

NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone *Nematostella vectensis*

Jacob F. Warner, Vincent Guerlais, Aldine R. Amiel, Hereroa Johnston, Karine Nedoncelle and Eric Röttinger*

ABSTRACT

For over a century, researchers have been comparing embryogenesis and regeneration hoping that lessons learned from embryonic development will unlock hidden regenerative potential. This problem has historically been a difficult one to investigate because the best regenerative model systems are poor embryonic models and vice versa. Recently, however, there has been renewed interest in this question, as emerging models have allowed researchers to investigate these processes in the same organism. This interest has been further fueled by the advent of high-throughput transcriptomic analyses that provide virtual mountains of data. Here, we present *Nematostella vectensis* Embryogenesis and Regeneration Transcriptomics (NvERTx), a platform for comparing gene expression during embryogenesis and regeneration. NvERTx consists of close to 50 transcriptomic data sets spanning embryogenesis and regeneration in *Nematostella*. These data were used to perform a robust *de novo* transcriptome assembly, with which users can search, conduct BLAST analyses, and plot the expression of multiple genes during these two developmental processes. The site is also home to the results of gene clustering analyses, to further mine the data and identify groups of co-expressed genes. The site can be accessed at <http://nvvertx.kahikai.org>.

KEY WORDS: Embryogenesis, Regeneration, Transcriptome, Database, Cnidarian, *Nematostella vectensis*

INTRODUCTION

A long-standing question in the field of regeneration is to what extent regenerative programs recapitulate development. Comparing gene expression during these two processes provides clues as to how genes activated during embryogenesis are re-deployed during regeneration. The majority of studies performing this comparison focus on the role of individual or small groups of genes (Binari et al., 2013; Carlson et al., 2001; Gardiner et al., 1995; Imokawa and Yoshizato, 1997; Kaloulis et al., 2004; Katz et al., 2015; Millimaki et al., 2010; Özpolat et al., 2012; Reitzel et al., 2007; Torok et al., 1998; Wang and Beck, 2014). Studies comparing transcriptomes of embryogenesis and limb regeneration in axolotls and zebrafish have been successful in identifying differentially expressed genes involved in these processes (Habermann et al., 2004; Mathew et al., 2009). A comparison of whole-body regeneration to embryogenesis has yet to be performed, and could help with further improving our understanding of how genes are used during

embryogenesis and re-used during regeneration. One organism is especially amenable to this line of study: the sea anemone *Nematostella vectensis* (Fig. 1A) (Layden et al., 2016; Reitzel et al., 2007).

Nematostella has been used as a research model for embryonic development (Finnerty et al., 2004; Kusserow et al., 2005; Matus et al., 2006; Rentzsch et al., 2006; Wikramanayake et al., 2003). *Nematostella* reproduce sexually and, after fertilization, the zygote undergoes a series of cleavages to form a blastula. Gastrulation occurs at the animal pole and, shortly thereafter, the embryo enters a swimming planula stage during which the pharynx and internal structures, termed mesenteries, develop. After several days, this planula larva settles, develops tentacles, and enters a juvenile stage. *Nematostella* development research entered the age of genomics with the sequencing of its genome by Putnam and colleagues in 2007 (Putnam et al., 2007; Technau and Schwaiger, 2015). Since then, a large number of developmental genes have been identified in *Nematostella*, and many commonalities between *Nematostella* and bilaterian development have emerged (Amiel et al., 2017; Burton and Finnerty, 2009; Darling et al., 2005; Genikhovich et al., 2015; Layden and Martindale, 2014; Layden et al., 2012; Leclère et al., 2016; Matus et al., 2008; Meyer et al., 2011; Reitzel et al., 2007; Röttinger et al., 2012). With the advent of high-throughput transcriptomics, several studies have examined gene expression during embryogenesis at the whole-genome level (Helm et al., 2013; Fischer and Smith, 2013; Fischer et al., 2014; Tulin et al., 2013), firmly establishing *Nematostella* as an embryonic model.

More recently, *Nematostella* has shown to be a powerful model for regeneration. Upon bisection, *Nematostella* is capable of regenerating the missing body half after ~6 days postamputation (Bossert et al., 2013). Following subpharyngeal amputation (head removal), regeneration occurs via a highly dynamic process: first, there is an initial wound healing phase of ~6 h, then regeneration follows a stereotypic program in which the mesenteries fuse and, via subsequent cell proliferation, reform the missing pharynx and tentacles over the course of 6 days (Amiel et al., 2015). This process has been shown to be both cell proliferation dependent (Passamaneck and Martindale, 2012) and utilize dynamic tissue rearrangement, with large portions of unamputated tissue contributing to the reformed tissue (Amiel et al., 2015). The existence of adult stem cells and the role they might play in regeneration have yet to be uncovered. This process is known to use several developmental signaling pathways originally deployed during embryogenesis (DuBuc et al., 2014; Schaffer et al., 2016; Trevino et al., 2011). It remains unclear, however, if these pathways are deployed the same way, i.e. with similar or divergent regulatory logic. One way to address this question is to systematically compare gene expression profiles during embryonic development and regeneration, to identify groups of genes originally used during embryogenesis that are re-used during regeneration. To facilitate this line of study, we created *N.vectensis* Embryogenesis and

Université Côte d'Azur, CNRS, INSERM, Institute for Research on Cancer and Aging, Nice (IRCAN), 06107 Nice, France.

*Author for correspondence (eric.rottinger@unice.fr)

© J.F.W., 0000-0002-9197-1587; A.R.A., 0000-0002-3049-2822; H.J., 0000-0001-8477-596X; K.N., 0000-0003-4096-4886; E.R., 0000-0002-2938-6774

Received 23 December 2017; Accepted 25 April 2018

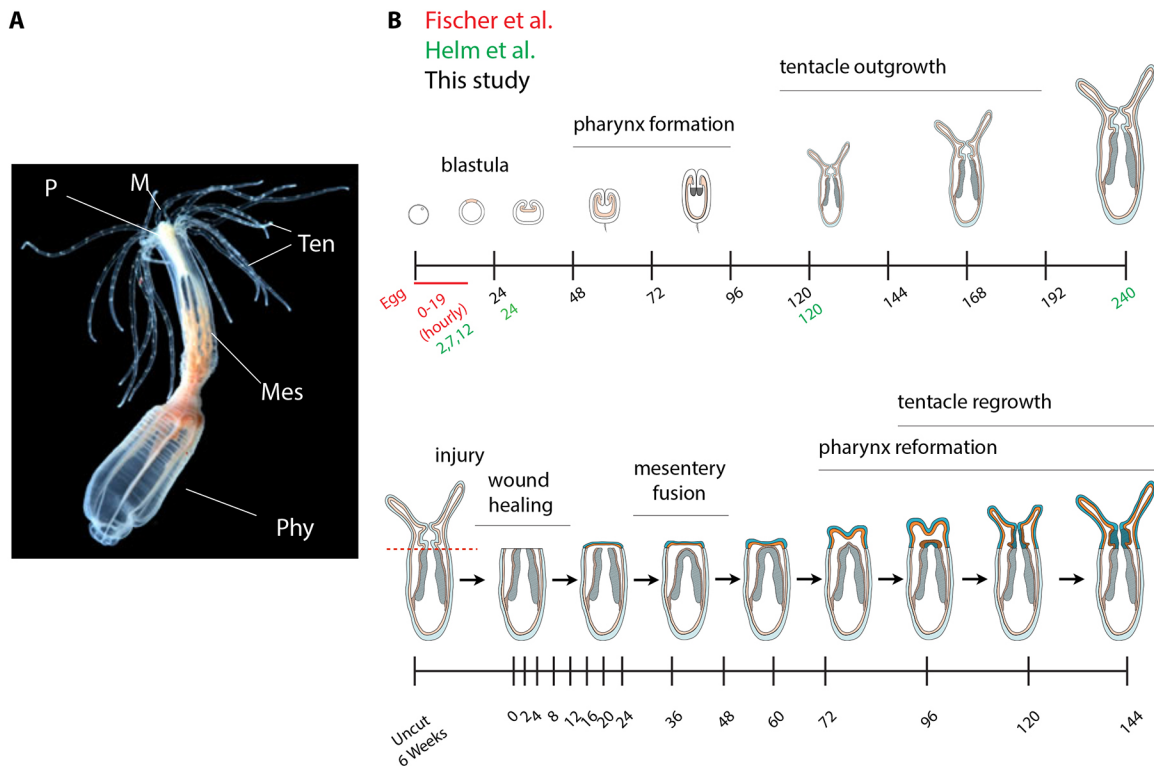


Fig. 1. Sampling strategy of data sets used in NvERTx. (A) *Nematostella* anatomy. *Nematostella* is a small sea anemone (~5 cm), with a mouth (M), pharynx (P) tentacles (Ten) and a body column with internal structures called mesenteries (Mes), the posterior section of which is termed the physis (Phy). (B) Schematic of RNAseq samples included in the NvERTx database. Three data sets spanning embryogenesis are included: one including data from Fischer and Smith (2013) and Fischer et al. (2014), sampled hourly from 0 hpf to 19 hpf; one including data from Helm et al. (2013), sampled at 2, 7, 12, 24, 120 and 240 hpf; and data from this study, sampled daily from 24 hpf to 192 hpf (250 embryos/time point, biological duplicates). Regeneration was sampled from 6-week-old animals at -1 (uncut), 0, 2, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 96, 120 and 144 hpa (300 juveniles/time point, biological triplicates).

