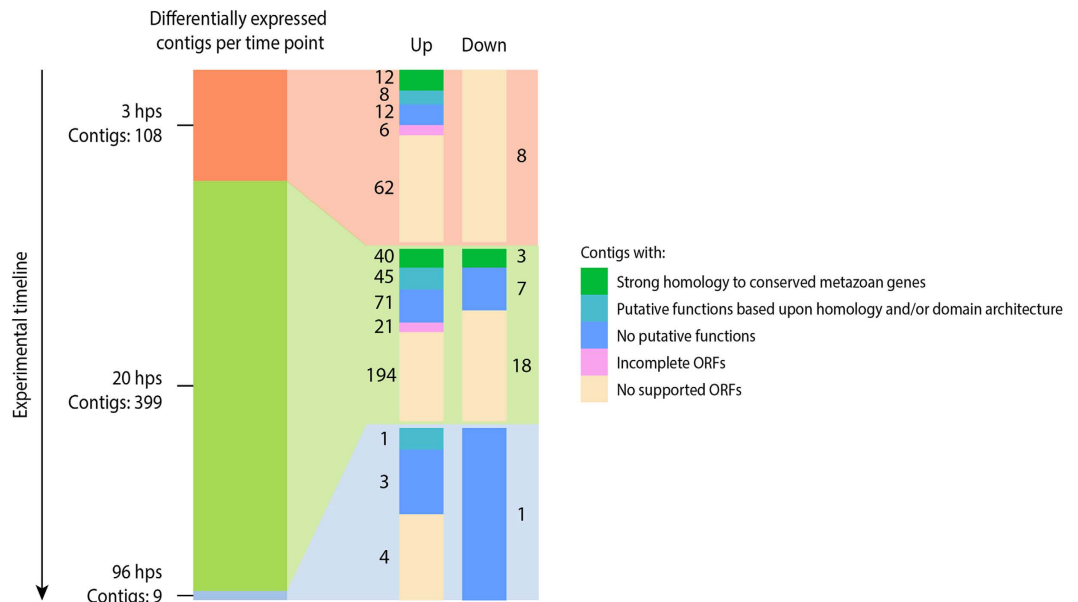
Of the 485 differentially expressed contigs, 108, 399, and 9 contigs were categorised as being differentially expressed at 3 hps, 20 hps, and 96 hps, respectively (Fig. 4). Notably, the majority of contigs were upregulated as



**Figure 3. Heat map and sample correlation matrix of differentially expressed genes following injury of *Calliactis polypus*.** (a) Heat map showing the RNA-Seq expression levels of 796 differentially expressed contigs ( $\log_2 |FC| \geq 2$ ,  $FDR \leq 0.01$ ,  $n = 3$ ) across the four time points. Expression values are  $\log_2$ -transformed median-centred FPKM. Yellow and purple colour intensity indicate contig upregulation and downregulation, respectively. Dendrogram clustering on the X-axis indicates sample similarity, whereas dendrogram clustering on the Y-axis groups contigs with similar expression profiles over time. Coloured bars on the Y-axis correspond to subclusters shown in (c). (b) Sample correlation matrix displaying the overall similarity in expression profile for each sample at the four different time points. Yellow colour intensity indicates increasing sample correlation, whereas purple colour intensity indicates decreasing sample correlation. Dendrogram clustering on the X and Y-axis indicates the overall similarity of these samples. (c) Contigs with similar RNA-Seq expression patterns were subclustered by cutting the Y-axis dendrogram of (a) at 50% of its height, resulting in 6 subclusters. Each contig's expression, measured in  $\log_2$ -transformed median-centred FPKM, is plotted in grey, with the mean expression profile for each cluster plotted in blue. Coloured bars lining the top of each graph correspond to the sections indicated on (a). T1–T4; time 1 – time 4 (0 hrs, 3 hrs, 20 hrs, 96 hours post sectioning), A1–A3; anemone 1 – anemone 3, FPKM; fragments per kilobase of transcript per million mapped reads.

opposed to downregulated. Moreover, contig differential expression increased greatly from 3 hps to 20 hps, with the number of differentially expressed contigs decreasing markedly by 96 hps. The group of contigs which contained identifiable protein-coding ORFs (208 in total) could be divided into two similarly sized groups with and without annotated functions. The 281 contigs lacking an identifiable protein-coding ORF thus represent a large proportion of the differentially expressed contigs, for which no functional roles can be inferred.

