

OPEN

The *Eupentacta fraudatrix* transcriptome provides insights into regulation of cell transdifferentiation

Alexey V. Boyko^{1*}, Alexander S. Girich¹, Ekaterina S. Tkacheva¹ & Igor Yu. Dolmatov^{1,2}

The holothurian *Eupentacta fraudatrix* is a unique organism for studying regeneration mechanisms. Moreover, *E. fraudatrix* can quickly restore parts of its body and entire organ systems, yet at the moment, there is no data on the participation of stem cells in the process. To the contrary, it has been repeatedly confirmed that this process is only due to the transformation of terminally differentiated cells. In this study, we examine changes in gene expression during gut regeneration of the holothurian *E. fraudatrix*. Transcriptomes of intestinal anlage of the three stages of regeneration, as well as the normal gut, were sequenced with an Illumina sequencer (San Diego, CA, USA). We identified 14,617 sea urchin protein homologs, of which 308 were transcription factors. After analysing the dynamics of gene expression during regeneration and the map of biological processes in which they participate, we identified 11 factors: Ef-EGR1, Ef-ELF, Ef-GATA3, Ef-ID2, Ef-KLF1/2/4, Ef-MSX, Ef-PCGF2, Ef-PRDM9, Ef-SNAI2, Ef-TBX20, and Ef-TCF24. With the exception of TCF24, they are all involved in the regeneration, development, epithelial-mesenchymal transition, and immune response in other animals. We suggest that these transcription factors may also be involved in the transdifferentiation of coelomic epithelial cells into enterocytes in holothurians.

Regeneration is a unique phenomenon in which the origin, evolution, and mechanisms are still poorly understood. In particular, stem cells are currently considered the main cellular source for most animals¹. At the same time, even in the most primitive multicellular animals, such as sponges, regeneration can occur not only with the involvement of stem cells (archeocytes), but through the dedifferentiation of specialized cells, choanocytes^{2,3}. In humans, with damage to the liver, lungs, and pancreas, specialized cells of these organs can dedifferentiate, entering the mitotic cycle, and participating in their repair⁴.

In echinoderms, regeneration apparently occurs through dedifferentiation or transdifferentiation of specialized cells, which has been confirmed by many studies^{5–9}. There is, however, no direct evidence of the presence of stem cells in echinoderms or their involvement in morphogenesis¹⁰. In addition, there has recently been data that regeneration in these animals can occur with complete suppression of mitotic activity^{9,11}, which indicates participation of specialized cells of organ remnants, rather than stem cells. In connection with this, and the ability to regenerate, echinoderms are convenient model objects for studying the involvement of differentiated cells in regeneration.

Among echinoderms, holothurians are capable of restoring almost all organs and tissue types, and as such are of special interest^{5,12,13}. In addition, their unique feature is the ability to eject (eviscerate) their intestine in response to stressful stimuli; the lost part of the digestive tract regenerates by formation of two anlagen on the anterior and posterior parts of the body, which grow towards each other¹⁴. This process in most studied species of holothurians occurs as a function of the cells of luminal epithelia of the cloaca (posterior anlage) and the remaining part of the esophagus (anterior anlage), which is due to dedifferentiation, migration, proliferation, and differentiation^{15–18}.

Gut regeneration mechanisms in *E. fraudatrix* are even more unusual. In this species, evisceration occurs through the anterior part of the body, which is lost (the pharynx and esophagus) with the entire digestive

¹National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, Palchevsky 17, 690041, Vladivostok, Russia. ²Far Eastern Federal University, Suhanova 8, Vladivostok, 690950, Russia. *email: alteroldis@gmail.com

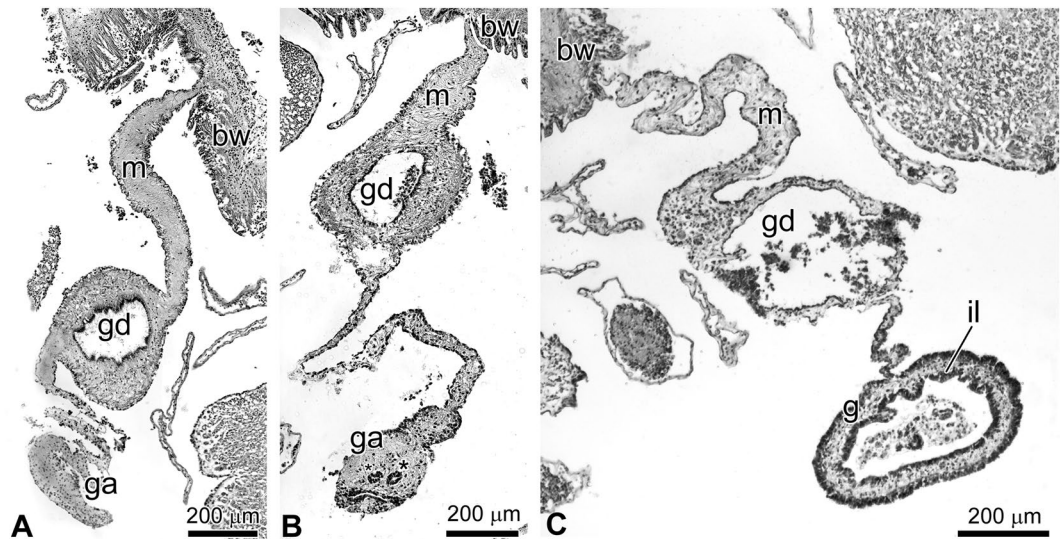


Figure 1. Structure of gut on different stages of regeneration in holothurian *Eupentacta fraudatrix*. (A) Connective-tissue thickening on the edge of mesentery on day 3 post-evisceration (first stage). (B) Gut anlage on day 6 post-evisceration (second stage); arrow indicates site of migration of coelomic epithelial cells into connective-tissue thickening, asterisks show clusters inside the intestinal anlage. (C) Digestive tube on day 12 post-evisceration (third stage). bw – body wall, g – digestive tube, ga – gut anlage, gd – gonoduct, il – intestinal lining, m – mesentery.

system^{19–21}. In this situation, the anterior part of the intestine is formed with cells from the coelomic epithelium, as a derivative of the mesoderm⁷. Coelomic epithelium in echinoderms mainly consists of peritoneocytes and myoepithelial cells²², and contains neurons and bundles of axons that form the basiepithelial nerve plexus. After evisceration, the peritoneal and myoepithelial cells of the coelomic epithelium of gut mesentery, as well as of the anterior part of the animal, dedifferentiate. This is manifested as a loss of specialized structures, e.g., bundles of tonofilaments (peritoneocytes) and myofilaments (myoepithelial cells). Myofilaments are assembled into specific spindle-like structures and gradually degrade in the cytoplasm, or are ejected into the intercellular space⁸. At the same time, dedifferentiated cells of the coelomic epithelium migrate to the damaged edge of the mesentery, forming a cluster of flattened epithelial cells; neurons and nerve processes are probably destroyed. In the process of regeneration, part of the dedifferentiated cells of the coelomic epithelium undergo transdifferentiation, mitotically divide, and transform into enterocytes⁷.

Until recently, *E. fraudatrix* was the only species for which transdifferentiation of mesodermal cells into enterocytes was reported⁷. Recently, the ability of muscle cells to transform into enterocytes has been found in zebrafish²³. It is not yet clear how similar the mechanisms of transdifferentiation in holothurians and fish actually are. One of the key genes that trigger transdifferentiation in zebrafish is *oct4*²³. The holothurian ortholog of this gene, *oct-1/2/11*, does not show noticeable expression during intestinal regeneration in *Holothuria glaberrima*²⁴. Moreover, the molecular basis of cell transdifferentiation mechanisms in *E. fraudatrix* remains unknown. In order to understand, to a first approximation, the possible molecular processes that occur during transdifferentiation in holothurians, we decided to use RNA-seq. In this work, we tried to obtain the complete transcriptome and determine the differentially expressed genes (DEGs) at the stage of active cell transdifferentiation and transcription factors (TFs), which are candidates for the role of transdifferentiation regulator.