Regeneration Transcriptomics (NvERTx) – a quantitative gene expression database for comparing embryogenesis and regeneration in the sea anemone *N. vectensis*. NvERTx comprises several data sets spanning embryogenesis (Helm et al., 2013; Fischer and Smith, 2013; Fischer et al., 2014; Tulin et al., 2013; this study) and regeneration (this study). We used pooled RNA sequencing (RNAseq) data from *Nematostella* embryogenesis and regeneration to generate a *de novo* transcriptome assembly. Using this assembly, we then quantified each of the RNAseq data sets and clustered the transcripts to discover groups of genes that share similar expression. This tool can be used to find transcript sequences, identify co-expressed genes, and directly compare expression profiles during embryogenesis and regeneration. All of these data can be found in a searchable database that is accessible at www.nvertx.kahikail.org.

RESULTS AND DISCUSSION

NvERTx database: *de novo* assembly, annotation and data quantification

NvERTx is a quantitative gene expression database for embryogenesis and regeneration, consisting of several RNAseq data sets. It includes previously published data sets spanning very early embryogenesis to polyp from Helm et al. (2013) [sampled 2, 7, 12, 24, 120 and 240 h postfertilization (hpf)] and Fischer and Smith (2013) and Fischer et al. (2014) (sampled 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19 hpf). To complement these and sample time points during tentacle genesis and pharynx formation, we generated an additional data set sampled at 24, 48, 72, 96, 120, 144, 168 and 192 hpf. Together, these data sets cover the major hallmarks of *Nematostella* development, including blastula (12–24 hpf), gastrula (24–48 hpf),

planula (48–120 hpf) and juvenile (120–244 hpf) stages. The regeneration RNAseq data were sampled from 6-week-old juveniles after subpharyngeal amputation at -1 (uncut), 0, 2, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 96, 120 and 144 h postamputation (hpa) (Fig. 1B). We chose these time points because they span the most important events of regeneration, including wound healing (0–6 hpa), pharynx reformation (24–72 hpa) and tentacle reformation (72–120 hpa) (Fig. 1A,B) (Amiel et al., 2015), all stages for which we have embryonic data spanning the initial development of these structures. For each of these data sets, we obtained the raw sequencing reads and used these as input for our quantitative workflow.

To quantify the RNAseq data, we performed a *de novo* transcriptome assembly, which we term NvERTx. We assembled this transcriptome using the short-read assembler Trinity (Haas et al., 2013), with combined paired-end reads from our regeneration data set and additional embryonic data from Tulin et al. (2013), sampled at 0, 6, 12, 18 and 24 hpf, as input (see Materials and Methods for complete workflow). The resulting assembly includes 234,381 transcripts with an N50 of 1678 and an average length of 837.30 bp (see Table S1 for assembly statistics). Each transcript is identified with a unique NvERTx.4 number. To annotate the transcriptome, we first identified the 231,294 transcripts with an open reading frame (ORF) and extracted the coding sequences using OrfPredictor (Min et al., 2005). We then compared the resultant protein sequences with NCBI's nonredundant protein database (NR) and the UniProt database using the Basic Local Alignment Search Tool (BLAST)-like tool PLAST (Nguyen and Lavenier, 2009). From these analyses we identified 85,475 transcripts with a significant hit to NR (e-value <5e-5), and 69,335 transcripts with a significant hit to the UniProt databases

(e-value <5e-5). Additionally, we compared each transcript with the current *Nemve1* gene predictions (<https://genome.jgi.doe.gov/Nemve1/Nemve1.home.html>) using nucleotide BLAST (BLASTn), and identified 102,581 transcripts with a significant hit (e-value <5e-5) (Kent, 2002). Of these 102,581 *Nemve1* hits, 19,565 represent unique *Nemve1* 'genes'. We found that 110,531 transcripts did not hit any of the three databases, a proportion typical of Trinity assemblies (Conesa et al., 2016), and could be noncoding sequences or other assembly artifacts (see Table S2 for full annotation statistics). For each of the 234,381 NvERTx transcripts, we provide all available annotations. We then used the transcriptome assembly to quantify each of the RNAseq data sets. We did this by aligning the reads using Bowtie2 (Langmead et al., 2009) and quantifying each transcript using RSEM and edgeR (Li and Dewey, 2011; Robinson et al., 2010). Transcripts with the same best *Nemve1* hit were combined to obtain 'gene-level' quantification. To validate these data, we compared our RNAseq quantifications with results obtained by performing quantitative reverse transcription PCR (RT-qPCR) for several target genes using our in-house generated data sets, and observed a high level of concordance in their expression profiles (Figs S2 and S3; see Materials and Methods, 'RNAseq quality assurance' section). Finally, the three embryonic data sets were corrected for batch effects using the *sva* R package with time point as a categorical covariate (Leek et al., 2012) (see Materials and Methods for details). The quantified data sets are reported for each transcript in the database.

Retrieving expression plots, count tables and sequences

The NvERTx database can be accessed using multiple points of entry: by searching annotations (Fig. 2Ai), using BLAST (Fig. 2Aii) or exploring the co-expression clusters (Fig. 2Aiv) (see 'Exploring gene expression clusters' section). To demonstrate this, we use the *Nematostella* Brachyury protein (NvBra, NCBI GenBank ID AAO27886.2) as an example. To retrieve transcripts corresponding to *Nvbra*, we can use the search function to query the transcript annotations and enter the *Nemve1* gene model ID from the current genome assembly (*jgi|Nemve1|770*), the NCBI GenPept accession number (*gi|122058623*), the NCBI GenBank accession number (AAO27886.2), or the gene name ('Brachyury') (Fig. 2Ai). Using any of these queries identifies several transcripts that correspond to *Nvbra* (Fig. 2B). Multiple matching transcripts reflect the different isoforms predicted by the transcriptome assembler. We can confirm which transcripts correspond to *Nvbra* by examining the annotations that are reported with the search results. It is normal for different isoforms to have slightly different annotations as each transcript was individually annotated. Clicking the NvERTx.4 numbers fills in the field on the left of the screen, enabling the user to directly compare their expression profiles by clicking 'Plot!' (Fig. 2Aiii).

Similarly, the BLAST tool can be used to retrieve NvERTx.4 transcripts using either a nucleotide (BLASTn) or protein (tBLASTn) query. The reported alignments can be then used to identify the correct transcripts. For example, using tBLASTn against the transcriptome with *Mus musculus* Brachyury (NCBI GenBank accession number AAI20808.1), we find several homologous transcripts, the first of which is *Nvbra* (Fig. 2C). Again, clicking the NvERTx.4 number fills in the field on the left of the screen, enabling the user to obtain the expression plots and annotations by clicking 'Plot!' (Fig. 2Aiii).

Once the NvERTx.4 IDs for *Nvbra*, NvERTx.4.100808 and NvERTx.4.100809, are selected we can query the database by clicking 'Plot!' on the left (Fig. 2Aiii). The first page that appears displays the transcripts' expression during regeneration and embryogenesis (Fig. 3A). We can see that the expression of

Nvbra exhibits two peaks during regeneration, beginning at 8 hpa and 60 hpa, while during embryogenesis, *Nvbra* is expressed early, rapidly peaks at 20 hpf, then decreases throughout development. Note that the expression profiles for the two transcripts are perfectly superimposed and appear as one. This is because transcripts corresponding to the same *Nemve1* best hit are quantified equivalently as they are from the same 'gene'. To distinguish transcript isoforms from separate genes we can compare the individual transcripts in the 'alignment' tab, where a MUSCLE sequence alignment is reported (Edgar, 2004) (Fig. 3F). In this case, we observe that NvERTx.4.100809 is a longer assembled isoform of *Nvbra*. The results tabset also includes the normalized count tables (Fig. 3B), transcript annotations (Fig. 3C), the sequences in FASTA format (Fig. 3D), and links to bibliographical resources including PubMed articles citing the protein and a PaperBlast query (Price and Arkin, 2017) (Fig. 3E). In the annotations tab, we can also see which co-expression cluster the transcript belongs to for embryogenesis and regeneration. Exploring these clusters is very useful for identifying co-expressed genes.

Exploring gene expression clusters


Co-expression analysis is particularly useful to identify genes that function in the same gene regulatory module. Co-expressed genes can also represent groups of genes that participate in a similar biological function. One method for identifying co-expressed genes is to cluster genes or transcripts by expression profile. For NvERTx, we performed fuzzy c-means clustering to regroup genes by expression profile, and provide those clusters in the 'Co-expression clusters' section of the site. These clusters can be used to identify transcripts that are co-expressed with a given gene of interest. Furthermore, comparing the membership of gene clusters during embryogenesis and regeneration can be used to identify groups of genes that function similarly during these processes.

The gene expression clusters can be browsed by either clicking a cluster in the 'Co-expression clusters' section (Fig. S1A) or by following a direct link from the annotation table of a transcript to its cluster (Fig. 3C). Using our previous example, *Nvbra*, from the annotations results, we can see that *Nvbra* participates in regeneration cluster two (R-2). When exploring R-2, we see all of the transcripts found in the cluster sorted by membership score. The score reflects how strongly a gene's expression matches the cluster core. By plotting several high-scoring transcripts, we can identify groups of co-expressed genes. For example, when we plot NvERTx.4.40781 (best NR hit: XP_015758878.1 forkhead box protein G1-like *Acropora digitifera*), NvERTx.4.57897 (best NR hit: AOP31964.1 dickkopf3-like 1 *N. vectensis*) and NvERTx.4.100808 (best NR hit: AAO27886.2 Brachyury protein *N. vectensis*), we see that they are indeed co-expressed with an initial expression peak between 4 hpa and 16 hpa, followed by a gradual rise from 36 hpa onward (Fig. S1Bii). By contrast, these genes are not co-expressed during embryogenesis and exhibit divergent expression patterns (Fig. S1Bi), raising the hypothesis that this particular grouping of genes is unique to regeneration. Using this method to find co-expressed genes is an effective way of identifying potential gene-regulatory modules and gene batteries. Importantly, assessing whether these genes are co-expressed during regeneration and embryogenesis can shed light on how these gene batteries are used or re-used during these two processes.