**Functional annotation of contigs. Genes of interest.** Functional annotation of the differentially expressed contigs identified a number of genes with potentially important roles in wound healing and regeneration within *C. polypus*. This included many genes with significant identity to known transcription factors involved in tissue patterning, such as *Sox E1*, *Bicaudal-C homologue 1 (BICC1)*, *Doublesex* and *mab-3-related transcription*



**Figure 4. Number and type of differentially expressed contigs at the three treatment time points when compared to baseline expression.** The leftmost bar represents the number of differentially expressed contigs at its labelled time point, with the rightmost paired bars representing the number of contigs with relation to their expression status (up or down) and the type of contig as noted in the legend. hps; hours post sectioning, ORF; open reading frame.

factors *A* and *F* (*DMRT A/F*), *Pax D2*, and *Nuclear factor 1 X-type (NF1X)*, as well as uncharacterised proteins with domain architecture indicative of transcription factor activity (PF00170: bZIP domain, PF01342: SAND domain, zinc finger domains with high identity to known zinc finger transcription factor domains). In addition to these, we also identified genes involved in fibroblast growth factor (FGF) signalling, namely a FGF protein and the FGF feedback inhibitor *Sprouty*.

Stress response and wound healing genes were represented in our study, including *Biliverdin reductase A*, *DnaJ-like 1*, and *Heat shock factor 1*, and *Hspa12B*.

Genes involved in cellular energy dynamics included *Solute carrier family 25 member 36-like*, *Graves' disease carrier protein*, *Phosphoenolpyruvate carboxykinase [GTP]*, *LYR motif-containing protein 5*, *Thiopurine S-methyltransferase*, *Amidophosphoribosyltransferase*, and *PAICS*.

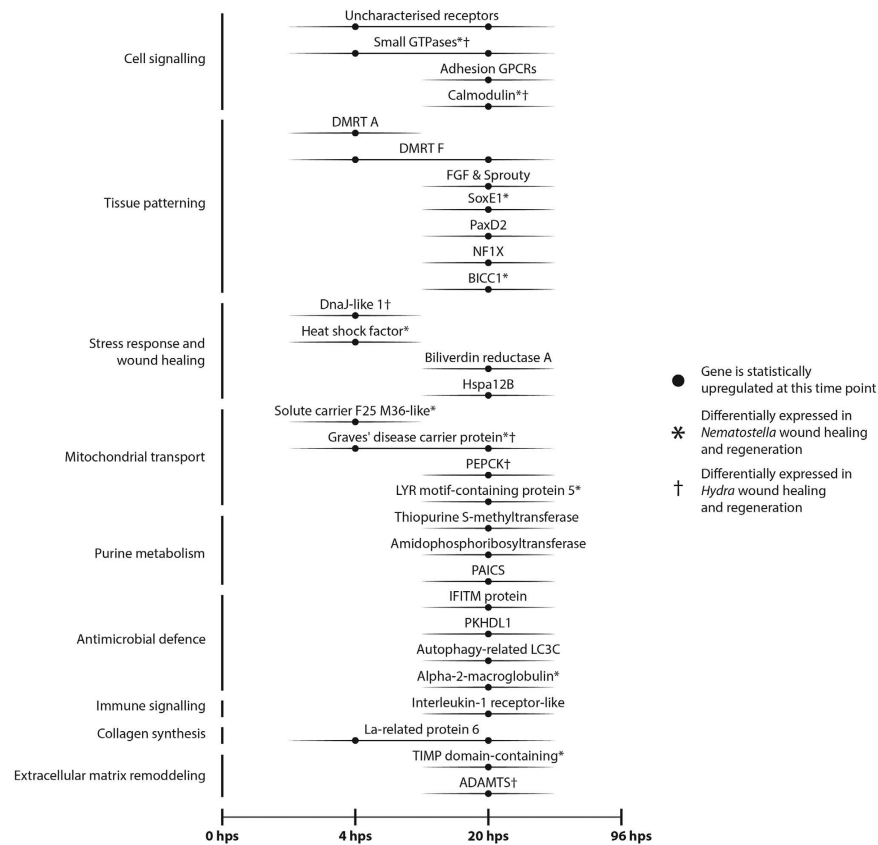
Genes proposed to be involved in immunity were differentially expressed, including *Alpha-2-macroglobulin (A2M)*, two putative interleukin-1 receptor-like genes, and a number of genes with putative antimicrobial roles such as *Autophagy-related protein LC3C*, *Fibrocystin-like gene (PKHDL1)*, as well as those containing domains commonly associated with immune responses (PF00530: Scavenger receptor cysteine-rich, PF04505: Interferon-induced transmembrane protein).