Results and Discussion

Formation of digestive epithelium by coelomic epithelial cells. The morphological features of regeneration of internal organs in the holothurian *E. fraudatrix* have been well studied^{7,12,19–21}. Nevertheless, the successive stages of digestive tube formation are characterized only superficially and are interpreted somewhat differently in various publications^{7,19}. Thus, in our work, we conditionally distinguish three stages of anterior gut regeneration, depending on the phases of transdifferentiation of the coelomic epithelial cells.

The first stage (day 3 post-evisceration) is characterised by formation of thickening on the edge of the mesentery in the anterior part of the holothurians. This thickening represents the intestinal anlage, consisting of connective tissue (Fig. 1A), and is covered by the coelomic epithelial cells that migrate from the mesentery; numerous mesenchymal-like cells are visible inside. During the regeneration of *E. fraudatrix*, the connective-tissue thickening gradually, lengthens and spreads back along the mesentery edge.

The second stage of regeneration (days 5–7 post-evisceration) is characterised by active transdifferentiation of cells. On the ventral side of the anlage, coelomic epithelial cells form folds and migrate into the connective tissue with the formation of numerous cell clusters inside the intestinal anlage (Fig. 1B). During migration, the structure of cells changes, with fragments of tonofilaments and spindle-like structures finally disappearing from their cytoplasm^{7,12}. On days 8–9 post-evisceration, microvesicles and secretory vacuoles, characteristic of enterocytes, are found in their apical part¹². Transforming cells retain their intercellular junctions, i.e., they do not undergo the epithelial-mesenchymal transition^{7,12}.

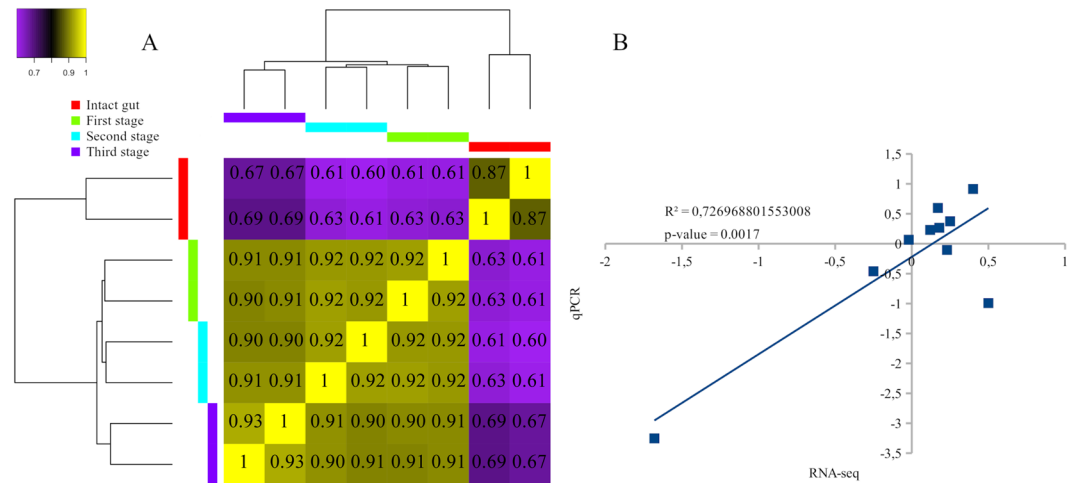


Figure 2. Correlation map of all RNA-seq samples (A) and gene expression correlation between RT-qPCR (y-axis) and RNA-seq (x-axis) data. (B) The squared Pearson correlation coefficient, p-value, and linear regression line for 5 genes are indicated on plot. (B) Each point on the plot (B) corresponds to the logarithm (base 2) of the fold change in the second or third stages of regeneration relative to the first stage.

The third stage of regeneration (days 10–12 post-evisceration) is characterised by the presence of well-formed luminal epithelium (Fig. 1C), due to merging of cell clusters. Subsequently, the anterior anlage continues to grow towards the posterior body end along the mesentery edge. The luminal epithelial cells gradually differentiate and acquire a structure characteristic of enterocytes. The unification of anlagen occurs towards the middle of the animal's body on days 25–27 post-evisceration, with the formation of a continuous digestive tube⁷.

De novo transcriptome assembly. As a result of sequencing of eight libraries, corresponding to the three stages of regeneration and the normal gut of the holothurian *E. fraudatrix*, a total of 413 million (412,999,924) raw paired-end reads were obtained. After filtering and trimming the adapters, 95% paired-end reads were retained with an average quality of 35.7 units by PhredScore 33, and an average read length of 99.4 nucleotides (see Supplementary Table S1). All of these reads, as well as the unpaired reads that remain after filtering and read error corrections, accounted for 5% of the initial number of reads, were used in the assembly. As a preassembly stage, two iterations of read error corrections were performed, reducing the number of paired-end reads. On average, 5% of paired reads are lost. As a result of the assembly, filtering of contaminant and non-coding sequences, and subsequent clustering to identify isoforms, a total of 83,960 transcript isoforms and 70,538 genes were obtained.

In addition, all sequences were removed in which the predicted protein matched the species protein which did not belong to deuterostomes. These sequences satisfied two conditions: the bitscore above 200; the lack of a match with the available protein sequences of *Apostichopus japonicus*, *Parastichopus parvimensis*, *Cladolabes schmelztzii*, *Patiria miniata*, *Lytechinus variegatus*, and *Strongylocentrotus purpuratus*^{25–27} with a bitscore, which was more than 80% of the match bitscore in the nonredundant protein (NRP) NCBI database. We assessed the completeness of the assembly using BUSCO v3²⁸. As result, 98.1% of the core metazoan genes (based on 978 core essential genes) and 83% of the single-copy orthologs among them, were identifiable in the transcriptome. Only 21% of reads were mapped back to the assembly. This value is underestimated, due to strict alignment parameters and the presence of only protein coding regions in the assembly.

Differential expression analysis. After aligning the reads to assembly and counted mapped events (see Supplementary Table S2), we performed a correlation analysis (Fig. 2A). We found a high correlation of replicates within samples, at almost 91%, while the correlation from the regeneration stages was significantly higher, versus the sample from an intact gut. Validation of expression levels of the 5 genes showed a mean correlation between qPCR and RNA-seq data (Fig. 2B).

As a result of the search for DEGs, we found a total of 17,227 and 15,342 genes with notable changes in expression levels for all samples, relative to an intact gut and the second stage, respectively (see Supplementary Data S2). Of these, 13,234 and 11,942 sequences, respectively, had significant hits in the NRP NCBI database (Fig. 3). Thus, most of the DEGs are detected when compared to the three stages of regeneration with the intact sample. The same is evidenced by the map of correlations between samples (Fig. 2A). The observation indicates essential changes in the work of genes after damage and, accordingly, the artificiality of comparing gene expression levels in regenerating and intact tissues. This could probably enable identification of the regulatory genes, whose expression is activated or repressed during regeneration. In this study, the emphasis was on finding the probable transdifferentiation regulators, rather than analysing the entire regeneration process. In this regard, only DEGs obtained in comparison to the second stage will be considered below. In addition, an obvious increase in the number of unique genes in the second stage, when transdifferentiation occurs, is interesting. All of the above only confirms expectations, but is not significant in itself.

