A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors

Highlights

- Creation of a transcriptome with near-complete sequence data for 88% of axolotl genes
- Expression analyses identify tissue-enriched transcripts for key tissues
- The RNA-binding protein cirbp plays a cytoprotective role in limb regeneration
- Knockdown and overexpression of kazald1 in blastema cells impair limb regeneration

Authors

Donald M. Bryant, Kimberly Johnson, Tia DiTommaso, ..., Aviv Regev, Brian J. Haas, Jessica L. Whited

Correspondence

bhaas@broadinstitute.org (B.J.H.), jwhited@bwh.harvard.edu (J.L.W.)

In Brief

Discovery of genes driving axolotl limb regeneration has been challenging, due to limited genomic resources. Bryant et al. have created a transcriptome with near-complete sequence information for most axolotl genes, identified transcriptional profiles that distinguish blastemas from differentiated limb tissues, and uncovered functional roles for cirbp and kazald1 in limb regeneration.

Accession Numbers

GSE92429
A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors

Donald M. Bryant,1,6 Kimberly Johnson,1,6 Tia DiTommaso,1 Timothy Tickle,2 Matthew Brian Couger,3 Duygu Payzin-Dogru,1 Tae J. Lee,1 Nicholas D. Leigh,1 Tzu-Hsing Kuo,1 Francis G. Davis,1 Joel Bateman,1 Sevara Bryant,1 Anna R. Guzikowski,1 Stephanie L. Tsai,4 Steven Coyne,1 William W. Ye,1 Robert M. Freeman, Jr.,1 Leonid Peshkin,6 Clifford J. Tabin,2 Aviv Regev,2 Brian J. Haas,2,* and Jessica L. Whited1,7,*

1Harvard Medical School, Harvard Stem Cell Institute, and Department of Orthopedic Surgery, Brigham & Women’s Hospital, 65 Landsdowne St., Cambridge, MA 02139, USA
2Broad Institute of MIT and Harvard and Klarman Cell Observatory, 7 Cambridge Center, Cambridge, MA 02142, USA
3Department of Microbiology and Molecular Genetics, Oklahoma State University, 307 Life Sciences East, Stillwater, OK 74078, USA
4Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
5Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA
6Co-first author
7Lead Contact
*Correspondence: bhaas@broadinstitute.org (B.J.H.), jwhited@bwh.harvard.edu (J.L.W.)
http://dx.doi.org/10.1016/j.celrep.2016.12.063

SUMMARY

Mammals have extremely limited regenerative capabilities; however, axolotls are profoundly regenerative and can replace entire limbs. The mechanisms underlying limb regeneration remain poorly understood, partly because the enormous and incompletely sequenced genomes of axolotls have hindered the study of genes facilitating regeneration. We assembled and annotated a de novo transcriptome using RNA-sequencing profiles for a broad spectrum of tissues that is estimated to have near-complete sequence information for 88% of axolotl genes. We devised expression analyses that identified the axolotl orthologs of cirbp and kazald1 as highly expressed and enriched in blastemas. Using morpholino anti-sense oligonucleotides, we find evidence that cirbp plays a cytoprotective role during limb regeneration whereas manipulation of kazald1 expression disrupts regeneration. Our transcriptome and annotation resources greatly complement previous transcriptomic studies and will be a valuable resource for future research in regenerative biology.

INTRODUCTION

The limited capacity of humans to regenerate many tissues, organs, and appendages is a formidable clinical hurdle (Ziegler-Graham et al., 2008). Conversely, some animals, including invertebrates such as planaria and vertebrates such as amphibians, have remarkable regenerative capacity. Among those, many salamanders, including axolotls, can regenerate entire limbs throughout life (reviewed in Whited and Tabin, 2009). Elucidating the molecular mechanisms that enable such profound regenerative capacity may provide key insights relevant to human regenerative medicine.

The axolotl community has made significant strides in advancing our understanding of limb regeneration, but our knowledge of the molecular mechanisms that underlie axolotl regeneration is still very limited. Unbiased genomics and transcriptomics can often unlock the molecular components of systems that have not been genetically tractable. Unfortunately, the axolotl genome remains mostly unsequenced and poses major challenges at ~32 Gb in size (Keinath et al., 2015; Smith et al., 2009; Straus, 1971). RNA sequencing followed by de novo transcriptome assembly (Haas et al., 2013; Robertson et al., 2010; Schulz et al., 2012) has offered investigators an alternative for identifying near-full-length transcripts and performing differential gene expression analyses without genome mapping. Recent axolotl transcriptome studies (Knapp et al., 2013; Li et al., 2014; McCusker et al., 2015; Monaghan et al., 2009; Stewart et al., 2013; Voss et al., 2015; Wu et al., 2013) have focused on and significantly advanced our understanding of the changes in transcription over time in the regenerating portion of the limb. However, an important missing component of the existing datasets is deep transcriptional information about each of the presumed parent tissue types within the limb, which contribute progenitors and serve as the template for the future regenerate limb. Thus, examining the transcripts that define them in the differentiated, homeostatic state will be critical for future comparisons with progenitor cells along the temporal path of regeneration.