Differentially expressed genes

Comparing which genes are differentially expressed during embryogenesis and regeneration can provide important clues about

A



NvERTx - HOME **Co-Expression Clusters** **BLAST** **DE Genes** **About/ FAQ**

NvERTx - An embryogenesis & regeneration gene expression plotter

Welcome to NvERTx, an embryonic and regenerative transcriptome exploration tool. To learn more about the assembly check out the [About Page](#)

To get started you can enter a search term (gene name or JGI ID) in the sidebar or if you have a sequence you can BLAST it below. You can also explore the transcript clusters by clicking below.

References for regeneration and novel embryonic datasets: [Warner et al\(a\), 2017](#), [Warner et al\(b\), 2017](#). Re-analyzed embryonic datasets: [Fischer et al. 2014](#), [Tulin et al. 2013](#), and [Helm et al. 2013](#)

Search by gene name

Brachyury

Search!

Search by NvERTx ID

eg : NvERTx.4.100046 or 1C

eg : NvERTx.4.100046

eg : NvERTx.4.100046

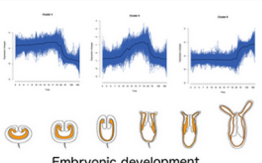
eg : NvERTx.4.100046

eg : NvERTx.4.100046

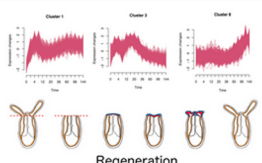
☐ Check for CPM (Regen only)

Plot!

Mine co-expression clusters



Embryonic development



Regeneration

BLASTn

Switch to tBLASTn

B

Search results

[click on a row to view the other nr hits](#)

Compare !

ID	Nemve1 hit	Nemve1 e-value	Uniprot id	Uniprot description	top nr_hit eval
<input type="checkbox"/> NvERTx.4.100808	gjl Nemve1 770 gw.137.1.1	0.0	Q17134	Brachyury protein homolog 1	gjl 122058623 gb AAO27886.2 Brachyury-like protein [Nematostella vectensis], 2e-119
<input type="checkbox"/> NvERTx.4.100809	gjl Nemve1 770 gw.137.1.1	0.0	A0A146ZVK6_FUNHE	Brachyury protein {ECO:0000313 EMBL:JAR70374.1}	None

C Blastplus results

TBLASTN 2.6.0+

Description

Compare !

contig	query	length	e-value	score	ident %
NvERTx.4.100809	AAI20808.1	3509	1.81112e-112	919.0	52.4
NvERTx.4.145171	AAI20808.1	985	2.94102e-57	490.0	47.5
NvERTx.4.145170	AAI20808.1	1116	1.15925e-56	490.0	48.7
NvERTx.4.73614	AAI20808.1	1420	7.95126e-55	484.0	45.8
NvERTx.4.73613	AAI20808.1	1624	8.87954e-55	488.0	45.8
NvERTx.4.145168	AAI20808.1	2491	1.86442e-54	497.0	51.9
NvERTx.4.159335	AAI20808.1	1335	2.02584e-53	473.0	48.9

Fig. 2. See next page for legend.

Fig. 2. NvERTx points of entry. (A) Screenshot of the NvERTx portal. (Ai) Users can search for genes using the gene name, *Nemve1* accession number, or NCBI GenBank accession number. (Aii) The transcriptome can also be searched using BLASTn or tBLASTn. (Aiii) Multiple transcripts can be queried simultaneously. (Aiv) Users can also directly explore co-expression clusters from embryogenesis and regeneration to identify groups of co-expressed genes. (B) Screen shot of results from searching the annotations using the term 'Brachyury'. (C) tBLASTn results using *Mus musculus* Brachyury as a query (GenBank AAI20808.1) identify several homologous transcripts. The top scoring isoform is reported first.

how genes are initially deployed during embryonic development and if/how they are reused during regeneration. Systematically analyzing genes differentially expressed during regeneration has allowed for the discovery of 'regeneration-specific' genes in other models, including axolotl, zebrafish and newts (Bryant et al., 2017; Knapp et al., 2013; Looso et al., 2013; Mathew et al., 2009). To facilitate these lines of inquiry using NvERTx, we provide the results of intra-data-set pairwise differential gene-expression testing on the summarized *Nemve1* genes. For each data set, we compare each time point to t_0 , which is defined as 0 hpa for the regeneration data, 7 hpf (the estimated beginning of zygotic transcription) for the Fischer and Smith (2013), Fischer et al. (2014) and Helm et al. (2013) data, and 24 hpf (the first time point sampled) for the in-house embryonic data. The results of this testing can be queried on the DE Genes page, and an interactive volcano plot displaying the negative \log_{10} false-discovery rate (FDR) as a function of fold change is generated (Fig. 4Ai,Bi). Each transcript is plotted as a single point, and multiple isoforms for a single gene are superimposed. Users can use the plot tools to click or drag-select the transcripts that are then displayed in the table below (Fig. 4Aii,Bii). These transcripts can then be compared by ticking the boxes in the table and querying the database as described above.

For example, when we select the three transcripts with the most dramatic fold change during embryogenesis from 7 hpf to 24 hpf (Fig. 4Ai, red dotted line box), we identify several transcripts corresponding to three predicted proteins (Fig. 4Aii). Querying the database shows that indeed these genes are highly expressed at 24 hpf (Fig. 4Aiii). Conversely, these genes show little variation during regeneration (Fig. 4Aiv). Likewise, when we examine differential expression during regeneration at 24 hpa, three genes show a large fold change (Fig. 4Bi, red dotted line box). Selecting these genes shows two predicted proteins and a T-box transcription factor (Fig. 4Bii). Comparing these genes, we see that all three are indeed upregulated at 24 hpa (Fig. 4Biv). Of these three, only the two predicted proteins show significant variation during embryogenesis (Fig. 4Biii, NvERTx.4.229217, NvERTx.4.119508), while no embryonic data are found for NvERTx.4.207772, meaning that this gene is not expressed at a detectable level in any of the embryonic data sets and could represent a 'regeneration-specific' gene.

Conclusion and future directions

NvERTx provides a platform to efficiently compare gene expression during embryogenesis and regeneration in *Nematostella*. Additionally, the comprehensive transcriptome provides high-quality transcript models that can be used to identify gene sequences. Using co-expression clusters, one can explore groups of genes that share similar expression patterns during embryogenesis and regeneration. Finally, mining the differentially expressed genes enables the identification of embryogenesis or 'regeneration-specific' genes. All of these tools are aimed at inspiring and building hypotheses concerning embryogenesis and regeneration for *Nematostella* and non-*Nematostella* researchers alike. Users can use their own groups of co-expressed genes to test

for conservation of regeneration gene batteries, or to explore gene expression clusters to identify genes that share expression, and examine these in their own models.

As this web application is intended to complement and expand upon existing resources, we provided transcript models that have been annotated using a variety of gene/protein databases (NR, trEMBL, *Nemve1*). Sequencing technologies are evolving to achieve longer reads, and assemblers will soon provide more robust transcriptomes; in the future, we plan to take advantage of these technologies to improve our transcript models. We also plan to grow the database as future data sets examining embryogenesis and regeneration emerge. Finally, we foresee merging this resource with an existing spatial gene expression database found at <http://www.kahikai.org/index.php?content=genes> (Ormestad et al., 2011). This will enable the identification of syn-expression groups, genes that are co-expressed both spatially and temporally (Niehrs and Pollet, 1999), and further facilitate studies comparing differential gene usage during embryogenesis and regeneration.

MATERIALS AND METHODS

Animal culture, spawning and amputation

Adult *N. vectensis* were cultured at 16°C in the dark in 1/3 strength artificial sea water (ASW) as previously described (Amiel et al., 2017). Spawning was induced by feeding the animals with oysters the day before and transferring the animals to a light table for 12 h. Regeneration experiments were performed using 6-week-old juveniles as previously described (Amiel et al., 2017).