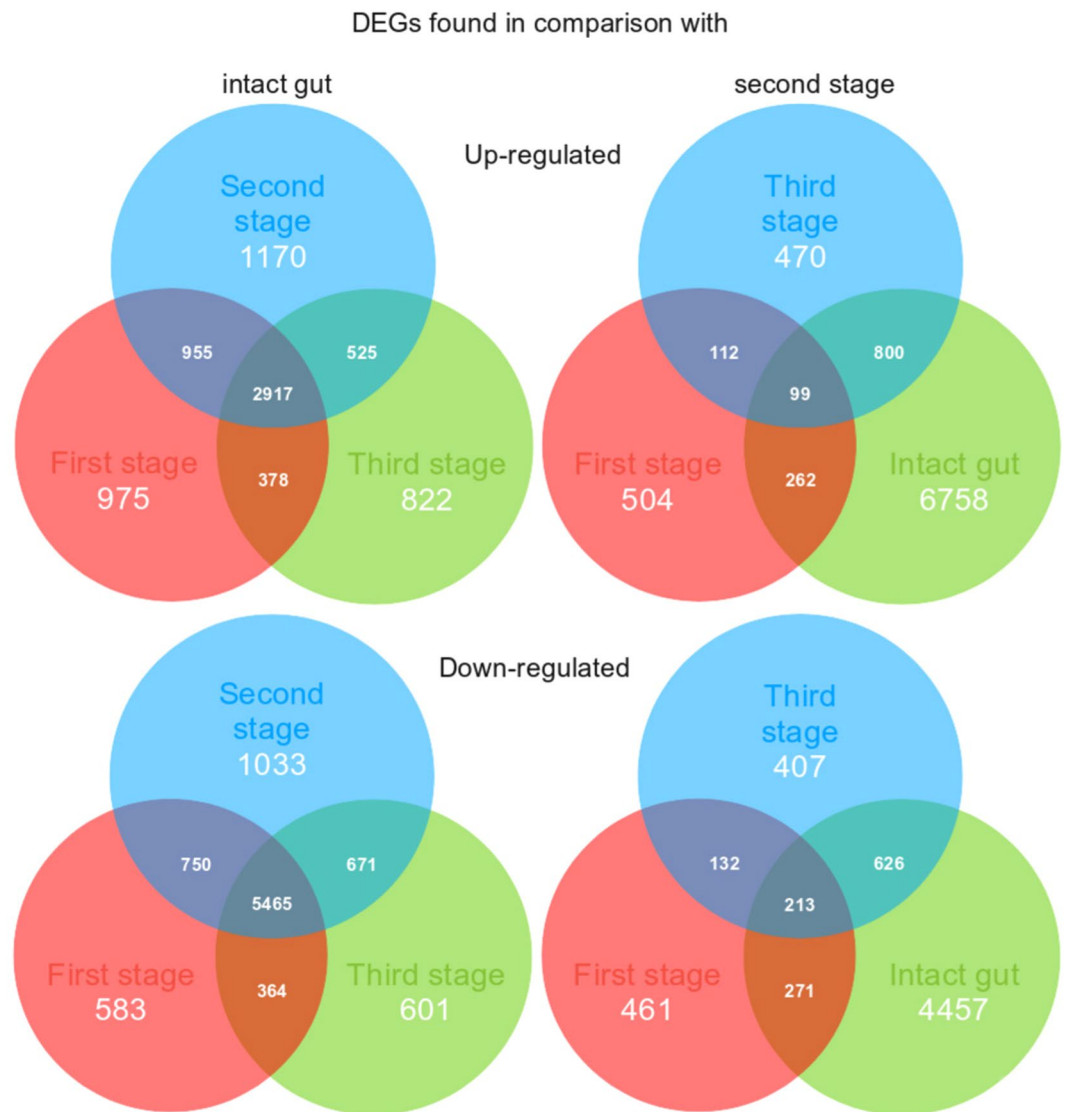


Figure 3. Venn diagrams for up-regulated and down-regulated DEGs.

Annotation. The annotation of 83,960 sequences by the BLASTx search against to the NRP NCBI database, resulted in the identification of 48,677 sequences and 37,790 genes with significant hits, including 7,118 sequences having only unnamed hits (see Supplementary Table S3). Hits belonged to 1,022 organisms, and almost 80% of sequence matched echinoderm proteins (see Supplementary Table S4). In addition, there is a high probability for the presence of a few contaminant sequences. However, the lack of a genome for this or any related holothurian species does not allow us to accurately determine the source of these sequences.

The annotation against human and sea urchin proteins led to the identification of 10,358 and 14,617 orthologs, respectively, as we used a new approach to find the best orthologs. This approach allows for the extraction of a larger number of protein–sequence pairs, versus that of the classical reciprocal search²⁹. Thus, the reciprocal method provides only 7,791 and 9,292 pairs compared to human and sea urchin proteins, respectively. The difference between our own versus reciprocal methods is that only the best hits are analysed in the latter. That is, if there is a protein family in an evolutionarily distant species, with a member of one of the species having better alignment with all members of the protein family of an other species, the reciprocal approach enables a match for this member only. In our case, protein pairs were distributed automatically, based on alignment quality, so members of the best pairs were immediately removed from the hit list of all other pairs (algorithm and code described in Supplementary Note).

Search of the most likely candidates for the role of transdifferentiation regulators. A search for homologs of sea urchin and human TFs resulted in the identification of 308 and 918 TFs, respectively (see Supplementary Table S5). Of these, 265 TF homologs were found among proteins both humans and sea urchin. Given all TFs, 20 were the most interesting in terms of expression in the second stage of regeneration. All homologs were verified using the NRP NCBI database, but for zinc finger TFs were confirmed only matching with zinc finger proteins. Then, 8 putative zinc finger proteins were removed, as none had an unconditional match

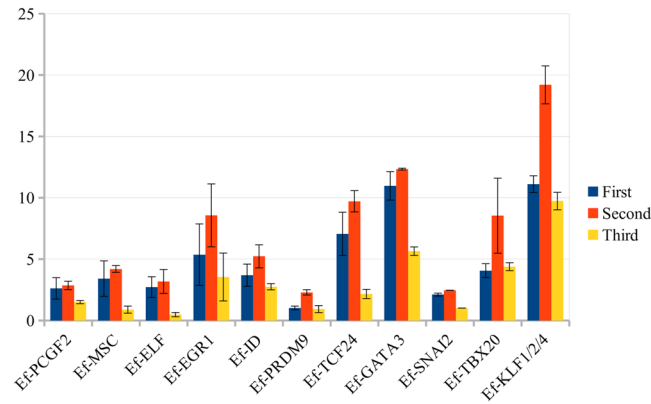


Figure 4. TPM expression values of the most likely candidates for the role of transdifferentiation regulators. All these TFs have a significant peak of expression at second stage of regeneration.

to known zinc finger proteins. Also, Ef-ELF4 was removed, since the region corresponding to the ETS domain of Ef-ELF2 and Ef-ELF4 is identical, while the rest of the Ef-ELF4 sequence, unlike the Ef-ELF2, has no significant alignment to the NRP NCBI database proteins. The Ef-ELF4 sequence is apparently an assembly error, which is also confirmed by sea urchin data, for which only one *elf* gene was found to be homologous to the human *elf1*, *elf2*, and *elf4* genes³⁰. For this reason, by analogy with the sea urchin *elf* gene, we refer below to Ef-ELF2 as Ef-ELF. The expression profiles of the 11 remaining TFs, with TPM values at each stage, are shown in Fig. 4. With the exception of Ef-ID2, they all had a p-value less than 0.05. They belong to 6 TF classes: tryptophan cluster (Ef-ELF), C2H2 zinc finger (Ef-PRDM9, Ef-EGR1, Ef-KLF1/2/4 (KLF2), Ef-SNAI2), bHLH (Ef-TCF24, Ef-MSK, Ef-ID2), C4 zinc finger (Ef-GATA3), polycomb group ring finger (Ef-PCGF2), and T-box (Ef-TBX20).