Genes related to cell signalling also represented a large proportion of functionally annotated contigs, including two adhesion G protein-coupled receptors, three Ras-related small GTPases, the calcium signalling molecule *Calmodulin*, and eight uncharacterised proteins with receptor-related protein domain (PF00001: 7 transmembrane receptor [rhodopsin family]), five of which shared moderate identity to known neurotransmitter receptors annotated on UniProtKB/Swiss-Prot (e.g., dopamine receptors, alpha and beta adrenergic receptors, galanin receptors, among others).

Finally, genes involved in extracellular matrix-related functions were identified, which included a putative cnidarian homologue of the type 1 collagen synthesising *La-related protein 6 (LARP6)*, in addition to three tissue inhibitors of metalloproteinase (TIMP) and four A disintegrin and metalloproteinase with thrombospondin motif (ADAMTS) proteins.

The differential expression patterns of these genes are summarised in Fig. 5, and the full annotation report is available as Supplementary Data 1 (2. annotation\_report).

**Uncharacterised genes.** The contig with the largest fold change when compared to control, as well as the highest overall log<sub>2</sub>-transformed median-centred expression level (three replicate average  $\approx 3.85$ ), was an upregulated, uncharacterised sterile alpha motif domain-containing protein (PF07647: SAM domain). This was followed by the immune-related *A2M*, which had the third largest fold change when compared to control and second highest overall log<sub>2</sub>-transformed median-centred expression level (three replicate average  $\approx 3.67$ ). The remaining contigs in the top 10 according to either of these two metrics (i.e., fold change compared to control, and log<sub>2</sub>-transformed median-centred expression level at a specific time point) were all uncharacterised sequences without putative functions. It is worth noting, however, that within the top 10 of these two metrics were a group of uncharacterised genes with a predicted protein domain which is typically associated with prokaryotic enterotoxins (PF05791:



**Figure 5. Key candidate genes identified through our differential gene expression study likely to be important for wound healing and regeneration.** Note that no contigs with putative functions were found to be differentially expressed at 96 hours post sectioning (hps). Genes noted with an asterisk (\*) or dagger (†) were also differentially expressed in a study of *N. vectensis*<sup>9</sup> or *Hydra*<sup>28</sup>, respectively. Abbreviations - Adhesion GPCR; Adhesion G protein-coupled receptor, DMRT A/F; Doublesex and mab-3 related transcription factor A/F, FGF; Fibroblast growth factor, Sox E1; Sry-related HMG box E1, PaxD2; Paired box D2, NF1X; Nuclear factor 1X-type, BICC1; Bicaudal-C homologue 1, Hspa12B; Heat shock protein family A member 12B, Solute carrier F25 M36-like; Solute carrier family 25 member 36-like, PEPCK; Phosphoenolpyruvate carboxykinase, PAICS; Phosphoribosylaminoimidazole carboxylase/phosphoribosylaminoimidazole succinocarboxamide synthetase, IFITM protein; Interferon induced transmembrane protein, PKDHL1; Polycystic kidney and hepatic disease-like 1 (or 'Fibrocystin-like'), Autophagy-related LC3C; Autophagy-related protein light chain 3 C, TIMP domain-containing; Tissue inhibitors of metalloproteinase domain-containing, ADAMTS; A disintegrin and metalloproteinase with thrombospondin motifs.