Here we combined RNA sequencing (RNA-seq) of diverse tissues with de novo transcriptome assembly, computational analysis, and experimental validation to develop a systematic map of the axolotl transcriptome. This assembly facilitated identification of specific transcripts and classes of genes whose expression is associated with successful limb regeneration. We experimentally validated our transcriptome’s accuracy by analyzing mRNA expression of identified transcripts using in situ hybridization. Furthermore, we experimentally modified the expression of cirbp.
A. Select tissues isolated and RNA-Seq performed
- Unamputated limb
- Hand
- Lower arm
- Upper elbow arm
- Gill
- Testis
- Ovary
- Embryo
- Bone

Build a comprehensive de novo assembly of transcriptome

Perform pairwise comparisons between transcripts expressed by each tissue

Map to transcriptome and assign orthology

Validate and perform functional experimentation

B. Number of transcripts vs. Min FPKM
- E90 = 26,378

C. Contig N50 vs. Total expression (%)
- E90N50 = 3511

D. Heatmap showing correlation between different tissues and cell types
- Pearson correlation

(legend on next page)
and kazald1 (two blastema-enriched transcripts) and uncovered functional roles for these genes in axolotl limb regeneration. The transcriptome and analyses that we provide significantly complement prior research and will be an important resource for future studies of limb regeneration, as well as for inquiries using the axolotl that extend beyond those of limb regeneration.

RESULTS

An RNA-Seq Catalog for Limb Tissues and Regeneration
To build a reference map of axolotl limb regeneration, we profiled 42 samples across 16 different tissues (Figure 1A). First, we profiled intact, unamputated limbs to reflect the starting—and end—point for limb regeneration. We sampled four positions along the proximal (shoulder) to distal (fingertip) axis to identify any location-specific transcriptional differences. Second, we sequenced the blastema, a bud-like outgrowth at the tip of the regenerating limb that contains activated progenitor cells that regenerate the internal structures of the limb. To identify transcripts whose regulation distinguishes blastema cells, we removed the regenerated epithelium of blastemas at the medium-bud stage before it began to differentiate into the various tissues of the regenerating limb. Third, it is thought that axolotls are able to redeploy transcriptional programs that guide embryogenesis during limb regeneration, and a recent study demonstrated that genes with roles in germline cell renewal are required for axolotl limb regeneration (Zhu et al., 2012). To facilitate the systematic identification of germline and embryonic transcriptional programs that are reactivated during regeneration, we generated transcriptional profiles for testes, ovaries, and embryos (one-cell to pre-hatch stage). Fourth, following amputation, individual tissues within the limb, including skeletal muscle, cartilage, bone, and blood vessels, may contribute activated progenitor cells to the regenerating limb (Kragl et al., 2009; Muneoka et al., 1986; Sandoval-Guzman et al., 2014; reviewed in Knapp and Tanaka, 2012). We therefore sampled and analyzed each of these tissues to define tissue-specific expression and marker transcripts for the differentiated tissue types in limbs. We also included transcripts from heart and gill filaments in our assembly to increase the comprehensiveness of the transcriptome and to provide resources for researchers interested in pursuing questions outside of limb regeneration. Fifth, the blastema is an autonomous unit, programmed from the onset with spatial coordinates that instruct the regrowth of precisely the portion of the limb that has been lost (Crawford and Stocum, 1988a, 1988b; Echeverri and Tanaka, 2005; McCusker and Gardiner, 2013; Mercader et al., 2005; Stocum and Melton, 1977), such that amputation of the hand (a “distal” amputation) results in regeneration of a hand, but amputation of an entire arm (a “proximal” amputation) has been lost (Crawford and Stocum, 1988a, 1988b; Echeverri in other species, of which over a third (9,950) are significantly enriched in the E90 set. Trinotate reported 109,180 transcripts matching protein’s length. These transcripts likely represent fully matching 13,501 proteins across at least 80% of the predicted coding regions separately for protein homology by BLAST search against Swiss-Prot, and annotates the coding regions for domain content, signal peptides, and transmembrane domains. Trinotate reports its annotations in a convenient tab-delimited file (Data S3).

We applied Trinotate to the 1.6 million transcripts in our assembly, finding a large number of likely full-length transcripts enriched in the E90 set. Trinotate reported 109,180 transcripts matching 29,529 unique Swiss-Prot proteins, with 27,056 Trinity transcripts matching 13,501 proteins across at least 80% of the matching protein’s length. These transcripts likely represent fully or near-fully reconstructed transcripts with detectable homologs in other species, of which over a third (9,950) are significantly enriched (22-fold; p value < 2.2 × 10^-16, Fisher’s exact test) among the E90 transcript set.

By subjecting our transcriptome to BUSCO analysis (Simão et al., 2015), we identified by this measure that our transcriptome and analyses that we provide significantly complement prior research and will be an important resource for future studies of limb regeneration, as well as for inquiries using the axolotl that extend beyond those of limb regeneration.