Ef-ELF (ELF2) belongs to the Ets family, whose members are powerful regulators of cell proliferation, angiogenesis, hematopoiesis, tumour transformation, and differentiation^{31,32}. However, the functions of ELF1, ELF2, and ELF4 in chordates are associated with lymphoid and endothelial cell differentiation, hematopoietic stem and progenitor cell survival, and tumour suppression^{33–36}. Sea urchin *elf* expression is detected everywhere in late gastrula, being more concentrated in the gut³⁰. The latter assumes that the ELF ortholog in echinoderms may be involved in the regulation of the early stages of digestive epithelium formation. Our data agree with this assumption. In *E. fraudatrix*, Ef-elf is expressed at early regeneration stages, during the formation of the gut luminal epithelium (Fig. 4). In the third stage, when the gut tube is already formed, its expression is sharply reduced.

Ef-KLF1/2/4 is probably a homolog of human KLF1, 2, and 4, which is confirmed by the data of Mashanov *et al.*²⁴. The KLF proteins play diverse roles in cell proliferation, differentiation, and development³⁷. In particular, KLF2 and KLF4 are key TFs that maintain a stem cell-like state and somatic cell reprogramming³⁸. In this regard, the expression of Ef-klf1/2/4 in *E. fraudatrix* may indicate its involvement in the regulation of differentiation and/or transdifferentiation. At all three stages of regeneration, this gene showed a high level of expression. As mentioned above, various transformation processes of the coelomic epithelium occur in this period. On the mesenteric surface and lateral sides of the gut tube, myoepithelial cells and peritoneocytes dedifferentiate, migrate, proliferate, and redifferentiate. Their dedifferentiation, embedding in connective tissue, proliferation, and transdifferentiation occur on the ventral side of the anlage (Fig. 1). A two-fold increase in the expression of Ef-klf1/2/4 at the second stage, compared to the first and third stages (Fig. 4) may indicate activation of cell reprogramming and maintenance of the dedifferentiated state during transdifferentiation, when gut epithelium is formed. According to Mashanov *et al.*²⁴, expression of *klf1/2/4* does not change during regeneration of the intestine and nervous system in the holothurian *H. glaberrima*. Differences in the dynamics of *klf1/2/4* expression between *H. glaberrima* and *E. fraudatrix* may reflect different formation mechanisms for intestinal epithelia in these species. In *H. glaberrima*, intestinal epithelium forms from dedifferentiated cells of the esophagus mucosa, without transdifferentiation¹⁶.

The expression of the *Ef-prdm9* gene is relatively low, compared to the other TFs considered here (Fig. 4). However, its activity at the second stage is more than twice as high as that during other stages. This may suggest its involvement in the regulation of some transdifferentiation aspects. Proteins of the PRDM family are important epigenetic regulators in the development, cell differentiation, and pluripotency in the mouse³⁹. In *E. fraudatrix*, Ef-PRDM9 may be involved in the processes of differentiation and/or maintenance of the dedifferentiated cell state.

EGR1 is involved in the cell cycle progression of various tumour types as well as hepatic regeneration in mammals^{40,41}. Recent research has shown that EGR1 is a “putative pioneer factor to directly activate wound-induced genes” in whole-body regeneration of acoels⁴². Ef-EGR1 probably performs similar functions. However, the dynamics of its expression during gut regeneration in *E. fraudatrix* (Fig. 4) shows that Ef-EGR1 may also play a role in the regulation of transdifferentiation.

SNAI proteins play a crucial role in the epithelial-mesenchymal transition (EMT) and in repressing the mesenchymal-epithelial transition (MET), being important in embryonic and tumour development^{43–45}. The co-expression of *Ef-snai2* and *Ef-id2* (see below) may indicate partial involvement of the EMT mechanisms in the transformation of coelomic epithelial cells into enterocytes in *E. fraudatrix*. Although transforming cells do not lose contact with neighboring cells⁷, they can acquire some mesenchymal features when migrate to the intestinal

anlage, similar to those during the growth of tubular organs⁴⁶. Epithelial cells at the migration tip have a partially mesenchymal character. These cells maintain intercellular junctions, but are flattened and form lamellipodia^{47,48}.

To our knowledge, no animal studies on the function of TCF24 in regeneration, development, or carcinogenesis have been carried out. The dynamics of *Ef-tcf24* expression in *E. fraudatrix* suggest its involvement in gut regeneration. The expression of *Ef-tcf24* at the second stage is 30 times as high as that of intact individuals (Fig. 4). Moreover, the number of *Ef-tcf24* transcripts reduces abruptly at the third stage, when the embedding of coelomic epithelium and its transformation into gut epithelium is completed. Based on these facts, we can assume this TF to be involved in the regulation of transdifferentiation or dedifferentiation of the coelomic epithelium.

Musculin (MSC) is involved in myogenesis and stem cell functioning in mammals. It represses myogenesis and acts as a lineage-restricted repressor of embryonic skeletal muscle development⁴⁹. MSC regulates LIF-induced gene expression of some renoprotective factors in the adult kidney-side population cells that are in a pluripotent state⁴⁹. In *E. fraudatrix*, *Ef-MSC* can perform similar functions. Its relatively low expression (Fig. 4) can be explained by the small number of cells in which *Ef-msc* is activated. It probably represses the myogenic differentiation of coelomic epithelium cells and maintain their pluripotent state in the formation of digestive epithelium.

Another TF, ID2, is not a transcription factor, despite the presence of the HLH-domain, as it lacks the basic DNA-binding domain. Yet it forms a heterodimer with bHLH TFs, inhibiting their function: in a sea urchin, this is noted as a TF²⁷, and its functions and expression seemed noteworthy to us for consideration. ID2 prevents precocious differentiation of epithelial cell progenitors during mouse gut development⁵⁰. Downregulation of ID2 induces dedifferentiation of epithelial cells⁵¹. *Ef-ID2* could perform similar functions during regeneration in *E. fraudatrix*, such as involvement in the dedifferentiation of coelomic epithelial cells and maintenance of their dedifferentiated state. This protein, with *Ef-SNAI2*, probably regulates cell migration and EMT, similar to their homologues in mammals^{52,53}.

The expression of *Ef-pcgf2* during gut regeneration in *E. fraudatrix* is relatively low (Fig. 4), and the difference in its magnitude between stages is negligible. Based on this, we can assume this TF would be involved in the processes that occur throughout the three regeneration stages. It has been shown that PCGF2 is a chromatin-modifying protein that inhibits EMT by downregulation of ZEB1 and ZEB2, in human breast cancer, and negatively regulates stem cell-like properties in human gastric cancer^{54,55}. In *E. fraudatrix*, its homolog may be involved in the maintenance and retention of the “mesodermal” properties of coelomic epithelial cells on the gut anlage surface.

Expression of *Ef-gata3* at the first and second stages is almost twice as high as at the third stage, which suggests that this TF is involved in the regulation of gut lining formation. Like its homologs in other animals^{56–58}, *Ef-GATA3* is apparently involved in the transformation processes of the mesoderm (coelomic epithelium cells).