Transcriptomic data sets

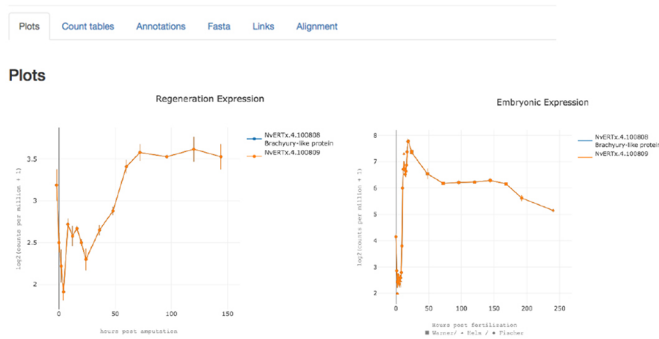
The sequences that served as input into our *de novo* assembly consisted of two data sets: one spanning the first 24 h of embryogenesis originally reported by Tulin et al. (2013), and another spanning the first 144 h of regeneration generated in-house (see 'Library preparation and sequencing' section). The embryonic data set was downloaded from the Woods Hole Open Access Server (<http://darchive.mblwhoilibrary.org/handle/1912/5613>, last accessed 1 June, 2017) and includes Illumina HiSeq 100 bp paired-end sequencing prepared from *Nematostella* embryos at 0, 6, 12, 18 and 24 hpf. The regeneration data set includes Illumina NextSeq 75 bp paired-end sequencing from regenerating *Nematostella* at -1 (uncut), 0, 2, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 96, 120 and 144 hpa (see 'Library preparation and sequencing' section for details). The sequence data that were used for transcript quantification are composed of four data sets. Three of the data sets spanned embryogenesis: one originally reported by Fischer and Smith (2013) and Fischer et al. (2014) (<http://darchive.mblwhoilibrary.org/handle/1912/5981>, last accessed 1 June, 2017) (Illumina HiSeq 100 bp paired-end replicates sampled hourly from 0 hpf to 19 hpf), a second embryonic data set originally reported in Helm et al. (2013) [NCBI Sequence Read Archive (SRA) project, PRJNA189768] (Illumina HiSeq 50 bp single-end replicates sampled from 2, 7, 12, 24, 120 and 240 hpf), and a third embryonic data set generated in house, sampled at 24, 48, 72, 96, 120, 144, 168 and 192 hpf (Illumina MiSeq 75Bp single-end replicates). The regenerative data used for quantification are the same as those used for the transcriptome assembly (Illumina NextSeq 75 bp single-end triplicates). Raw reads for the in-house generated data sets can be found in the NCBI SRA BioProject, under accession numbers PRJNA418421 and PRJNA419631.

Library preparation and sequencing

Two novel RNAseq data sets, one spanning embryonic, larval and postmetamorphic development and the other regeneration, were generated for this study. For the embryonic data set, ~250 embryos per time point were cultured in 1/3 strength ASW at 18°C. At each time point, 24, 48, 72, 96, 120, 144, 168 and 192 hpf, the embryos were transferred to 500 ml Tri Reagent and homogenized for 15 s with a pestle. The resulting lysate was snap frozen in liquid nitrogen and stored at -80°C. This was repeated to obtain duplicates for each time point. After all the samples were collected,

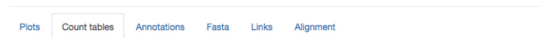
A

Results



B

Results



Count tables

Regeneration average counts (hours post amputation) -- log2											
ID	UC	0h	2h	4h	8h	12h	16h	20h	24h	36h	48h
NvERTx.4.100808	3.19	2.5	2.22	1.91	2.72	2.58	2.67	2.5	2.3	2.65	2.1
NvERTx.4.100809	3.19	2.5	2.22	1.91	2.72	2.58	2.67	2.5	2.3	2.65	2.1

Warner et al. average counts (hours post fertilization)

ID	24h	48h	72h	96h	120h
NvERTx.4.100808	7.38	6.54	6.18	6.21	6.22
NvERTx.4.100809	7.38	6.54	6.18	6.21	6.22

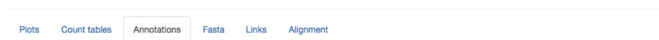
Fischer et al. counts (hours post fertilization)

ID	0h	1h	2h	3h	4h	5h	6h	7h	8h	9h	10h	11h	12h
NvERTx.4.100808	4.15	2.86	2.44	2.7	2.46	2.38	2.6	2.59	2.79	3.8	6.0	6.71	6.75
NvERTx.4.100809	4.15	2.86	2.44	2.7	2.46	2.38	2.6	2.59	2.79	3.8	6.0	6.71	6.75

Helm et al. counts (hours post fertilization)

C

Results



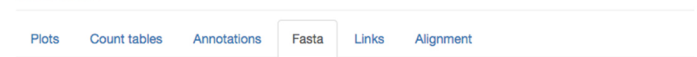
Annotations

click on a row to view the other nr hits

ID	Nvmer1 hit	Nvmer1 eval	Regen clust	Regen score	Embryo clust	Embryo score	Uniprot id	Uniprot description
NVERTx.4.100808	igf1[Nvmer1]770gw.137.1.1	0.0	R-2	0.78	E-7	0.98	Q17134	Brachyury protein homolog 1
NVERTx.4.100809	igf1[Nvmer1]770gw.137.1.1	0.0	R-2	0.78	E-7	0.98	A0A146ZVK6_FUNHE	Brachyury protein (ECO:0000313)EMBL:JAR70374.1

D

Results



Fasta

```

>NvERTx.4.100808
GGATAGAAAAATTTACCGAAAAGACACAAGTCTTTACTTGAAAAATTCAGAAATCTTGCTCTGATGCTAC
TTGAAGAGCCGTGAAGAGGCAAAAAGATAAGGCTTCTTGGCGACAAGTATTATGCACTTATTCTTGACA
CAGCTTTGTTCAATCTTTCAATGCGAAATGGTCTGAAGGAGGTAAGGGGGATCACTAATGTTTCGTTTCT
GGTTACATATCATAAAAGTTGGCGCGTCGGATAACAACCGTACAGTCGTCTCGCATCTTTTCCGAGACTCAG
AAAAATAAGTACAATCCGTTCTGCTAAGGCTTCTTGGACGCTAAAGAACGGCAGGAACAAAGAGGCTCTGG
TCTGTGCGAGGTCACGCGGCTACCGCATCATCACCTTCTCGTATCCTCATCTCTCC
CGCCCTACCCCAACCCATACCATAAGCGGACGGATTATCTTCTCTGTCACCCCAACAGTACTACGCCACG
CCTAGCTGTGCCAAGCAGCGCTCGCCATCACTCACAAGCGCGGGAATCCACGCCGGGTCTCACACTCTCACTTA
CGAAGAGAAAACAGACAAATCCAAGTTTCTGCAAGCGGGGGGCTGACTCTCCATTCAAGCTGGAACACCATACC
ATTGATTGAGGGGCTGCGCAAGTCAATTTAGTATATTGCTTTTGGCCCTGTTTCAGAGACCTTTTCGG
GAAATGGGGTAACAACCTAGAAATTTCTGATCACTTCTAACTTTCAACCGGAAGTGGCTATCTGTGCGGTT
AAAATAAATTCGTTCAAAATGACGCTTAACAATTTGTTAATAGATATCGAAAAACGAATATTTAAATGATTTTA
TCTGTTCAAATCATTTTTCATGTTCTTTGATTCTAAATAGACTGTATAGAAAAGCCTACTTCAGAAAGTAG
ATATTCAGACATTGAATCGGTTAAATAAATGTTT
>NvERTx.4.100809
CTACCAAAAACCGTTTTTTTTTTTTTTTTTTTATTTGAATGCGAACTTCCAAATATCAATACGCGAACATAGGG

```

E

Results

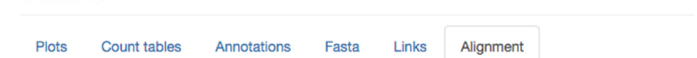


Links

ID	link	PaperBlast
NvERTx.4.100808	Link to PubMed (122058623)	Link to PaperBlast (AAC27886.2)
NvERTx.4.100809	No reference to NCBI found	

F

Results



MUSCLE Alignment

NvERTx.4.100808	-----
NvERTx.4.100809	cattccccgactcgcttaacttcggagcgactggatgaacagcctgttgcttttc
NvERTx.4.100808	-----CAGATCATGTAACTCTCTACA
NvERTx.4.100809	aaagtcacactcacaaataacaaactctgggggCAGATCATGTAACTCTCTACA
NvERTx.4.100808	CAAGTACGAACCCGGTTACATATCATAAAAGTTGGCGCTCGGATAACAACCGTACAGT
NvERTx.4.100809	CAAGTACGAACCCGGTTACATATCATAAAAGTTGGCGCTCGGATAACAACCGTACAGT
NvERTx.4.100808	CGTCTCGATTCTTTTCCGAGACTCAGTTTATAGAGTACCGCTTATCAAAACGAAGA
NvERTx.4.100809	CGTCTCGATTCTTTTCCGAGACTCAGTTTATAGAGTACCGCTTATCAAAACGAAGA
NvERTx.4.100808	GATCAGAGTCTAAAAATAAGTACAATCGTTTCGTAAGGCTTTCTGGACGCTAAGA
NvERTx.4.100809	GATCAGAGTCTAAAAATAAGTACAATCGTTTCGTAAGGCTTTCTGGACGCTAAGA
NvERTx.4.100808	ACGGCAGGAACAAAAGGAGGCTCTGGAGCAAGCTCGGAATCGCACTCAGCTTACTCCA
NvERTx.4.100809	ACGGCAGGAACAAAAGGAGGCTCTGGAGCAAGCTCGGAATCGCACTCAGCTTACTCCA
NvERTx.4.100808	ATATGGATGGTTCTGTGAGGGTCCACGCCGCTACCCGATCATCATCCCTTCTCT
NvERTx.4.100809	ATATGGATGGTTCTGTGAGGGTCCACGCCGCTACCCGATCATCATCCCTTCTCT

Fig. 3. Example results for the transcripts NvERTx.4.100808 and NvERTx.4.100809. (A) Expression plots for regeneration (left) and embryogenesis (right). The two transcripts are isoforms of the same genes so their expression profile plots and counts are equivalent. (B) Count data from each of the data sets. (C) Transcript annotations. (D) Sequences in FASTA format (the second sequence is shown cropped owing to space limitations). (E) Bibliographical resources including PubMed links and PaperBlast queries. (F) MUSCLE alignment to compare similar transcripts. The alignment shown is cropped to display the homologous region.

the RNA lysate was extracted using two phenol-chloroform extractions and precipitated with isopropanol. For the full extraction protocol, see Layden et al. (2013). The resulting nucleic acids were treated with the TURBO DNA-free kit from Invitrogen (AM1907) for 10 min at 37°C. The resulting RNA was quantified with a Qubit spectrometer, and RNA integrity was checked on an Agilent Bioanalyzer 2100. Then, 100 ng of RNA was used as input for an Illumina TruSeq Stranded mRNA Library Prep for NeoPrep Kit, and the libraries were prepared using an Illumina NeoPrep system. The 75 bp single-end sequencing was carried out on the NextSeq500 sequencer

of the Institute for Research on Cancer and Aging, IRCAN Genomics Core Facility, Nice, France.