A De Novo Assembled Axolotl Transcriptome
We assembled 42 RNA-seq samples totaling ~1.3 billion 100-base paired-end reads (Table S1) using Trinity (Grabherr et al., 2011; Haas et al., 2013). We first combined reads from all samples and performed in silico normalization, retaining 6.6% of reads for assembly (86 million paired-end reads). We then used Trinity (Table S1) to generate an assembly of 1,554,055 transcript contigs clustered into 1,388,798 “gene” groupings (Data S1 and S2; median transcript length: 288 bases, N50 of 606 bases). Although our in silico normalization filtered out the vast majority of the reads prior to assembly, 80% of the original reads mapped back to the assembly, with most mapping as properly paired reads (Figure S1A). The majority of the transcriptome corresponds to lowly expressed transcript contigs (Figure 1B), as 90% of the total transcription is represented by an “E90 transcript set” of 26,378 transcripts (20,506 genes, E90-N50 of 3,511) (Table S1; Figure 1C).

Functional Annotation with Trinotate and Assessment of Transcriptome Completeness
Annotating de novo assembled transcripts in the absence of an assembled genome, as in the case of the axolotl, remains a bioinformatics challenge with few available tools. We therefore developed Trinotate (http://trinotate.github.io), an annotation protocol and toolkit for de novo assembled transcriptomes. Tri- notate extracts predicted coding regions using TransDecoder (http://transdecoder.github.io), searches both the entire transcripts and the predicted coding regions separately for protein homology by BLAST search against Swiss-Prot, and annotates the coding regions for domain content, signal peptides, and transmembrane domains. Trinotate reports its annotations in a convenient tab-delimited file (Data S3).

We applied Trinotate to the 1.6 million transcripts in our assembly, finding a large number of likely full-length transcripts enriched in the E90 set. Trinotate reported 109,180 transcripts matching 29,529 unique Swiss-Prot proteins, with 27,056 Trinity transcripts matching 13,501 proteins across at least 80% of the matching protein’s length. These transcripts likely represent fully or near-fully reconstructed transcripts with detectable homologs in other species, of which over a third (9,950) are significantly enriched (22-fold; p value < 2.2 × 10^-16, Fisher’s exact test) among the E90 transcript set.

By subjecting our transcriptome to BUSCO analysis (Simão et al., 2015), we identified by this measure that our transcriptome
contains near-complete gene sequence information for 88% of the genes in the axolotl genome (Table 1). We finally applied the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007), which further indicated that our transcriptome is more than 98% complete (Table S1).

### Differential Expression Analysis Identified Tissue-Specific Gene Expression and Splicing

We estimated the expression profiles for all transcripts across all samples, and found excellent correlation between biological replicates (average Pearson correlation between replicates: 0.90; Figure 1D). As expected, samples derived from similar tissues (e.g., sections of the arm segments; bone and cartilage; proximal and distal blastemas) had more highly correlated expression than more distinct tissues such as the testes, which exhibited the most divergent transcriptional profile.

We next identified differentially expressed transcripts, conservatively focusing on transcripts identified as differentially expressed by multiple methods (Figure S1B) and restricted to samples with biological replicates (this excludes gill filament and embryos, which were used only for assembly). We identified 60,355 transcripts corresponding to 41,697 genes as at least 2-fold significantly differentially expressed (FDR ≤ 0.05) between at least two tissue-type comparisons (Data S4 and S5). Finally, we used a graph-based analysis to identify genes that are differentially expressed across each set of tissues (Figure 2A). In this approach, a graph is constructed for each differentially expressed transcript such that each tissue type is represented by a node, and an edge is drawn from a significantly upregulated tissue node to the downregulated tissue node. The graph is then partitioned to identify the maximal set of upregulated tissues for that transcript. Visualization of the graphical output for a given transcript intuitively presents the tissue(s) with the highest expression level for the transcript of interest (Figure 2A).

Of the differentially expressed transcripts, 14,594 transcripts (11,523 genes) were most tissue enriched given all pairwise tissue comparisons (Figure 2B; Table S2). Testes and ovaries had the most tissue-enriched transcripts (7,795 and 2,247, respectively, together 69%), followed by skeletal muscle tissue (961 transcripts) and the combination of all arm-segmented tissues (elbow, forearm, upper arm, and hand; 892) compared to each of the remaining tissues, including those tissues that compose the arm (e.g., bone). The tissue-specific transcripts were enriched for physiologically relevant functional categories based on enrichment analysis of Gene Ontology (GO) terms assigned from Swiss-Prot annotations of homologous proteins using Trinotate (e.g., spermatogenesis was enriched in testis-specific genes [FDR < 5.5e^{-4}], muscle contraction for skeletal muscle [FDR < 5.1e^{-5}], etc.; Table S2).