TBX20 is a crucial cardiogenic TF in mammals⁵⁹. The *tbx20* gene is also expressed in the nervous system and embryonic lateral mesoderm of the mouse^{59,60}. It is probable that *Ef-TBX20* in *E. fraudatrix* is also involved in the myogenic cell differentiation of the coelomic epithelium and the formation of gut muscles. However, the expression of *Ef-tbx20* at the second stage of regeneration is almost twice as high as that of the first and third stages. In this regard, its participation in the regulation of transdifferentiation can also be assumed.

Obviously, other TFs, which do not show such unambiguous dynamics as those listed above, can participate in the regulation of transdifferentiation. In fact, gene activation can be short-term or done by a small number of cells. In the former case, given sufficiently long intervals between stages, we could not record any sharp variation in expression. In the latter case, due to the use of the whole anlage, rather than single cells or equal cell numbers, the average number of transcripts of such genes may be small, so there may be no difference between certain stages of regeneration. Among these genes, a homolog of human *sox17* deserves special attention. In vertebrates, it is a marker of the formation of the endoderm and gut^{61,62}. According to Campbell *et al.*²³, *SOX17* is involved in the transdifferentiation of muscle cells into enterocytes in zebrafish. In *E. fraudatrix*, the *sox17* homolog is expressed as early as the first stage of regeneration, and subsequently the number of its transcripts increases (see Supplementary Table S5). This could indicate that the initial stages of determination of enterocyte precursors are launched on the surface of the anlage, before the cells migrate to the connective tissue.

Network of enrichment biological processes and pathways. By constructing a network associated with 11 TFs listed above, we obtained 490 enrichment pathways and biological processes, combining 3,168 homologs of human proteins (Fig. 5). Of them, 204 and 71 Gene Ontology (GO) terms had down- and upregulated genes, respectively, at the second stage relative to both the first and third stages. The obtained network primarily attracts attention by several blocks of closely interrelated processes associated with morphogenesis, confirming the involvement of the TFs we revealed during gut regeneration in *E. fraudatrix*.

The central part is occupied by terms, regarding the regulation of morphogenesis, intercellular interactions, and interaction of cells with the extracellular matrix (block 1). Of interest is the increase in expression of the genes involved in cell fate determination and commitment, over three stages. Indeed, in *E. fraudatrix*, the intestinal and coelomic epithelia form and their cells specialize in this period (days 3–11 post-evisceration)^{7,12}. Thus, in this block, genes that trigger or control the fate commitment change of coelomic epithelial cells.

The second group includes terms related to hematopoiesis and immune response. The expression of a large number of immune-related genes may indicate postinjury activation of protective functions in *E. fraudatrix*. This is not surprising, as the animals must protect the internal environment from pathogens that can penetrate from seawater during evisceration.

Terms related to neurogenesis in the third group is also logical, as the coelomic epithelium of echinoderms contains a large number of nerve processes and neurons that form the basiepithelial nerve plexus. During the digestive system's regeneration of *E. fraudatrix*, this plexus forms again in the coelomic epithelium of the intestine.

The fourth group contains processes related to cardiac development, mesenchymal morphogenesis, and EMT. Although echinoderms lack a heart, this group is noteworthy, as most processes in it are characterised by

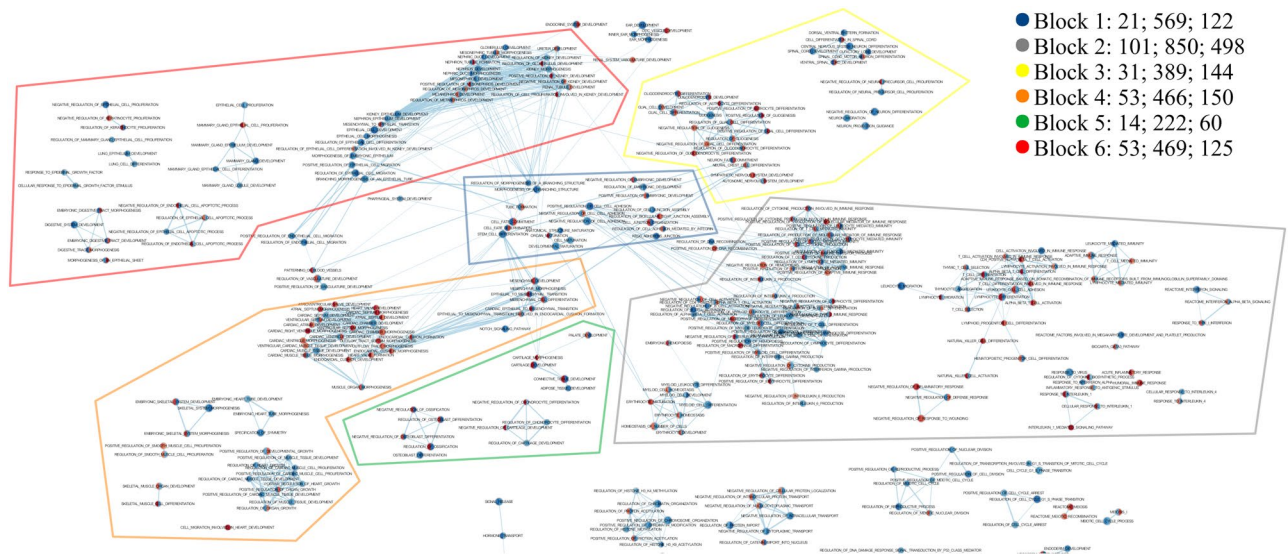


Figure 5. Network of enrichment biological processes and pathways associated with the most likely candidates for the role of transdifferentiation regulators. Nodes represent biological process (gene set). Edges represent overlap between pair of gene sets. Node size and edge width depend on the number of genes. Node fill represents enrichment scores of terms at the second stage of regeneration relative to the first (right half) and third stages (left half), respectively. The color gradient represents an increase (red color) or decrease (blue color) in the level of expression (depend on the enrichment score) in the second stage, relative to the first or third. A block description is a block number, the number of terms, and the number of all and unique genes in a block. The full version of the network in Cytoscape format is represented in the Supplementary Data S1.

increased gene expression at the second stage of regeneration. An analysis of the GO terms available shows they are associated with processes of mesodermal cells reorganization, such as migration, proliferation, and EMT, occurring in heart development of mammals⁶³. In *E. fraudatrix*, similar processes occur in coelomic epithelium, which is transformed during intestinal regeneration. Most genes associated with EMT have a decreasing expression of dynamics, with a maximum at the first stage of regeneration. Activation of EMT mechanisms during gut regeneration in *E. fraudatrix* is unusual and to some extent contradicts morphological data; the intercellular junctions are not destroyed during formation of intestinal epithelium, and epithelial cells are not transformed into mesenchymal cells. This is explained by the conversion process of part of the coelomic epithelium into mesenchyme, which is not associated with the enterocyte formation, as described by the holothurian *H. glaberrima*⁶⁴. In addition, when coelomic epithelium is immersed in the connective-tissue anlage of the intestine in *E. fraudatrix*, cells are likely to acquire some mesenchymal features, as mentioned^{46–48}. It is likely that partial and full EMT are regulated by the same genes.

The fifth block contains the GO terms associated with extracellular matrix remodelling, and the differentiation of chondrocytes and osteoblasts. Gene expression data are consistent with morphological events occurring during the intestinal regeneration of *E. fraudatrix*. The formation of the connective tissue anlage of the gut occurs in the early stages post-evisceration, manifested as higher gene activity in the first stage. The terms associated with the differentiation of chondrocytes and osteoblasts possibly reflect the processes of fibroblast changes, as well as the active formation of extracellular matrix of the gut wall.