*Bacillus haemolytic enterotoxin*). This domain, to our knowledge, has not been previously annotated in any eukaryotic proteins on the nr protein database (as of 16/05/16) with the exception of two proteins in *E. pallida* (KXJ22104, KXJ22064) and one in *Acropora digitifera* (XP\_015750752), both of which are cnidarians. Based upon preliminary bioinformatic evidence, these are not likely to represent contaminants in our data (refer to Supplementary Data 1 [4. in\_depth\_investigation]).

For those contigs where we could not identify putatively conserved ORFs (281), we found 71 similarly assembled contigs in a previously assembled *C. polypus* transcriptome, representing approximately 25% of this group of contigs. Refer to Supplementary Data 1 (7. no\_orf\_check) for a list of these contigs.

## Discussion

In our study we have compiled a rich dataset and identified a large number of candidate genes involved in wound healing and regeneration in the sea anemone *C. polypus*. Manual annotation of contigs reveals that conserved genes from various metabolic pathways are differentially expressed in *C. polypus* after wounding. These include genes related to tissue patterning and embryogenesis, stress response, energy dynamics, the immune system, cell signalling, and extracellular matrix (ECM) remodelling (Fig. 5). Most of these genes were upregulated after wounding, with very few downregulated genes identified to which we could attribute functional roles (Fig. 4). We also observe a large number of uncharacterised, putatively lineage-specific genes, that could not be functionally annotated (Fig. 4). We propose that wound healing and regeneration in *C. polypus* is enabled by a combination of conserved genes as well as genes that are specific to Cnidaria and, possibly, to Actiniaria.

**Time point selection.** Overall, the time points we have selected for sampling have provided us with baseline data related to the genes involved in *C. polyopus* regeneration. Sampling at 3 hps was chosen due to secretion of a mucus layer over the wound. While we don't find evidence of changes in expression of transcripts typically considered to be involved in mucus secretion at 3 hps, we do identify a mucin-like gene that is downregulated at 20 hps. This gene may be involved in the production of mucus within the first few hours of *C. polyopus* regeneration. We find that 20 hps represents a time of heightened transcriptomic activity, characterised by genes involved in a variety of different cellular pathways. At 96 hps, we did not find any transcripts that could explain the observed pulsating behaviour. We do, however, note that baseline expression levels are largely restored by this time point (Fig. 4). This is significant, as the morphology of the organism indicates that wound healing and regeneration are still occurring after 96 hps (Supplementary Figs S1 and S2). Based on this observation, we hypothesise that a transition from transcriptomic activity to proteomic activity may occur between 20 hps and 96 hps. If this is the case, future studies of cnidarian regeneration should use proteomic techniques in combination with RNA-Seq analysis.