For the regenerative data set, ~350 6-week-old *Nematostella* juveniles per time point were amputated below the pharynx. At each time point, -1, 0, 2, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 96, 120 and 144 hpa, the juveniles were transferred to 500 ml Tri Reagent and homogenized for 15 s with a pestle. The resulting lysate was snap frozen in liquid nitrogen and stored at -80°C. This was repeated for each of the three replicates. After all samples were collected, the RNA was extracted as described above for the embryonic samples.

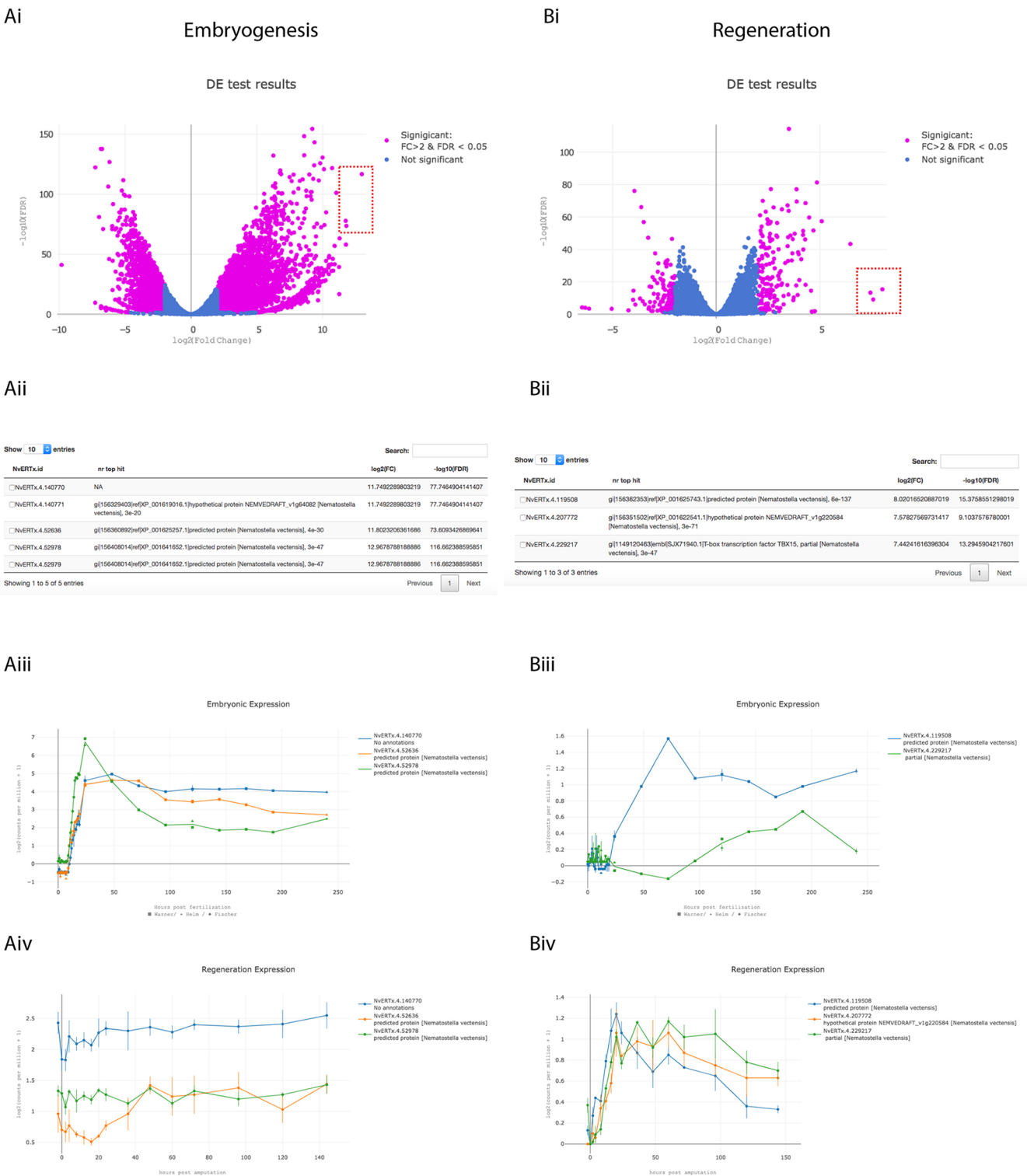


Fig. 4. Exploration of differentially expressed genes using the 'DE Genes' tool. Intra-data-set differential expression testing for summarized Nemve1 gene models. Each time point is compared with t_0 , defined as 0 hpa for the regeneration data, 7 hpf (the estimated beginning of zygotic transcription) for the data from Fischer and Smith (2013), Fischer et al. (2014) and Helm et al. (2013), and 24 hpf (the first time point sampled) for the in-house embryonic data. (Ai,Bi) When a user selects a comparison, a volcano plot displaying the $-\log_{10}(FDR)$ as a function of fold change is generated. Significant transcripts [$FC > \log_2(2)$ and $FDR < 0.05$] are colored magenta. Using the plot tools, multiple transcripts can be selected (red dotted line box, Ai,Bi) and a table is generated showing the corresponding transcripts (Aii,Bii). A user can compare these transcripts by ticking the corresponding box which fills the form on the left of the page. (Aiii,Aiv,Biii,Biv) The expression profiles shown here are then found on the results page.

Samples were stored in GenTegra-RNA stabilization reagent (GTR5100-S) and shipped to the NextGen Sequencing Core at the University of Southern California, CA, USA, for library preparation and sequencing. The samples were

prepared with a KAPA Stranded RNA Kit (KR0960). Two replicates were sequenced as 75-bp single-end sequencing, and one replicate as 75-bp paired-end sequencing, on an Illumina NextSeq500 sequencer.

RNAseq quality control

All reads from each data set were processed equivalently. Reads were first quality filtered to remove low-quality reads and adapter trimmed using trimmomatic (Bolger et al., 2014) and cutadapt (Martin, 2011), respectively.

Trinity de novo assembly

For the *de novo* assembly, paired-end reads from regeneration (−1, 2, 4, 8, 12, 16, 20, 24, 36, 48, 60, 72, 96, 120 and 144 hpa) and embryonic (0, 6, 12, 18 and 24 hpf) (Tulin et al., 2013) data sets were filtered of ribosomal sequences by aligning to *Nematostella* mitochondrial and ribosomal sequences using Bowtie2 (Langmead et al., 2009) and retaining the unmapped reads. These surviving reads were inputted into Trinity (v2.4.0) for assembly (Haas et al., 2013). To annotate the assembly, the ORFs were found using OrfPredictor (Min et al., 2005) and the resulting peptide sequences were compared with the NCBI NR database using the BLAST-like tool PLAST (Nguyen and Lavenier, 2009), with an e-value cutoff of 5e-5. The transcriptome was also compared with the UniProt databases, Swiss-Prot and trEMBL, using translated BLAST (BLASTx) and protein BLAST (BLASTp), respectively, with an e-value cutoff of 5e-5. The annotations were then compiled using the script totalannotation.py from the ‘Simple Fool’s Guide to Population Genomics via RNAseq’ (De Wit et al., 2012).

Quantification

To quantify the RNAseq data, single-end reads for each data set, regeneration and the three embryonic data sets [from Fischer and Smith (2013), Fischer et al. (2014), Helm et al. (2013) and this study], were aligned to the Trinity assembly using Bowtie2 (Langmead et al., 2009). Read counts were quantified using RSEM (Li and Dewey, 2011). To account for the many isoforms per gene reported by Trinity, transcripts were compared with the Nemve1 filtered gene models using BLASTn, and counts for transcripts with the same Nemve1 hit were combined. Transcripts with low-read counts, those that did not have fewer than five counts in at least 25% of the samples, were excluded. Each data set was then normalized separately using the R package edgeR and the counts per million (cpm) mapped reads were calculated (Robinson et al., 2010). Intra-data-set differential expression testing for each Nemve1 gene model was carried out by comparing each time point to t_0 using edgeR [$t_0=7$ hpf for the Helm et al. (2013), Fischer and Smith (2013) and Fischer et al. (2014) data sets; $t_0=24$ hpf for the in-house embryonic data set; and $t_0=0$ hpa for the regeneration data set]. A significantly differentially expressed gene is defined as having an absolute fold change (FC) >2 and a false discovery rate (FDR) <0.05. To correct for batch effects of the embryonic data sets after normalization, the R function ComBat from the SVA packages was used on $\log_2(\text{cpm}+1)$ transformed data using time point as a categorical covariate (Leek et al., 2012).