We experimentally validated by in situ hybridization and RT-PCR some of the tissue-enriched expression predictions to demonstrate our ability to correctly predict diverse tissue-specific gene expression patterns (Figures 2C–2J). For example, as predicted by our analysis, speriolin (spermatogenesis and centriole associated 1; spatc1) is enriched in germ cells within the testes (Figure 2C). This is consistent with limited studies on Speriolin that demonstrate this protein associates with centrosomes in spermatocytes and exhibits testis-specific expression in mice (Goto and Eddy, 2004; Goto et al., 2010).

Finally, we identified differentially expressed transcripts that are also associated with differential isoform usage across tissues. Of the 60,355 differentially expressed transcripts (41,697 genes), there are 29,750 transcripts (11,092 genes) with putative alternatively spliced isoforms. For 7,300 genes, the alternative isoforms have different patterns of tissue enrichment, suggestive of alternate functional roles (Table S3). Most of these genes (4,097 of 7,300) have at least one isoform significantly enriched in either testes or ovaries, with far fewer genes with evidence for tissue-specific isoforms in other tissues (e.g., 96 genes involving alternate isoform enrichment between heart and skeletal muscle). Notably, only 346 genes of the 11,523 genes with strong tissue-specific expression (Figure 2B) show evidence for tissue-specific isoform usage, the vast majority (315 genes) in either testes or ovaries. These data suggest axolotls may be less reliant upon alternative splicing in directing tissue-specific

### Table 1. BUSCO Analysis of Transcriptome Completeness

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>This manuscript</td>
<td>A. mexicanum</td>
<td>C:88%[D:53%],F:4.5%,M:7.3%,n:3023</td>
<td>C:42%[D:0.9%],F:17%,M:40%,n:3023</td>
<td>C:89%[D:42%],F:2.7%,M:7.8%,n:3023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiang et al., 2016</td>
<td>A. mexicanum</td>
<td>C:88%[D:53%],F:4.5%,M:7.3%,n:3023</td>
<td>C:42%[D:0.9%],F:17%,M:40%,n:3023</td>
<td>C:89%[D:42%],F:2.7%,M:7.8%,n:3023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakamura et al., 2014</td>
<td>C. pyrhogaster</td>
<td>C:82%[D:34%],F:4.6%,M:13%,n:3023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S1 contains near-complete gene sequence information for 88% of the genes in the axolotl genome. We finally applied the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007), which further indicated that our transcriptome is more than 98% complete.
Figure 2. Differential Gene Expression Analysis across Each Set of Tissues Identifies Transcripts Most Enriched in Specific Tissue Types

(A) Graph illustrating the methodology for the identification of genes that are tissue enriched in the context of all tissue pairwise comparisons using kazal-type serine peptidase inhibitor domain 1 (kazald1) as an example. Directed edges are drawn from upregulated to downregulated tissues, and fold changes in expression are indicated by edge colors.

(B) Heatmap showing all transcripts that are enriched in specific tissue types.

(C–J) RNA in situ hybridization performed on tissue sections.

(C) speriolin (spatc1) is enriched in the germ cells in the testes but is not detectable in adjacent support cells (asterisks).

(D) tropomyosin 1 (tpm1) is enriched in cardiomyocytes within the heart and not detectably expressed by other heart cell types such as epicardium (asterisks).

(E) ttn (ttn) is enriched in limb skeletal muscle but is not detectable in adjacent cartilage (cart) and epidermis (epiderm).

(F) kelch repeat and BTB domain-containing protein 10 (khl4l) is highly enriched in skeletal muscle and not detectable in adjacent tissues such as epidermis (epiderm) and fascia.

(legend continued on next page)
gene expression than has been documented in mammals (for example, Wang et al., 2008).

A Transcriptional Program for Blastemas Implicates the RNA Life Cycle in Limb Regeneration

We next analyzed the 159 transcripts (151 genes) differentially induced in all blastema samples (proximal and distal) as compared to all other sampled tissues (Figure 3A; Table S4). Notably, many of the blastema-specific genes (64/151) do not have a similar sequence identified in another organism in the UniRef90 database (E ≤ 10^-3; BLASTX). Nonetheless, several blastema-specific genes have known blastema relevance (e.g., twist family bhlh transcription factor 1 [twist1] [Satoh et al., 2008]) or are predicted to encode proteins in pathways implicated in salamander limb regeneration (retinol-binding protein 2a [rbp2a] [Maden, 1983]; matrix metallopeptidase 11 [mmp11] [Yang et al., 1999]). In particular, we identified the T box transcription factor tbx5, previously reported as enriched in forelimb blastemas (Khan et al., 2002), discriminating its tissue-specific expression among 13 total tbx genes detected in this axolotl transcriptome. Furthermore, when analyzing forelimb blastemas, our differential expression analysis did not erroneously uncover tbx4, which has been documented to only be upregulated in hindlimb blastemas (Khan et al., 2002).