The sixth block is quite large and contains GO terms associated with kidney development and epithelium morphogenesis. In the regeneration in *E. fraudatrix*, it should be considered as a group of genes involved in epithelial transformation, since holothurians lack kidneys, or any specialized excretory organs²². In *E. fraudatrix*, homologs of these genes are apparently involved in the reorganization of migrating clusters of epithelial cells into the tubular luminal intestinal epithelium – as shown by the presence of GO terms, such as “nephron tubule formation,” “mesonephric tubule morphogenesis,” etc. In the context of this article, genes of the sixth block are less interesting, being associated with processes of morphological reorganization of epithelium, not with transdifferentiation.

In any case, due to the difficulty of interpreting the processes identified for humans, the main function of this analysis was to obtain evidence of the involvement of the TF set in proliferation, cell differentiation, and embryonic development. This consideration has been formed as a function of two factors. First, it is impossible to find an ortholog for every human gene in the *E. fraudatrix* genome. The number of genes and more important, the composition of gene families is too different between such evolutionarily distant organisms. At best, we may consider one gene homologous to several human genes. In this regard, the exact function of genes cannot be predicted by human homologs. Second, human biological processes cannot be directly applied to evolutionary and structurally distant organisms. Thus, we consider it critically important to avoid proposing definite hypotheses about the functions of proteins and processes, based only on the homology and functional annotation of evolutionarily distant species. In other words, any analysis based on homology must be interpreted carefully.

Conclusions

Our study shows that the regeneration of the digestive system in *E. fraudatrix* causes a significant change in the expression of many genes, including TFs. Among the latter, 11 can be distinguished, the expression of which increases on days 5–7 post-evisceration. In this period, coelomic epithelium is immersed in the connective tissue of the intestinal anlage, with the coelomic epithelial cells transforming into enterocytes. We suggest that the 11 above-noted TFs as the most likely candidates for the role of transdifferentiation regulators in holothurians. To continue with the study on the mechanisms of transdifferentiation, it will be necessary to identify cells in which these TFs are expressed, and how gene cascades are triggered by each TF.

Methods

Animals. Adult mature individuals of the holothurian *Eupentacta fraudatrix* (D'yakonov *et al.*, 1958) were collected in the Peter the Great Bay, Sea of Japan, and kept in 3 m³ tanks during one week with running aerated seawater at 16 °C. Evisceration was induced by injection of 2% KCl. Then holothurians were transferred to 370-liter tanks with aerated seawater at 20–22 °C, where they regenerated their internal organs.

Sample collection and RNA extraction. Tissue of gut was sampled from the intact gut and on the third (first stage of regeneration), fifth–seventh (second stage), and tenth (third stage) days post-evisceration. A total of 10 individuals were selected from each of the stages. Each sample was represented in two biological replicates with 5 individuals, pooled together, per replicate. Fixation was performed in 3 mL RNAlater (Sigma, USA) for 24 h at 4 °C; then the material in RNAlater was stored at –20 °C. Before isolating total RNA, the tissue was precipitated in sterile seawater. Homogenization was carried out in ExtractRNA (Evrogen, Russia) with metal balls on a TissueLyser LT homogenizer (Quagen, Germany). Total RNA was isolated using phenol-chloroform extraction⁶⁵.

Transcriptome sequencing. All steps for preparing the libraries and sequencing were carried out at the Evrogen company (Evrogen, Russia). The libraries were prepared using a TruSeq Stranded mRNA Library Prep Kit (Illumina, USA), and fragments with a length of 250–450 nucleotides, including adapters, were selected. After testing the quality on an Agilent TapeStation system, paired-end sequencing (2 × 100) was performed on an Illumina HiSeq. 2500 sequencing system. Raw reads were loaded to the SRA NCBI database with the accession numbers from SRR8297983 to SRR8297990 for the three stages of regeneration and intact gut, respectively.

De novo transcriptome assembly. Raw reads from eight libraries (listed in Supplementary Table S1), obtained by us in FASTQ format, were processed using Trimmomatic 0.36 software⁶⁶ with “LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 AVGQUAL:25 MINLEN:21” parameters to achieve clean reads by removing those containing adapter sequences, poly-N sequences, or low-quality bases.

Then the clean reads were corrected and assembled using SPAdes 3.13 software⁶⁷ with 2 iterations for read correction step and with a k-mer length of 25, 33, and 49. Of all the obtained contigs, the Coding Sequences (CDSs) with a minimum length of 30 amino acid residues were extracted using TransDecoder 5.5.0 software⁶⁸. The code of TransDecoder was modified in such a way that stop-codon or the beginning of the sequence, but not “ATG” (Met) or the beginning of the sequence, was taken for the beginning of CDS. All CDSs were verified using BLAST-search⁶⁹ by the database containing protein sequences from the SwissProt (11.12.18) and the Echinobase (11.12.18) databases^{27,70}, as is described in the manual to TransDecoder.

Subsequently, the obtained sequences were assembled with HomoloCAP script as described in the article²⁶. The resulting sequences were filtered according to the NCBI requirements and uploaded to the TSA NCBI Database with the index GHCL00000000. Transcriptome completeness was assessed using BUSCO v3²⁸ in mode “protein” with a Metazoa v9 dataset.

Differential expression analysis. To find DEGs, the number of mapped reads was calculated in RSEM v1.3.1 software⁷¹; paired-end reads were aligned in Bowtie 2.3.4 software⁷² with the following parameters: “–no-mixed–no-discordant–gbar 50–end-to-end –k 200–very-sensitive–minins 50 –q–maxins 450–fr”. The procedure of detection of significant changes in expression was modified. Thus, for analysis we used sequences satisfying the following conditions: the total number of mapped reads was more than 50; more than tenfold coverage; a zero number of mapped reads allowed at any stage. After this filtration, differential expression was evaluated for the sequences in DESeq. 2 v1.18 software⁷³ (see Supplementary Data S2). The analysis was performed in two variants: in the first case, the control sample was the first stage of regeneration; in the second case, intact gut. Those DEGs, whose expression level was two times as high at sample than at the control, and with the p-value less than 0.05, were considered actual.

Validation of gene expression. In order to validate the changes in gene expression determined by RNA-seq, we selected 5 genes of TFs for qPCR analysis. Poly(A) RNA was extracted as described above, but from other individuals than used for sequencing. RNA was isolated independently three times (5 individuals per repeat) for each of the three stages of regeneration. Sequences of PCR primers are shown in the Supplementary Table S6. RNA was reverse transcribed with poly-dT primers as described in the protocol of MMLV RT kit (Evrogen, Russia). PCR reaction was performed using a qPCRmix-HS SYBR kit (Evrogen, Russia) according to the manufacturer's protocol for CFX96 Touch system (Bio-Rad, USA). The reactions were performed on three independent RNA samples per condition. Each sample was analyzed at least twice, making sure that the difference between technical replicates was lower than 0.5 Ct. The first regeneration stage was used as a control sample. All the expression values were normalized relative to the Ct geometric mean of Tubulin and EF1a genes. The delta-delta Ct method was used to evaluate expression. Primer efficacy was evaluated using ten serial dilutions (1:2) of cDNA combined from all samples in CFX Manager v3.1 software (Bio-Rad, USA). Specificity of the primers was determined by Sanger sequencing of target amplicon.