RNAseq quality assurance

To validate the accuracy of the in-house generated data sets (regeneration, embryonic) we performed RT-qPCR for several genes that exhibited variation across several time points. First, 500 ng of the same RNA used for library preparation for the regeneration and in-house embryonic data sets was used as input for a reverse transcription reaction using the Iscript Reverse Transcription Supremix Kit from Bio-Rad (1708840). The resultant cDNA was used as input for an RT-qPCR reaction using the FastStart Universal SYBR Green Master Mix from Sigma Aldrich (04913850001), and analyzed using an Applied Biosystems 7900HT 384-well plate qPCR machine, running 40 cycles with an annealing temperature of 60°C. Relative fold change (RFC) was calculated using the equation:

$$RFC = \frac{(E^{\Delta Ct_{\text{Target}}})}{(E^{\Delta Ct_{\text{Reference}}})}$$

ΔCt is the difference of the crossing threshold at the reference time point (defined as 24 hpf for the in-house embryonic data and 0 hpa for the regeneration data) and time point of interest, and E is the efficiency of the primer pair (Pfaffl et al., 2002). Actin was used as a reference gene (Forward: 5'-GGACAGGTCATCACCATTGGCAAC-3'; Reverse: 5'-CG-GATTCATACCCAGAAAGGAGG-3'; efficiency, 2.11). All other primer

sequences, efficiencies and resultant traces are listed in Figs S2 and S3. Overall, the RT-qPCR expression profiles correlate with those of the RNAseq data sets (Figs S2 and S3).

Fuzzy c-means clustering and gene ontology term enrichment

The expression profiles for each Nemve1 gene model were clustered using the R package mFuzz (Kumar and Futschik, 2007) on the batch-corrected combined embryonic data set and the regeneration data set separately. The cluster number was set to 9 for the regeneration data and 8 for the embryonic data sets as these numbers represented the point at which the centroid distance between clusters did not significantly decrease when new clusters were added (inflection point). For each cluster, gene ontology (GO) term enrichment was calculated using a Fisher’s exact test and the R package topGO on the GO terms identified from comparing the transcriptome with the UniProt database (all identified GO terms were used as a background model). The resulting GO term list was reduced and plotted using a modified R script based on REVIGO (Supek et al., 2011).

NvERTx website

The website was constructed using the Django python framework (<https://www.djangoproject.com/>). Plots are generated using the Plotly javascript library (<https://plot.ly/javascript/>). The source code for the website can be found at https://github.com/IRCAN/NvER_plotter_django. Data sets from the database can be found at http://ircan.unice.fr/ER/ER_plotter/about.

Acknowledgements

We thank Valerie Carlin for animal care; Christian Baudoin from the IRCAN Genomics Core Facility, and Charles Nicolet from the University of Southern California Sequencing Core Facility, for preparation and sequencing of the RNAseq libraries; and the IRCAN Genomics Core Facility for providing access to the NextSeq500 sequencer.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: J.F.W., E.R.; Methodology: J.F.W., V.G., E.R.; Software: J.F.W., V.G.; Validation: J.F.W., E.R.; Formal analysis: J.F.W., V.G.; Investigation: J.F.W., A.R.A., H.J., K.N.; Resources: E.R.; Data curation: J.F.W., V.G.; Writing - original draft: J.F.W., E.R.; Writing - review & editing: J.F.W., A.R.A., E.R.; Visualization: J.F.W., V.G., E.R.; Supervision: E.R.; Project administration: J.F.W., E.R.; Funding acquisition: E.R.

Funding

This work was supported by ATIP-Avenir (Institut National de la Santé et de la Recherche Médicale and Centre National de la Recherche Scientifique) funded by the Plan Cancer (Institut National Du Cancer) [C13992AS], Seventh Framework Programme [631665], Association pour la Recherche sur le Cancer [PJA 2014120186 to E.R.; PDF20141202150 to J.F.W.], Fondation pour la Recherche Médicale [SPF20130526781 to A.R.A.], Ligue Contre le Cancer [to K.N.] and Ministère de l’Enseignement Supérieur et de la Recherche [to H.J.].

Data availability

In-house data sets generated for this study are available at the NCBI SRA BioProject, under accession numbers PRJNA418421 and PRJNA419631. The dataset originally reported by Helm et al. (2013) is available at the NCBI SRA BioProject, under accession number PRJNA189768. The dataset originally reported by Fischer and Smith (2013) is available in the Woods Hole Open Access Server (<http://darchive.mblwhoilibrary.org/handle/1912/5981>). The dataset originally reported by Tulin et al. (2013) is available in the Woods Hole Open Access Server (<http://darchive.mblwhoilibrary.org/handle/1912/5613>).

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.162867.supplemental>

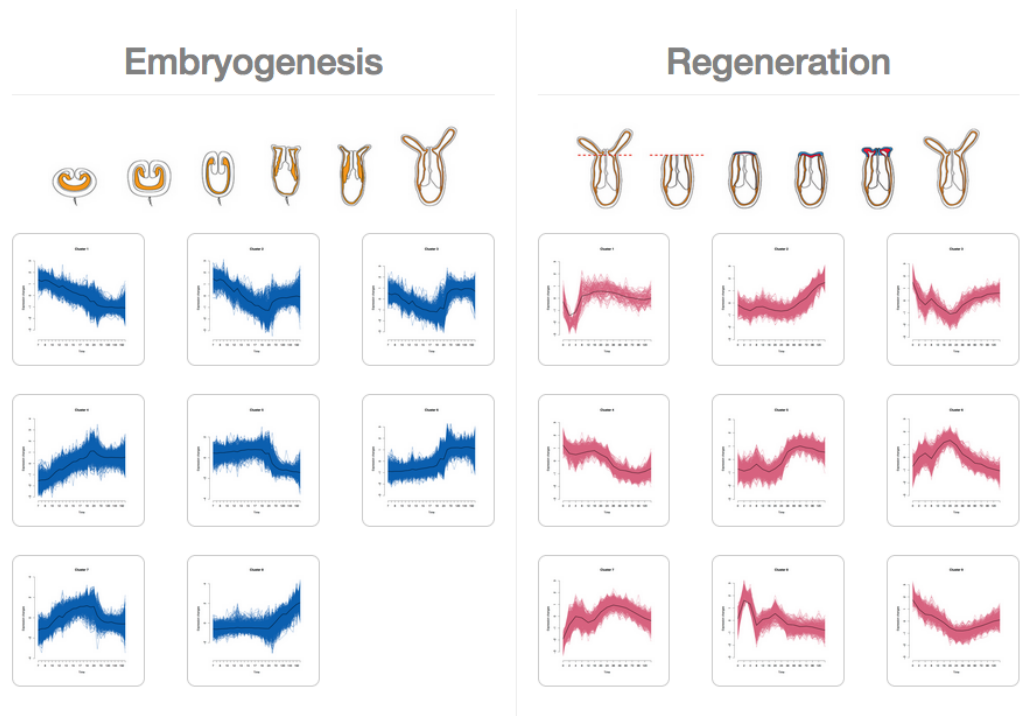
References

Amiel, A. R., Johnston, H. T., Nedoncelle, K., Warner, J. F., Ferreira, S. and Röttinger, E. (2015). Characterization of morphological and cellular events underlying oral regeneration in the sea anemone, *Nematostella vectensis*. *Int. J. Mol. Sci.* **16**, 28449–28471.