Several blastema-specific genes are predicted to encode RNA-binding proteins, highlighting a potential role for the regulation of the RNA life cycle in regeneration (Figure 3A, highlighted). Among the blastema-induced transcripts are three predicted heterogeneous nuclear ribonucleoprotein (hnrbp) transcripts, an RNA-binding motif (rmb3) ortholog, a serine/arginine-rich RNA splicing factor (srsf1) ortholog, and the axolotl ortholog of cold-inducible RNA-binding protein (cirbp). Interestingly, a new ortholog of cirbp was recently identified in a subtractive hybridization screen as upregulated in limb blastemas compared to non-amputated limbs (Jiang et al., 2014), raising the possibility that cirbp and its homologs may be important regulators in regeneration across species. We also validated several of the blastema-enriched genes by in situ hybridization (Figure 3B), and confirmed they are not appreciably expressed in unamputated limbs (Figure 3C).

We further explored the expression of cirbp during limb regeneration and found that cirbp expression was apparent as early as 3 days post-amputation (Figure 4A), persisted as the blastema consolidates, and grows through palette formation (Figures 4B–4D). Past studies have demonstrated that CIRBP proteins regulate cell fate by inhibiting pro-apoptotic pathways during blastema-repressed transcripts is higher in skeletal muscle than in bone and cartilage (e.g., casq1), suggesting that blastema cells may be more actively repressing a muscle program than a skeletal program.

Functional Studies Indicate a Critical Role for kazald1 in Limb Regeneration

The most blastema-enriched transcript identified by our analyses was kazald1, whose transcription was >10-fold upregulated in blastemas versus all other tissues examined (Figure 5A). Kazd1 is predicted to have a Kazal-type serine protease inhibitor domain, a follistatin-like domain, and an immunoglobulin domain. We found that kazald1 expression is not detectable in the intact limb pre-amputation, increases shortly after amputation, is maintained during the blastemal stages of regeneration, and is dramatically downregulated near the end of regeneration (Figures 5B and 5C). Notably, kazald1 is not detectable in the developing limb bud within limb bud progenitor cells (Figure 5C). Therefore, its function in blastemas may be regeneration specific, in contrast to other genes whose expression is similar between limb buds and blastemas (for example, sonic hedgehog [Torok et al., 1999]; hoxd-8, -10, and -11 [Torok et al., 1998]).

We hypothesized that diminishing kazald1 expression during regeneration would have deleterious consequences and found that kazald1-targeting morpholinos delayed limb regeneration (Figures 5D–5F; Figure S4) and resulted in significantly smaller blastemas as compared to controls (Figure 5F: n = 48 limbs per condition; Figures S4B–S4D: n = 48 control MO2-treated limbs, n = 44 kazald1-MO2-treated limbs). Further, the kazald1-morpholino-treated animals had delayed chondrification and differentiation of the digits at the palette-to-early-digit stage, with less than 50% total Alcian blue-staining area relative to controls (Figures 5G–5K: n = 46 control MO1-treated limbs, n = 45
Figure 3. Identification and Validation of Blastema-Enriched Transcripts

(A) We identified 159 transcripts (151 genes) enriched in the blastema (the combination of proximal and distal blastema tissue) as compared to all other tissues. Those predicted to encode proteins with RNA-binding/regulation properties are highlighted in yellow.

(legend continued on next page)
kazald1-MO1-treated limbs; Figures S4E–S4I: n = 48 control MO2-treated limbs, n = 44 kazald1-MO2-treated limbs). These data strongly suggest that kazald1 is critical for the proper progression of limb regeneration.

Motivated by the finding that kazald1 is essential for normal regeneration, we leveraged our transcriptome to create kazald1-targeted deletions in both embryos and unamputated limbs via electroporation (Supplemental Information; Figure S5) to further complement our studies with morpholino knockdown. Although we were able to successfully edit the kazald1 gene, we did not achieve a sufficient level of editing efficiency to study the effect of genetic loss of function of kazald1 on limb regeneration with a high degree of confidence (17%–46% efficiency; Figure S5); we did not observe regenerative defects in mosaic animals or electroporated limbs, most likely due to low editing efficiency (data not shown). Nevertheless, future studies with complete genetic loss of function of kazald1 are warranted and will prove highly valuable to further examining this gene’s mechanistic role in axolotl limb regeneration.

kazald1’s intense expression within the transient blastema structure suggests that it has a time-delineated function during regeneration. To test this, we used a replication-incompetent retrovirus (Whited et al., 2013) to constitutively overexpress kazald1 in regenerating limbs (Figure S6A). A majority (15/22) of kazald1-infected limbs regenerated abnormally compared to control EGFP infections (0/18) (Figure S6B). The most commonly observed defect following kazald1 overexpression was syndactyly (8/15 abnormal limbs), and other defects included missing digits, truncated digits, severe clinodactyly, and ectopic tissue growth on the palm and digits (Figure S6C).

Our morpholino and retroviral experiments highlight a functional role for kazald1 during limb regeneration, as perturbing its expression during this process adversely affected the progression and outcome of limb regeneration. Another report also discovered enriched expression of kazald1 in a regenerative context in vitro (Athippozhy et al., 2014). However, in this study, the expression of the transcript was not profiled in intact or regenerating limbs, and the bulk of the expression was attributed to nerves. Here we show that the blastema cells themselves produce large amounts of kazald1 transcript. More studies will be needed to determine whether blastema cells are stimulated by the ingrowing regenerating nerve fibers to upregulate this transcript.