**Conserved cnidarian regenerative properties.** Comparison of our *C. polyopus* study to two previous studies<sup>9,28</sup> of cnidarian regeneration using transcriptomic approaches revealed key mechanisms that appear to be conserved throughout Actiniaria and Cnidaria (see Supplementary Data 1 [6. comparison\_regen\_studies]). Unsurprisingly, the differentially expressed genes we identify in *C. polyopus* most closely correspond to the gene cohort reported by DuBuc *et al.*<sup>9</sup> in *N. vectensis* regeneration, with a smaller overlap when compared to Petersen *et al.*<sup>28</sup> study of *Hydra*. Genes found to be differentially expressed in all three species include a Ras-related small GTPase as well as *Calmodulin*, highlighting the role of intracellular signal transduction during regeneration. We also note that *Calmodulin* is upregulated in rat liver regeneration models<sup>29</sup>, which indicates a potentially conserved role of *Calmodulin* in regenerative processes throughout Eumetazoa. Finally, we find a putative homologue of the *Graves' disease carrier protein* which, based upon evidence from human and yeast models<sup>30</sup>, might play an important role in accumulating coenzyme A in the mitochondria of these three animals. Genes differentially expressed in *C. polyopus* and either *N. vectensis* or *Hydra* also include those related to tissue patterning, stress responses, mitochondrial transport, antimicrobial defence, and ECM remodelling (Fig. 5).

The genes noted in Fig. 5 are not the only genes we find in common between *C. polyopus* and *N. vectensis* or *Hydra*. Other contigs without ascribed functions include a putative *Protein phosphatase inhibitor-2* and a cohort of uncharacterised genes. One such uncharacterised gene contains a PF00170: bZIP domain, suggestive of transcription factor activity.

**Overview of genes involved in *Calliactis polyopus* regeneration.** *Early response.* Contigs identified in subclusters 2 and 4 (Fig. 3c) appear to represent a group of genes whose expression increases immediately following severe wounding. A high proportion of these genes are putative transcription factors or are involved in cellular stress responses, such as heat shock proteins. The early upregulation of stress response genes is expected given the severity of wounding during the experiment. With respect to transcription factors, we hypothesise that many of these may coordinate the regulation of genes differentially expressed at 20 hps, which are represented in subclusters 1, 3, and 6 (Fig. 3c).

Genes involved in the early response of *C. polyopus* regeneration also appear to be highly upregulated. Subcluster 2 contains many of the uncharacterised genes within the top 10 of the two expression metrics mentioned in the Results, which includes the SAM domain-containing protein and two of the three *Bacillus* haemolytic enterotoxin (PF05791) domain-containing proteins. While it is difficult to assign any specific function to these genes, we postulate that genes that are highly upregulated immediately and remain elevated later into the regeneration response may be of particular importance for *C. polyopus* and, more generally, cnidarian regeneration. Subsequently, we believe that these genes in particular deserve further attention, as they may provide unanticipated insights into cnidarian regeneration.

*Later response.* The largest number of differentially expressed contigs was observed at 20 hps (subclusters 1, 3, and 6). While most of the cellular pathways identified in the transcriptomic response of *C. polyopus* at this time point are also reported in other cnidarian models (Fig. 5), there are certain cellular pathways and genes of interest which our results suggest may hold additional significance. The genes and pathways of interest include tissue patterning, mitochondrial transport and energy dynamics, innate immune system and ECM remodelling genes.

A number of upregulated tissue patterning genes were identified 20 hps, including *DMRT A* and *F*, for which evidence suggests the paralog *DMRT B* is involved in neurogenesis<sup>31</sup>. Additionally, we find the gene *Pax-D2* to be upregulated, which in *N. vectensis* is involved in neurogenesis<sup>32</sup>. Furthermore, we identify a FGF protein and the FGF feedback inhibitor *Sprouty* as upregulated, both of which are expressed in *N. vectensis* embryos and are thought to be involved in endoderm formation and neurogenesis<sup>33</sup>. Taken together, these genes indicate that neurogenesis is occurring during *C. polyopus* regeneration. In consideration of these and other genes involved in coordinating tissue patterning processes that are upregulated at 20 hps (i.e., *Sox E1*, *BICCI*, *NFIX*), we also propose that *C. polyopus* may be reverting back to an embryogenesis-like state, wherein whole-body tissue patterning processes are recapitulated to enable regeneration.