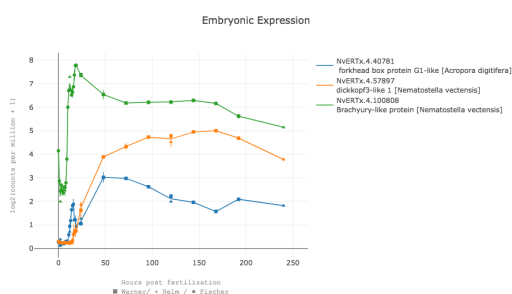
- Amiel, A. R., Johnston, H., Chock, T., Dahlin, P., Iglesias, M., Layden, M., Röttinger, E. and Martindale, M. Q. (2017). A bipolar role of the transcription factor ERG for cnidarian germ layer formation and apical domain patterning. *Dev. Biol.* **430**, 346–361.
- Binari, L. A., Lewis, G. M. and Kucenas, S. (2013). Perineurial glia require Notch signaling during motor nerve development but not regeneration. *J. Neurosci.* **33**, 4241–4252.
- Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
- Bossert, P. E., Dunn, M. P. and Thomsen, G. H. (2013). A staging system for the regeneration of a polyp from the aboral physa of the anthozoan Cnidarian *Nematostella vectensis*. *Dev. Dyn.* **242**, 1320–1331.
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G. et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**, 762–776.
- Burton, P. M. and Finnerty, J. R. (2009). Conserved and novel gene expression between regeneration and asexual fission in *Nematostella vectensis*. *Dev. Genes Evol.* **219**, 79–87.
- Carlson, M. R., Komine, Y., Bryant, S. V. and Gardiner, D. M. (2001). Expression of Hoxb13 and Hoxc10 in developing and regenerating Axolotl limbs and tails. *Dev. Biol.* **229**, 396–406.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X. et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13.
- Darling, J. A., Reitzel, A. R., Burton, P. M., Mazza, M. E., Ryan, J. F., Sullivan, J. C. and Finnerty, J. R. (2005). Rising starlet: the starlet sea anemone, *Nematostella vectensis*. *BioEssays* **27**, 211–221.
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., Therikildsen, N. O., Morikawa, M. and Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* **12**, 1058–1067.
- DuBuc, T. Q., Dubuc, T. Q., Traylor-Knowles, N., Traylor-Knowles, N. and Martindale, M. Q. (2014). Initiating a regenerative response: cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. *BMC Biol.* **12**, 1.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Finnerty, J. R., Pang, K., Burton, P., Paulson, D. and Martindale, M. Q. (2004). Origins of bilateral symmetry: hox and dpp expression in a sea anemone. *Science* **304**, 1335–1337.
- Fischer, A. H. L. and Smith, J. (2013). *Nematostella* High-density RNAseq time-course. *Woods Hole Open Access Server*. <http://dx.doi.org/10.1575/1912/5981>.
- Fischer, A. H. L., Mozzherin, D., Eren, A. M., Lans, K. D., Wilson, N., Cosentino, C. and Smith, J. (2014). SeaBase: a multispecies transcriptomic resource and platform for gene network inference. *Integr. Comp. Biol.* **54**, 250–263.
- Gardiner, D. M., Blumberg, B., Komine, Y. and Bryant, S. V. (1995). Regulation of HoxA expression in developing and regenerating axolotl limbs. *Development* **121**, 1731–1741.
- Genikhovich, G., Fried, P., Prünster, M. M., Schinko, J. B., Gilles, A. F., Fredman, D., Meier, K., Iber, D. and Technau, U. (2015). Axis patterning by BMPs: cnidarian network reveals evolutionary constraints. *Cell Rep.* **10**, 1646–1654.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M. et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512.
- Habermann, B., Bebin, A.-G., Herklotz, S., Volkmer, M., Eckelt, K., Pehlke, K., Epperlein, H. H., Schackert, H. K., Wiebe, G. and Tanaka, E. M. (2004). An Ambystoma mexicanum EST sequencing project: analysis of 17,352 expressed sequence tags from embryonic and regenerating blastema cDNA libraries. *Genome Biol.* **5**, R67.
- Helm, R., Siebert, S., Tulin, S., Smith, J. and Dunn, C. (2013). Characterization of differential transcript abundance through time during *Nematostella vectensis* development. *BMC Genomics* **14**, 266.
- Imokawa, Y. and Yoshizato, K. (1997). Expression of Sonic hedgehog gene in regenerating newt limb blastemas recapitulates that in developing limb buds. *Proc. Natl. Acad. Sci. USA* **94**, 9159–9164.
- Kaloulis, K., Chera, S., Hassel, M., Gauchat, D. and Galliot, B. (2004). Reactivation of developmental programs: the cAMP-response element-binding protein pathway is involved in hydra head regeneration. *Proc. Natl. Acad. Sci. USA* **101**, 2363–2368.
- Katz, M. G., Fargnoli, A. S., Kendle, A. P., Hajjar, R. J. and Bridges, C. R. (2015). The role of microRNAs in cardiac development and regenerative capacity. *Am. J. Physiol. Heart Circ. Physiol.* **310**, H528–H541.
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664.
- Knapp, D., Schulz, H., Rascon, C. A., Volkmer, M., Scholz, J., Nacu, E., Le, M., Novozhilov, S., Tazaki, A., Protze, S. et al. (2013). Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. *PLoS ONE* **8**, e61352.
- Kumar, L. and Futschik, M. E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **23**, 5–7.
- Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M. Q. et al. (2005). Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156–160.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Layden, M. J. and Martindale, M. Q. (2014). Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. *EvoDevo* **5**, 30.
- Layden, M. J., Boekhout, M. and Martindale, M. Q. (2012). *Nematostella vectensis* achaete-scute homolog NvashA regulates embryonic ectodermal neurogenesis and represents an ancient component of the metazoan neural specification pathway. *Development* **139**, 1013–1022.
- Layden, M. J., Röttinger, E., Wolenski, F. S., Gilmore, T. D. and Martindale, M. Q. (2013). Microinjection of mRNA or morpholinos for reverse genetic analysis in the starlet sea anemone, *Nematostella vectensis*. *Nat. Protoc.* **8**, 924–934.
- Layden, M. J., Rentzsch, F. and Röttinger, E. (2016). The rise of the starlet sea anemone *Nematostella vectensis* as a model system to investigate development and regeneration. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 408–428.
- Leclère, L., Bause, M., Sinigaglia, C., Steger, J. and Rentzsch, F. (2016). Development of the aboral domain in *Nematostella* requires β -catenin and the opposing activities of Six3/6 and Frizzled5/8. *Development* **143**, 1766–1777.
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883.
- Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
- Looso, M., Preussner, J., Sousounis, K., Bruckskotten, M., Michel, C. S., Lignelli, E., Reinhardt, R., Höffner, S., Krüger, M., Tsonis, P. A. et al. (2013). A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol.* **14**, R16.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12.
- Mathew, L. K., Sengupta, S., Franzosa, J. A., Perry, J., La Du, J., Andreassen, E. A. and Tanguay, R. L. (2009). Comparative expression profiling reveals an essential role for raldh2 in epimorphic regeneration. *J. Biol. Chem.* **284**, 33642–33653.
- Matus, D. Q., Pang, K., Marlow, H., Dunn, C. W., Thomsen, G. H. and Martindale, M. Q. (2006). Molecular evidence for deep evolutionary roots of bilaterality in animal development. *Proc. Natl. Acad. Sci. USA* **103**, 11195–11200.
- Matus, D. Q., Magie, C. R., Pang, K., Martindale, M. Q. and Thomsen, G. H. (2008). The Hedgehog gene family of the cnidarian, *Nematostella vectensis*, and implications for understanding metazoan Hedgehog pathway evolution. *Dev. Biol.* **313**, 501–518.
- Meyer, E. J., Ikmi, A. and Gibson, M. C. (2011). Interkinetic nuclear migration is a broadly conserved feature of cell division in pseudostratified epithelia. *Curr. Biol.* **21**, 485–491.
- Millimaki, B. B., Sweet, E. M. and Riley, B. B. (2010). Sox2 is required for maintenance and regeneration, but not initial development, of hair cells in the zebrafish inner ear. *Dev. Biol.* **338**, 262–269.
- Min, X. J., Butler, G., Storms, R. and Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **33**, W677–W680.
- Nguyen, V. H. and Lavenier, D. (2009). PLAST: parallel local alignment search tool for database comparison. *BMC Bioinformatics* **10**, 329.
- Niehhs, C. and Pollet, N. (1999). Synexpression groups in eukaryotes. *Nature* **402**, 483–487.
- Ormestad, M., Martindale, M. Q. and Röttinger, E. (2011). A comparative gene expression database for invertebrates. *EvoDevo* **2**, 17.
- Özpolat, B. D., Zapata, M., Daniel Frugé, J., Coote, J., Lee, J., Muneoka, K. and Anderson, R. (2012). Regeneration of the elbow joint in the developing chick embryo recapitulates development. *Dev. Biol.* **372**, 229–238.
- Passamanek, Y. J. and Martindale, M. Q. (2012). Cell proliferation is necessary for the regeneration of oral structures in the anthozoan cnidarian *Nematostella vectensis*. *BMC Dev. Biol.* **12**, 34.
- Pfaffl, M. W., Horgan, G. W. and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36.
- Price, M. N. and Arkin, A. P. (2017). PaperBLAST: text mining papers for information about homologs. *mSystems* **2**, e00039–e00017.
- Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V. et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94.

- Reitzel, A. M., Burton, P. M., Krone, C. and Finnerty, J. R.** (2007). Comparison of developmental trajectories in the starlet sea anemone *Nematostella vectensis*: embryogenesis, regeneration, and two forms of asexual fission. *Invertebr. Biol.* **126**, 99-112.
- Rentzsch, F., Anton, R., Saina, M., Hammerschmidt, M., Holstein, T. W. and Technau, U.** (2006). Asymmetric expression of the BMP antagonists chordin and gremlin in the sea anemone *Nematostella vectensis*: implications for the evolution of axial patterning. *Dev. Biol.* **296**, 375-387.
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K.** (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Röttinger, E., Dahlin, P. and Martindale, M. Q.** (2012). A framework for the establishment of a cnidarian gene regulatory network for "endomesoderm" specification: the inputs of β -catenin/TCF signaling. *PLoS Genet.* **8**, e1003164.
- Schaffer, A. A., Bazarsky, M., Levy, K., Chalifa-Caspi, V. and Gat, U.** (2016). A transcriptional time-course analysis of oral vs. aboral whole-body regeneration in the Sea anemone *Nematostella vectensis*. *BMC Genomics* **17**, 718.
- Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T.** (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800.
- Technau, U. and Schwaiger, M.** (2015). Recent advances in genomics and transcriptomics of cnidarians. *Mar Genomics* **24**, 131-138.
- Torok, M. A., Gardiner, D. M., Shubin, N. H. and Bryant, S. V.** (1998). Expression of HoxD genes in developing and regenerating axolotl limbs. *Dev. Biol.* **200**, 225-233.
- Trevino, M., Stefanik, D. J., Rodriguez, R., Harmon, S. and Burton, P. M.** (2011). Induction of canonical Wnt signaling by alsterpaullone is sufficient for oral tissue fate during regeneration and embryogenesis in *Nematostella vectensis*. *Dev. Dyn.* **240**, 2673-2679.
- Tulin, S., Aguiar, D., Istrail, S. and Smith, J.** (2013). A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for de novo assembly in emerging model systems. *EvoDevo* **4**, 16.
- Wang, Y.-H. and Beck, C. W.** (2014). Distal expression of sprouty (spry) genes during *Xenopus laevis* limb development and regeneration. *Gene Expr. Patterns* **15**, 61-66.
- Wikramanayake, A. H., Hong, M., Lee, P. N., Pang, K., Byrum, C. A., Bince, J. M., Xu, R. and Martindale, M. Q.** (2003). An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation. *Nature* **426**, 446-450.