Comparison of Tissues at Different Anatomical Locations Identifies Candidate Transcripts of Positional Memory along Intact and Limb Regeneration Sites

Axolotl limb tissues maintain a “positional memory” throughout life allowing for the precise regeneration of only lost limb elements following amputation, suggesting that specific genes may be differentially expressed along the length of the limb and control positional memory. To uncover such genes, we assayed samples of similar tissue type but of diverse positional/anatomical context. First, we profiled cartilage from the wrist (carpals) and the ends of long bones (humerus, radius, and ulna). Second, we collected samples along the length of the limb, and blastemas arising from either proximal (mid-humerus) or distal (distal radius/ulna) amputations. We then analyzed their expression profiles to identify differentially expressed genes for each context.

Arm Segments

We performed two separate analyses to address the transcriptional differences between different locations along the proximal-distal axis of the arm (Figure 6). We first performed pairwise comparisons to identify transcripts significantly differentially expressed between any pair of arm segments. In the second approach, we performed analyses to uncover factors that show graded enrichment across the proximal-distal axis.

Using the pairwise approach, we uncovered a relatively modest number of transcriptional differences between different segments of the forelimb (“arm”): 636 differentially expressed transcripts identified between at least one pair of arm segments and 83 transcripts involving all arm segments (Figures 6A and 6B; Table S5), especially in the hand (48/83). Many of these likely reflect differences in the cell-type composition of the different arm segments. For example, transcripts enriched in the forearm, elbow, and upper arm are also more highly enriched in skeletal muscle and bone tissue; transcripts enriched in upper arm are also enriched in bone and cartilage (Figure S7A). Interestingly, 27 of the 48 transcripts enriched in the hand relative to other arm segments are also hand specific relative to other non-arm tissues. These include keratins and other filamentous proteins (6/27) with known structural roles in skin, consistent with the increased surface area of the hand versus the rest of the arm. Notably, when comparing highly similar tissue types, we found little signal of position-specific expression. In particular, comparing cartilage between wrist and long bone, we found only minor distinctions (Figure S7B), such that the samples were very highly correlated (r = 0.98), with only 22 differentially expressed transcripts, mostly in lowly expressed or non-cartilage-specific genes.

The arm segment analysis highlighted key transcripts with known function in limb polarity and candidates with uncharacterized functions in limb polarity (Figure 6B). First, we found hoxa13, which is known to mark distal elements in developing appendages (Haack and Gruss, 1993). Second, we found and validated a specific keratin, krt17, as exclusively expressed in the hand, restricted to the tubercles on the ventral/palmar surface of the hand (Figures 6B and 6C). Additional hand-enriched transcripts include a homolog of alpha tectorin (tecta), which is required for hearing in mammals (reviewed in Legan et al., 2014). Finally, one of the few transcripts that distinguishes upper-arm segments from other segments of intact limbs encodes a predicted ortholog of Chordin-like 1 (crdl1) (Figure 6B), a diffusible morphogen (reviewed in Zakin and De Robertis, 2010).

(B and C) In situ hybridization for six highly blastema-enriched transcripts at 23 days post-amputation (DPA (B) and on intact limbs (C). (B) Lower: higher magnifications of the boxed areas (upper). Yellow lines mark the WE/BL boundary. WE, wound epidermis; BL, blastema; mus, muscle; epi, epidermis; cart, cartilage; n, nerve. Scale bars, 500 μm.

See also Figure S3, Table S4, and Data S1, S2, S3, S4, and S5.
We uncovered 875 transcripts demonstrating a significant (FDR < 0.05) gradient in expression intensity along the proximal-distal axis of the unamputated limb (Figure 6D; Table S5). Among these, we uncovered two transcripts orthologous to emx2 (empty spiracles homeobox 2) that were most highly expressed in the hand and had diminishing expression toward the proximal sample sites. The newt ortholog of emx2 has previously been shown to exhibit distal-enriched expression in unamputated limbs (Beauchemin et al., 1998). Conversely, we identified apcdd1l (orthologous to Apcdd1-like) as the most significantly graded transcript, with high levels in the proximal segments. Apcdd1 is an extracellular membrane-bound signaling protein that can inhibit Wnt signaling (Shimomura et al., 2010).