Annotation. Annotation was carried out by several protein databases with the standard e-value for BLASTP 2.7.0, which was $1e^{-5}$. Basic annotation was by the NR NCBI Database (11.12.18); the annotation for enrichment analysis was performed by BLASTP searching against human proteins from the Ensemble database v95⁷⁴; the annotation for finding the transcription factors was based on the human proteins from Ensemble and sea urchin proteins from the Echinobase project²⁷. Only TF homologs with a predicted protein length of more than 200 amino acid residues were taken into consideration. Orthologs of human proteins were identified using a custom Python script that implements modified reciprocal method for finding the best hit (algorithm and code described in Supplementary Note).

To identify the most likely candidates for the role of transdifferentiation regulators, only the TFs satisfying the three conditions were taken into consideration: the values of TPMs (Transcripts Per Kilobase Million) of the second stage higher than unity; the average value of LogFC (logarithm of fold change with base 2) of the second stage relative to the first and third stages more than half; the value of LogFC (logarithm of fold change) of the second stage relative to the first or third stages more than unity.

The enrichment analysis of biological processes and pathways was performed with GSEA software⁷⁵ in accordance with the EnrichmentMap protocol for RNA-seq data⁷⁶. The gene sets for biological processes and pathways were accessed from the MsigDB database v6.2⁷⁵. Only processes or pathways containing at least one of the 11 TFs selected as described in the results section were used. Then, results of the enrichment analysis were visualized using EnrichmentMap plug-ins in Cytoscape software⁷⁷ (see Supplementary Data S1).

Data availability

The raw reads, obtained by us, were uploaded to the SRA NCBI Database with the accession numbers from SRR8297983 to SRR8297990 for the three stages of regeneration and the normal gut, respectively. The assembly was uploaded to the TSA NCBI Database with the index GHCL00000000.

Received: 6 November 2019; Accepted: 15 January 2020;

Published online: 30 January 2020

References

- Lai, A. G. & Aboobaker, A. A. EvoRegen in animals: Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes. *Dev. Biol.* **433**, 118–131 (2018).
- Borisenko, I. E., Adamska, M., Tokina, D. B. & Ereskovsky, A. V. Transdifferentiation is a driving force of regeneration in *Halisarca dujardini* (Demospongiae, Porifera). *PeerJ* **3** (2015).
- Lavrov, A. I., Bolshakov, F. V., Tokina, D. B. & Ereskovsky, A. V. Sewing up the wounds: The epithelial morphogenesis as a central mechanism of calcareous sponge regeneration. *J. Exp. Zool. Part B Mol. Dev. Evol.* **330**, 351–371 (2018).
- Ahmed, E. *et al.* Lung development, regeneration and plasticity: From disease pathophysiology to drug design using induced pluripotent stem cells. *Pharmacol. Ther.* **183**, 58–77 (2018).
- Dolmatov, I. Y. Regeneration in echinoderms. *Russ. J. Mar. Biol.* **25**, 225–233 (1999).
- Dolmatov, I. Y. & Ginanova, T. T. Muscle regeneration in holothurians. *Microsc. Res. Tech.* **55**, 452–463 (2001).
- Mashanov, V. S., Dolmatov, I. Y. & Heinzeller, T. Transdifferentiation in holothurian gut regeneration. *Biol. Bull.* **209**, 184–193 (2005).
- García-Arraras, J. E. & Dolmatov, I. Y. Echinoderms: potential model systems for studies on muscle regeneration. *Curr. Pharm. Des.* **16**, 942–955 (2010).
- Kalacheva, N. V., Eliseikina, M. G., Frolova, L. T. & Dolmatov, I. Y. Regeneration of the digestive system in the crinoid *Himerometra robustipinna* occurs by transdifferentiation of neurosecretory-like cells. *PLoS One* **12**, 1–28 (2017).
- Vogt, G. Hidden Treasures in stem cells of indeterminate growing bilaterian invertebrates. *Stem Cell Rev. Reports* **8**, 305–317 (2012).
- Mashanov, V. S., Zueva, O. R. & García-Arraras, J. E. Inhibition of cell proliferation does not slow down echinoderm neural regeneration. *Front. Zool.* **14**, 1–9 (2017).
- Dolmatov, I. Y. & Mashanov, V. S. *Regeneration in holothurians*. (Dalnauka, 2007).
- Mashanov, V. S. & García-Arraras, J. E. Gut regeneration in holothurians: A snapshot of recent developments. *Biol. Bull.* **221**, 93–109 (2011).
- Hyman, L. H. The invertebrates: Echinodermata, the coelomate bilateria. in (1955).
- Shukalyuk, A. I. & Dolmatov, I. Y. Regeneration of the digestive tube in the holothurian *Apostichopus japonicus* after evisceration. *Russ. J. Mar. Biol.* **27**, 168–173 (2001).
- García-Arraras, J. E. *et al.* Cellular mechanisms of intestine regeneration in the sea cucumber, *Holothuria glaberrima* Selenka (Holothuroidea:Echinodermata). *J. Exp. Zool.* **281**, 288–304 (1998).
- Dolmatov, I. Y., Khang, N. A. & Kamenev, Y. O. Asexual reproduction, evisceration, and regeneration in holothurians (Holothuroidea) from Nha Trang Bay of the South China Sea. *Russ. J. Mar. Biol.* **38**, 243–252 (2012).
- Dolmatov, I. Y. New data on asexual reproduction, autotomy, and regeneration in holothurians of the order Dendrochirota. *Russ. J. Mar. Biol.* **40**, 228–232 (2014).
- Leibson, N. L. & Dolmatov, I. Y. Evisceration and regeneration of the inner complex in a sea cucumber *Eupentacta fraudatrix* (Holothuroidea, Dendrochirota). *Zool. Zhurnal* **68**, 67–74 (1989).
- Leibson, N. Regeneration of digestive tube in holothurians *Stichopus japonicus* and *Eupentacta fraudatrix*. in *Keys for Regeneration* (eds. Taban, C. H. & Boilly, B.) 51–61 (Karger, 1992).
- Dolmatov, I. Y. Regeneration of the aquapharyngeal complex in the holothurian *Eupentacta fraudatrix* (Holothuroidea, Dendrochirota). in *Keys for Regeneration* (eds. Taban, C. H. & Boilly, B.) 40–50 (Karger, 1992).
- Harrison, F. W. & Ruppert, E. E. *Microscopic anatomy of invertebrates. Volume 14: Echinodermata*. (Wiley-Liss, 1994).
- Campbell, C. *et al.* *In vivo* lineage conversion of vertebrate muscle into early endoderm-like cells. *bioRxiv Dev. Biol.*, <https://doi.org/10.1101/722967> (2019).
- Mashanov, V. S., Zueva, O. R. & García-Arraras, J. E. Expression of pluripotency factors in echinoderm regeneration. *Cell Tissue Res.* **359**, 521–536 (2015).
- Dolmatov, I. Y., Afanasyev, S. V. & Boyko, A. V. Molecular mechanisms of fission in echinoderms: transcriptome analysis. *PLoS One*, <https://doi.org/10.1371/journal.pone.0195836> (2018).
- Boyko, A. V., Girich, A. S., Eliseikina, M. G., Maslennikov, S. I. & Dolmatov, I. Y. Reference assembly and gene expression analysis of *Apostichopus japonicus* larval development. *Sci. Rep.* **9**, 1–11 (2019).
- Kudrarkar, P. & Cameron, A. Echinobase: an expanding resource for echinoderm genomic information. *Database* **2017**, 1–9 (2017).