Genes involved in mitochondrial transport and energy dynamics appear to play a role in *C. polyopus* regeneration (Fig. 5). While the relationship between mitochondrial energy production and regeneration has not been well established in the literature, the results of our study and another recently published study on neuronal axon regeneration<sup>34</sup> suggest that understanding this relationship may be important for understanding regeneration as a whole. Further genes that indicate a change in *C. polyopus* energy dynamics include those involved in purine metabolism. We observe the upregulation of two genes involved in *de novo* purine synthesis, including *Amidophosphoribosyltransferase* and *PAICS*. It is noteworthy that we do not find genes involved in salvage



synthesis of purines, the method of purine synthesis typically employed in cells, to be differentially expressed. Increases in *de novo* synthesis of purines have been observed within *Xenopus* tadpole tail regeneration<sup>35</sup> as well as in rat liver regeneration<sup>36</sup>. Given its implication in changes from oxidative phosphorylation to glycolysis to meet energetic demands<sup>37</sup>, transcriptomic data of *C. polypus* regeneration might indicate a shift towards glycolysis. This has been proposed in the *Xenopus* regeneration model<sup>35</sup> which, in conjunction with our findings, suggests a conserved role for *de novo* purine synthesis and glycolysis in eumetazoan regeneration.

Genes with a role in the innate immune system appear to have a slower activation in *C. polypus* in response to injury. Two interleukin-1 receptor-like proteins, involved in coordinating the activation of the innate immune system<sup>38</sup>, are upregulated at 20 hps. We also see *A2M* upregulated, a gene that was differentially expressed in DuBuc *et al.*'s<sup>9</sup> study on *N. vectensis* but not in Petersen *et al.*'s<sup>28</sup> study of *Hydra*. Due to its high level of differential expression, *A2M* is likely to have an important role in modulating the immune response during regeneration in actinarians.

In our study, ECM remodelling genes also constitute a major cellular pathway observed in the later transcriptomic response of *C. polypus*. Representatives of this pathway include TIMPs as well as ADAMTS genes, both of which represent gene families with various functions in ECM remodelling<sup>39,40</sup>. In addition, we report up-regulation of the cnidarian homologue of *LARP6*. This gene has been shown to result in an increase in type I collagen protein synthesis without affecting type I collagen mRNA abundance<sup>41</sup>. The potential increase in synthesis of type I collagen indicated in the transcriptome of *C. polypus* is in line with a previous histological study of the related *C. parasitica*, where increased amounts of collagen filaments were observed during wound healing<sup>8</sup>. Thus, while we do not detect any differentially expressed collagen transcripts in our study, our results corroborate this observation.

**The role of uncharacterised genes in regeneration and wound healing.** Our study has revealed a high proportion of differentially expressed contigs with uncharacterised ORFs (Fig. 4). These uncharacterised ORFs likely represent important lineage specific genes involved in regeneration in cnidarians. An increased scientific effort to functionally characterise the many genes found only in cnidarians<sup>42,43</sup> will be necessary for future efforts to understand regeneration in this phylum.

It is also important to note the large number of contigs (281) for which we could not find conserved ORFs. In an attempt to discern whether these contigs were likely to represent either sequencing artefacts or biologically valid transcripts, we found that 71 (~25%) of these had good matches to a previously assembled *C. polypus* transcriptome. Our hypothesis is that many of these contigs are likely to encode small ORFs. The study of small ORFs represents a relatively new field that has only recently been explored in animals and plants utilising ribosomal profiling techniques<sup>44,45</sup>. Future studies on cnidarian regeneration utilising ribosomal profiling may provide clues into the role of small ORF-coding transcripts in regeneration and wound healing processes.