A



Bi



Bii

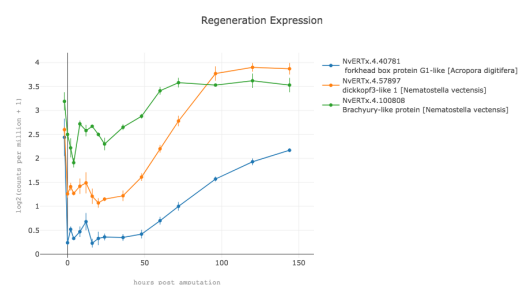


Figure S1: The NvERTx Co-Expression Clusters Page. A) Screen shot of the Co-expression Clusters page. Users can directly explore co-expression clusters to identify groups of genes that share expression patterns during embryogenesis (blue) or regeneration (red). B) Example output plots from NvERTx comparing multiple gene expression patterns. Three genes from regeneration cluster 2, *Nvbra* (NvERTx.4.100808, yellow), *Nvdickkopf3* (NvERTx.4.57897, red), and a FoxG1-like protein (NvERTx.4.40781, blue) are co-expressed during regeneration (i) but not embryogenesis (ii).

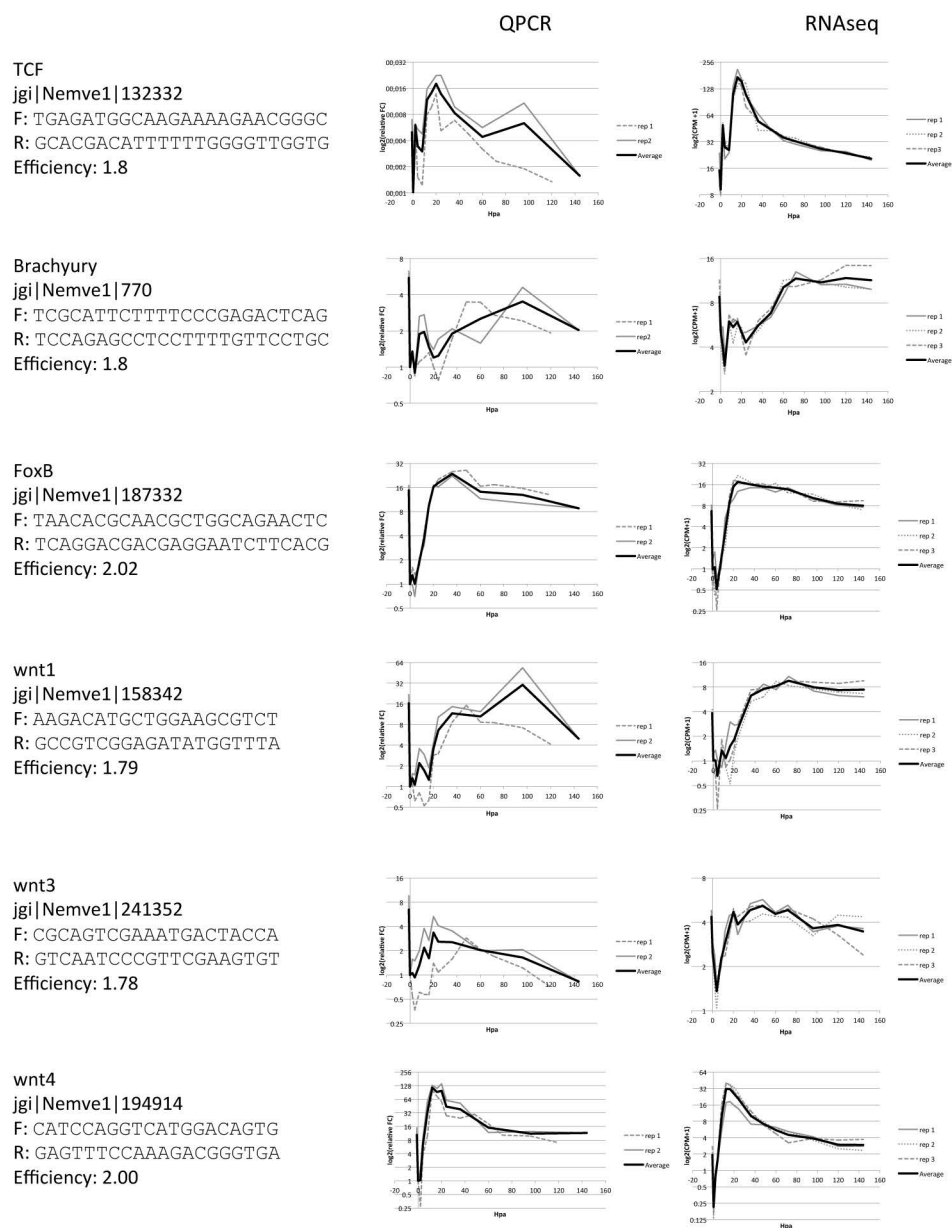


Figure S2: RT-QPCR of selected genes compared to regeneration RNAseq data. RT-QPCR traces (replicates grey, average black) are shown in terms of $\log_2(\text{relative fold change})$ versus hours post regeneration (Hpa). RNAseq traces (replicates grey, average black) are shown in terms of $\log_2(\text{counts per million} + 1)$ versus Hpa.

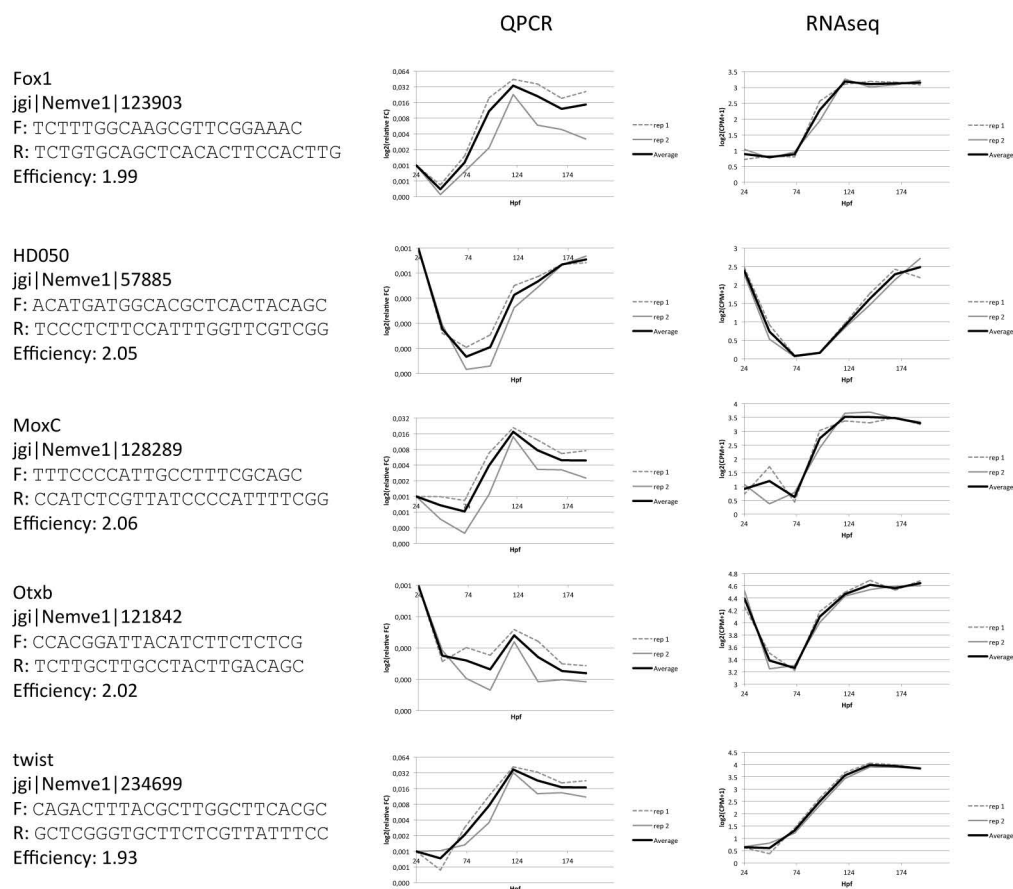


Figure S3: RT-QPCR of selected genes compared to embryogenesis RNAseq data (Warner dataset). RT-QPCR traces (replicates grey, average black) are shown in terms of $\log_2(\text{relative fold change})$ versus hours post fertilization (Hpf). RNAseq traces (replicates grey, average black) are shown in terms of $\log_2(\text{counts per million} + 1)$ versus Hpf.

Table S1: Assembly statistics from Trinity assembly.

<i>N50</i>	1678
<i>Median contig length</i>	384
<i>Average contig length</i>	837.30
<i>Total assembled bases</i>	196247212

Table S2: NvERTx.4 annotation statistics. 234,381 transcripts map to 19,565 unique Nemve1 genes.

<i>Total assembled transcripts</i>	234381
<i>Transcripts with ORF</i>	231294
<i>Transcripts without ORF</i>	3087
<i>Transcripts with hit to nr</i>	85475
<i>Transcripts with hit to uniprot</i>	69335
<i>Transcripts with hit to Nemve1</i>	102581
<i>Unique Nemve1 'genes'</i>	19565
<i>Transcripts without annotation</i>	110531