28. 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).
29. Ward, N. & Moreno-Hagelsieb, G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS One* **9**, 1–6 (2014).
30. Rizzo, F., Squarzoni, P., Archimandritis, A. & Arnone, M. I. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). *Dev. Biol.* **300**, 35–48 (2006).
31. Hsu, T., Trojanowska, M. & Watson, D. K. Ets proteins in biological control and cancer. *J. Cell. Biochem.* **91**, 896–903 (2004).
32. Oikawa, T. & Yamada, T. Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11–34 (2003).
33. Sashida, G., Bazzoli, E., Menendez, S., Liu, Y. & Nimer, S. D. The oncogenic role of the ETS transcription factors MEF and ERG. *Cell Cycle* **9**, 3457–3459 (2010).
34. Guan, F. H. X. *et al.* The antiproliferative ELF2 isoform, ELF2B, induces apoptosis *in vitro* and perturbs early lymphocytic development *in vivo*. *J. Hematol. Oncol.* **10**, 1–17 (2017).
35. Suico, M. A., Shuto, T. & Kai, H. Roles and regulations of the ETS transcription factor ELF4/MEF. *J. Mol. Cell Biol.* **9**, 168–177 (2017).
36. Xie, Y. *et al.* Reduced Erg dosage impairs survival of hematopoietic stem and progenitor cells. *Stem Cells* **35**, 1773–1785 (2017).
37. Ilsley, M. D. *et al.* Krüppel-like factors compete for promoters and enhancers to fine-tune transcription. *Nucleic Acids Res.* **45**, 6572–6588 (2017).
38. Yamane, M., Ohtsuka, S., Matsuura, K., Nakamura, A. & Niwa, H. Overlapping functions of Krüppel-like factor family members: targeting multiple transcription factors to maintain the naïve pluripotency of mouse embryonic stem cells. *Development* **145**, dev162404 (2018).
39. Vervoort, M., Meulemeester, D., Béhague, J. & Kerner, P. Evolution of Prdm genes in animals: Insights from comparative genomics. *Mol. Biol. Evol.* **33**, 679–696 (2016).
40. Lai, S. *et al.* Regulation of mice liver regeneration by early growth response-1 through the GGPPS/RAS/MAPK pathway. *Int. J. Biochem. Cell Biol.* **64**, 147–154 (2015).
41. Yan, L. *et al.* MiR-301b promotes the proliferation, mobility, and epithelial-to-mesenchymal transition of bladder cancer cells by targeting EGR1. *Biochem. Cell Biol.* **95**, 571–577 (2017).
42. Gehrke, A. R. *et al.* Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* (80-). **363**, eaau6173 (2019).
43. Kalinkova, L. *et al.* Decreased methylation in the SNAI2 and ADAM23 genes associated with de-differentiation and haematogenous dissemination in breast cancers. *BMC Cancer* **18**, 1–12 (2018).
44. Zhou, Y. *et al.* Transdifferentiation of type II alveolar epithelial cells induces reactivation of dormant tumor cells by enhancing TGF- β 1/SNAI2 signaling. *Oncol. Rep.* **39**, 1874–1882 (2018).
45. Chen, Y., Wang, K., Qian, C. N. & Leach, R. DNA methylation is associated with transcription of Snail and Slug genes. *Biochem. Biophys. Res. Commun.* **430**, 1083–1090 (2013).
46. Schöck, F. & Perrimon, N. Cellular processes associated with germ band retraction in *Drosophila*. *Dev. Biol.* **248**, 29–39 (2002).
47. Radice, G. P. The spreading of epithelial cells during wound closure in *Xenopus* larvae. *Dev. Biol.* **76**, 26–46 (1980).
48. Schöck, F. & Perrimon, N. Molecular mechanisms of epithelial morphogenesis. *Annu. Rev. Cell Dev. Biol.* **18**, 463–493 (2002).
49. Hishikawa, K. *et al.* Muscudin/MyoR is expressed in kidney side population cells and can regulate their function. *J. Cell Biol.* **169**, 921–928 (2005).
50. Nigmatullina, L. *et al.* Id2 controls specification of Lgr5 + intestinal stem cell progenitors during gut development. *EMBO J.* **36**, 869–885 (2017).
51. Veerasamy, M., Phanish, M. & Dockrell, M. E. C. Smad Mediated Regulation of Inhibitor of DNA Binding 2 and Its Role in Phenotypic Maintenance of Human Renal Proximal Tubule Epithelial Cells. *PLoS One* **8**, 1–8 (2013).
52. Kamata, Y. *et al.* Introduction of ID2 enhances invasiveness in ID2-null oral squamous cell carcinoma cells via the SNAIL axis. *Cancer Genomics and Proteomics* **13**, 493–498 (2016).
53. Chang, C., Yang, X., Pursell, B. & Mercurio, A. M. Id2 complexes with the SNAG domain of Snail inhibiting Snail-mediated repression of Integrin 4. *Mol. Cell Biol.* **33**, 3795–3804 (2013).
54. Lee, J. Y. *et al.* Loss of the polycomb protein Mel-18 enhances the epithelial-mesenchymal transition by ZEB1 and ZEB2 expression through the downregulation of miR-205 in breast cancer. *Oncogene* **33**, 1325–1335 (2014).
55. Wang, X.-F. *et al.* Mel-18 negatively regulates stem cell-like properties through downregulation of miR-21 in gastric cancer. *Oncotarget* **7** (2016).
56. Lentjes, M. H. *et al.* The emerging role of GATA transcription factors in development and disease. *Expert Rev. Mol. Med.* **18**, 1–15 (2016).
57. Materna, S. C., Ransick, A., Li, E. & Davidson, E. H. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. *Dev. Biol.* **375**, 92–104 (2013).
58. Riddle, M. R., Weintraub, A., Nguyen, K. C. Q., Hall, D. H. & Rothman, J. H. Transdifferentiation and remodeling of post-embryonic *C. elegans* cells by a single transcription factor. *Development* **140**, 4844–9 (2013).
59. De Benedittis, P. & Jiao, K. Alternative splicing of T-box transcription factor genes. *Biochem. Biophys. Res. Commun.* **412**, 513–7 (2011).
60. Takashima, Y. & Suzuki, A. Regulation of organogenesis and stem cell properties by T-box transcription factors. *Cell. Mol. Life Sci.* **70**, 3929–3945 (2013).
61. Alexander, J. & Stainier, D. Y. R. A molecular pathway leading to endoderm formation in zebrafish. *Curr. Biol.* **9**, 1147–1157 (1999).
62. Warga, R. M. & Nüsslein-Volhard, C. Origin and development of the zebrafish endoderm. *Development* **126**, 827–838 (1999).
63. Moorman, A. F. M. & Christoffels, V. M. Cardiac chamber formation: development, genes, and evolution. *Physiol. Rev.* **83**, 1223–1267 (2003).
64. García-Arrarás, J. E. *et al.* Cell dedifferentiation and epithelial to mesenchymal transitions during intestinal regeneration in *H. glaberrima*. *BMC Dev. Biol.* **11** (2011).
65. Wallace, D. M. Large- and small-scale phenol extractions. *Methods Enzymol.* **152**, 33–41 (1987).
66. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
67. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
68. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* **8** (2013).
69. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
70. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
71. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
73. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol.* **15**, 550 (2014).
74. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2017).

75. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102** (2005).
76. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
77. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (2003).

Acknowledgements

The results were obtained using the equipment of CKP «Primorsky aquarium», National Scientific Center of Marine Biology FEB RAS (Vladivostok, Russia). This study was supported by the Russian Foundation for Basic Research (grant №19-34-90015).

Author contributions

All authors conceived of and designed research; E.S.T. and A.S.G. collected animals and prepared RNA libraries; A.V.B. and I.Yu.D. analysed data, wrote the paper; all authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-58470-0>.

Correspondence and requests for materials should be addressed to A.V.B.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020