**Proximal versus Distal Blastema**

There are 305 transcripts (275 genes) differentially expressed between proximal and distal blastema tissue (Data S4 and S5), of which 81 transcripts (70 genes; Table S5) are also enriched in blastema tissue relative to other tissues (Figures 6E and 6F). We validated cd38 and shox, whose expression is significantly enriched in proximal blastemas (Figure 6G). Cd38 is predicted to encode the axolotl ortholog of mammalian CD38, a cell-surface glycoprotein and ADP-riboseyl cyclase (reviewed in Wei et al., 2014). Cd38 expression at both the mRNA and protein levels can be induced by exposure of cells to retinoic acid (Drach et al., 1993). Because retinoic acid is a proximalizing agent in limb regeneration (Maden, 1983), enrichment of cd38 transcripts in proximal blastemas may reflect activation of retinoic acid signaling. Shox (short-stature homeobox) encodes a predicted homeobox-containing transcription factor. Experimentally removing the most closely related mouse gene, Shox2, in limb buds results in the near-complete loss of the humerus, a proximal skeletal element, whereas remaining structures are mostly unaffected (Cobb et al., 2006). Future studies may reveal a functional role for these genes in controlling proximal-distal identity during limb regeneration.

**DISCUSSION**

Here we report the assembly and validation of an axolotl transcriptome centered on tissues relevant for limb regeneration that provides near-complete sequence information for exonic regions of ~88% of axolotl genes with a high degree of confidence (Table 1), generating a valuable annotated resource for researchers using axolotl. We used differential expression analysis within the context of a biologically informed design to identify tissue-specific and anatomically specific gene expression patterns that we were able to validate experimentally, demonstrating the validity and the breadth of experimental design that can be conducted with our catalog.

Within the blastema, the nexus of limb regeneration, we uncovered an enriched profile of transcripts for members of the heterogeneous nuclear ribonucleoprotein complex (hnRNP) such as cirbp, fus, roa1, and hnrnpd. In other biological contexts, RNA-binding proteins are capable of modulating large repertoires of mRNA targets, coordinating cellular responses, and maintaining cellular phenotypes (Kafasla et al., 2014; Zhang et al., 2013). During limb regeneration, modulation of
Figure 5. *kazald1*, the Most Robust Blastema Marker, Is Required for Limb Regeneration

(A) Differential tissue expression analysis identifies *kazald1* as the most blastema-enriched transcript compared to all other tissues sequenced. (B) RT-PCR performed on blastema cDNA samples throughout the course of regeneration for *kazald1* expression. *kazald1* was not detected in intact limbs and at 1 DPA, and has dramatically diminished by 60 DPA. (C) In situ hybridization for *kazald1* in the blastema over the course of regeneration (top panels). *Kazald1* is not detectable in regenerated limbs (35 DPA), intact limbs, or developing limb buds (lower panels). (D–F) Regenerating limbs at 19 DPA treated with control (D) or *kazald1*-targeting morpholino (E); quantified in (F). (G–J) Regenerating limbs at 28 DPA treated with control (G) or *kazald1*-targeting morpholino (H and H'). (I–K) Same specimens stained with Alcian blue to visualize cartilage; (I) is the skeletal preparation of the limb shown in (G); (J) is the skeletal preparation of the limb shown in (H); (J') is the skeletal preparation of the limb shown in (H'). Not that (G) and (I) are a control, while (H) and (J) and (H') and (J') are specimens treated with *kazald1*-targeting morpholino. Results are quantified in (K).

Scale bars, 500 μm (C) and 1 mm (D, E, and H–J'). ***p < 0.001. Error bars are SEM. Arrowheads mark the amputation plane in each image. See also Figures S4–S6 and Data S1, S2, S3, S4, and S5.
Figure 6. Transcripts Differentially Expressed in Proximal versus Distal Elements
(A) Schematic illustrating specific elements of the hand and arm.
(B) Differential gene expression analysis identifies transcripts that are enriched in distinct sections of the intact limb.
(C) RT-PCR validation of select transcripts identified by differential expression analysis.
(D) Gradient gene expression analysis identifies transcripts enriched in a gradient from proximal to distal or distal to proximal.
(E) Schematic illustrating amputation planes for sampling of proximal and distal blastemas.
(F) Differential expression analysis identifies transcripts that are enriched in proximal versus distal blastemas.
(G) In situ and RT-PCR validation of computational predictions of differentially expressed transcripts in proximal and distal blastemas.
See also Figure S7, Table S5, and Data S1, S2, S3, S4, and S5.
RNA-binding proteins and post-translational regulation of mRNA may constitute major mechanisms for executing the vast cellular changes necessary to regenerate a complex structure. Specifically, we highlighted the axolotl ortholog of cirbp as highly blastema enriched and uncovered a cytoprotective function for this gene during early blastema development. cirbp encodes a putative RNA-binding protein orthologous to mammalian cold-inducible RNA-binding protein (CIRBP) and highly similar to RBM3 (orthologous axolotl rbmx was also identified as blastema enriched). Both cirbp and rbm3 have been demonstrated to be transcriptionally upregulated in response to specific cellular stressors including cold shock, hypoxia, and UV irradiation in mammalian cells (Danno et al., 1997; Nishiyama et al., 1997; Wellmann et al., 2004). Our experiments suggest that one function of cirbp in regeneration is to protect progenitor cells from cell death, possibly to protect blastema progenitors from extensive cellular changes occurring following amputation. We also identified cirbp transcripts at the early-to-late-dig stage (35 days post-amputation) within the interdigital regions, which are presumed to be relatively undifferentiated and will be eliminated as the regenerating digits are sculpted (reviewed in Montero and Hurlé, 2010). Further, fus mRNA has been identified as a major target of mammalian CIRBP (Morf et al., 2012), both of which are significantly enriched in the blastema. Thus, understanding regulatory loops between RNA-binding proteins to modulate cell-fate pathways within blastema cells may be emerging as a theme in limb regeneration, supporting a need for future studies.