**Limitations.** It should be mentioned that a limitation of our study was that a sample collected 3 hps contained some degree of ribosomal contamination, which resulted in a poorer representation of genes and may have impacted the statistical ability to detect differentially expressed genes at this time point (refer to Supplementary Table S2). Importantly, a limitation in non-model species is that results often rely on inferring homology to genes studied in models such as human, mouse, or zebrafish. Given the large evolutionary distance between cnidarians and these models (cnidarians originated ~686–819 million years ago)<sup>46</sup> it would not be surprising if some of the functions we have tentatively attributed to certain genes show some divergence.

## Conclusion

In conclusion, our results indicate that wound healing and regeneration in *C. polypus* involves a large-scale response at the transcriptomic level. Many of the genes identified in our study can be broadly classified into key gene groups, such as those involved in tissue patterning, stress responses, cellular energy dynamics, the immune system, and ECM remodelling. For the sake of brevity, we have only discussed certain genes which provide an encompassing view of the overall processes likely to be involved in *C. polypus* regeneration. From this, we have been able to generate a number of important hypotheses for future regeneration research to investigate in both cnidarians and other animal groups, the results of which should deepen our understanding of how regeneration occurs in certain lineages.

## References

1. Bullard, K. M., Longaker, M. T. & Lorenz, H. P. Fetal wound healing: current biology. *World J. Surg.* **27**, 54–61 (2003).
2. Reinke, J. M. & Sorg, H. Wound repair and regeneration. *Eur. Surg. Res.* **49**, 35–43 (2012).
3. Bely, A. E. & Nyberg, K. G. Evolution of animal regeneration: re-emergence of a field. *Trends Ecol. Evol.* **25**, 161–170 (2010).
4. Bosch, T. C. G. Why polyps regenerate and we don't: towards a cellular and molecular framework for Hydra regeneration. *Dev. Biol.* **30**, 421–433 (2007).
5. Newmark, P. A. & Alvarado, A. S. Not your father's planarian: a classic model enters the era of functional genomics. *Nat. Rev. Genet.* **3**, 210–219 (2002).
6. Gardiner, D. M., Endo, T. & Bryant S. V. The molecular basis of amphibian limb regeneration: integrating the old with the new. *Semin. Cell Dev. Biol.* **13**, 345–352 (2002).
7. Brockes, J. P. & Kumar, A. Comparative aspects of animal regeneration. *Annu. Rev. Cell Dev. Biol.* **24**, 525–549 (2008).
8. Young, J. A. C. The nature of tissue regeneration after wounding in the sea anemone *Calliactis parasitica* (Couch). *J. Mar. Biol. Assoc. UK* **54**, 599–617 (1974).
9. DuBuc, T. Q., Traylor-Knowles, N. & Martindale, M. Q. Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. *BMC Biol.* **12**, 24 (2014).
10. Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
11. Hejnol, A. *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. Biol. Sci.* **276**, 4261–4270 (2009).
12. Rodríguez, E. *et al.* Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of hexacorals. *PLoS One* **9**, e96998 (2014).

13. Schaffer, A. A., Bazarsky, M., Levy, K., Chalifa-Caspi, V. & Gat, U. A transcriptional time-course analysis of oral vs. aboral whole-body regeneration in the sea anemone *Nematostella vectensis*. *BMC Genomics* **17**, 718 (2016).
14. Baumgarten, S. *et al.* The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc. Natl. Acad. Sci. USA* **112**, 11893–11898 (2015).
15. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
16. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
17. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
18. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
19. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
20. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
22. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
23. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
24. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43**, D257–D260 (2015).
25. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
26. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
27. Mott, R. Accurate formula for *P*-values of gapped local sequence and profile alignments. *J. Mol. Biol.* **300**, 649–659 (2000).
28. Petersen, H. O. *et al.* A comprehensive transcriptomic and proteomic analysis of *Hydra* head regeneration. *Mol. Biol. Evol.* **32**, 1928–1947 (2015).
29. Agell, N. *et al.* Calmodulin expression during rat liver regeneration. *Hepatology* **20**, 1002–1008 (1994).
30. Prohl, C. *et al.* The yeast mitochondrial carrier Leu5p and its human homologue Graves' disease protein are required for accumulation of coenzyme A in the matrix. *Mol. Cell Biol.* **21**, 1089–1097 (2001).
31. Parlier, D. *et al.* The *Xenopus* doublesex-related gene *Dmrt5* is required for olfactory placode neurogenesis. *Dev. Biol.* **373**, 39–52 (2013).
32. Matus, D. Q., Pang, K., Daly, M. & Martindale, M. Q. Expression of Pax gene family members in the anthozoan cnidarian, *Nematostella vectensis*. *Evol. Dev.* **9**, 25–38 (2007).
33. Matus, D. Q., Thomsen, G. H. & Martindale, M. Q. FGF signaling in gastrulation and neural development in *Nematostella vectensis*, an anthozoan cnidarian. *Dev. Genes Evol.* **217**, 137–148 (2007).
34. Zhou, B. *et al.* Facilitation of axon regeneration by enhancing mitochondrial transport and rescuing energy deficits. *J. Cell Biol.* **214**, 103–119 (2016).
35. Chen, Y., Love, N. R. & Amaya, E. Tadpole tail regeneration in *Xenopus*. *Biochem. Soc. Trans.* **42**, 617–623 (2014).
36. Itakura, M., Maeda, N., Tsuchiya, M. & Yamashita, K. Increased rate of de novo purine synthesis and its mechanism in regenerating rat liver. *Am. J. Physiol. Gastrointest. Liver Physiol.* **251**, G585–G590 (1986).
37. Zhao, H., French, J. B., Fang, Y. & Benkovic, S. J. The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *Chem. Commun.* **49**, 4444–4452 (2013).
38. van der Burg, C. A., Prentis, P. J., Surm, J. M. & Pavašovic, A. Insights into the innate immunome of actinians using a comparative genomic approach. *BMC Genomics* **17**, 850 (2016).
39. Brew, K. & Nagase, H. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochim. Biophys. Acta* **1803**, 55–71 (2010).
40. Kelwick, R., Desanlis, I., Wheeler, G. N. & Edwards, D. R. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol.* **16**, 113 (2015).
41. Manojlovic, Z. & Stefanovic, B. A novel role of RNA helicase A in regulation of translation of type I collagen mRNAs. *RNA* **18**, 321–334 (2012).
42. Babonis, L. S., Martindale, M. Q. & Ryan, J. F. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC Evol. Biol.* **16**, 114 (2016).
43. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. G. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25**, 404–413 (2009).
44. Bazzini, A. A. *et al.* Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **33**, 981–993 (2014).
45. Hellens, R. P., Brown, C. M., Chisnall, M. A. W., Waterhouse, P. M. & Macknight, R. C. The emerging world of small ORFs. *Trends Biochem. Sci.* **21**, 317–328 (2016).
46. Park, E. *et al.* Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Mol. Phylogenet. Evol.* **62**, 329–345 (2012).

## Acknowledgements

We acknowledge the National Council for Scientific and Technological Development (Brazil) for providing scholarship to Daniella H. Hock to study at the Queensland University of Technology and contribute towards this project's work.

## Author Contributions

P.J.P., A.P., Z.K.S. and D.H.H. conceived the idea and contributed to the design of the study. Z.K.S. and D.H.H. conducted the experimental work and performed the bioinformatic analysis. Z.K.S. wrote the main manuscript text and prepared figures. P.J.P. and A.P. guided all aspects of the project in their capacity as supervisors. All authors reviewed the manuscript.

## Additional Information

**Accession codes:** Raw sequencing reads generated from this study are deposited at the NCBI sequence read archive (SRA) under accession SRP091618.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Stewart, Z. K. *et al.* Transcriptomic investigation of wound healing and regeneration in the cnidarian *Calliactis polypus*. *Sci. Rep.* 7, 41458; doi: 10.1038/srep41458 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017