We also experimentally tested the consequence of manipulating expression of kazald1, the gene whose transcript best distinguishes blastema cells from all other samples sequenced in our data. We found that reducing expression of kazald1 via morpholinos resulted in an outward lag in the progression of limb regeneration compared to controls whereas constitutive expression of kazald1 within the blastema resulted in regeneration of deformed limbs, underscoring the importance of temporal and/or spatial regulation of this transcript. Interestingly, a similar gene has been implicated in regeneration in hydra, an invertebrate with extensive regenerative abilities (Chera et al., 2006), raising the possibility that an ancient and conserved role for kazald1 in regeneration exists. Other Kazal-type serine protease inhibitors in mammals serve as biomarkers for various forms of cancer, but their exact functions are only now emerging (reviewed in Fradet, 2009). Because aspects of limb regeneration share some similarity to cancer (for instance, activation of stem cells and dedifferentiation, cell proliferation, reactivation of developmental pathways, and the wound response) (reviewed in Pomerantz and Blau, 2013), links between the two processes warrant further investigation. Future work will be necessary to determine whether connections exist between the roles of Kazal-type inhibitors in other organisms and the role of kazald1 in axolotl limb regeneration.

Our assembly and experimentation will serve as a valuable resource for researchers using axolotls in a variety of fields, as well as for those studying regeneration in other models. Our differential gene expression analysis can provide starting points for studying genes that may drive the biological function of particular tissues. It can also be used to identify reliable markers of particular tissue types and, with further experimentation, possibly specific cell types within particular tissues. In many cases, our assembly will provide the full-length coding sequence for future functional studies, including the requisite sequence information to design genome-editing strategies and provide data to minimize potential off-targets. The axolotl transcriptome we report can serve as a basis for mining tissue-enriched or -repressed transcripts, and will also serve as a powerful resource for more directed future studies. All reconstructed axolotl transcripts, expression data, and feature annotations are available via our web portal at http://portals.broadinstitute.org/axolotomics.

EXPERIMENTAL PROCEDURES

Trinity De Novo RNA-Seq Assembly

Trinity version trinityrnaseq_r2013-02-25 was used for de novo transcriptome assembly and analysis as follows. The combined set of 1.3 billion RNA-seq reads spanning all sampled tissues was combined into a single pair of fastq files, quality trimmed using Trimomatic (Bolger et al., 2014) (parameters LEADING:5 TRAILING:5 MINLEN:38), and subsequently normalized using the in silico normalization step incorporated into Trinity (TRINITY_HOME/util/normalize_by_kmer_coverage.pl–seqType fq–JM 100G–left ALL_AXOLOTL_READS.left.fq–right ALL_AXOLOTL_READS.right.fq–ptrim.fq–pairs_together–JELLY_CPU 10–PARALLEL_STATS–max_cov 50). The resulting normalized reads were then assembled using Trinity (Trinity: pi-left left.fq–right right.fq–seqType fq–JM 100G–CPU 10).

Transcript Annotation Using Trinotate

We developed Trinotate as a bioinformatics protocol and software system for annotating putative functional characteristics of transcripts. Similarities to known proteins were detected by a BLASTX search (Camacho et al., 2009) (E ≤ 1e-5) of a comprehensive protein database formed by merging Swiss-Prot (Boeckmann et al., 2005) with UniRef90 (UniProt Consortium, 2019) protein databases downloaded from UniProt (versions available on July 5, 2014). Likely coding regions within transcripts were predicted using TransDecoder (http://transdecoder.github.io), and resulting protein products were searched for sequence similarities against the comprehensive protein database described above (BLASTP: E ≤ 1e-5) and for conserved protein domains using Hmmer (http://hmmer.org/) and PFam (Finn et al., 2014). Signal peptides were predicted using SignalP (Petersen et al., 2011), and transmembrane region predictions were predicted using TMHMM (Krogh et al., 2001). All results were parsed by Trinotate, stored in an SQLite relational database, and then reported as a tab-delimited transcript annotation summary file (Data S3). Gene Ontology identifiers were transfinitively assigned to transcripts based on available GO annotations of best-matching Swiss-Prot entries (Data S4 and S5).

Transcript Abundance Estimation and Differential Expression Analysis

RSEM (Li and Dewey, 2011) and MMSEQ (Turro et al., 2011) software were used to estimate expression values for transcripts, and edgeR (Robinson et al., 2010), EBSeq (Leng et al., 2013), and mmdiff (Turro et al., 2014) were used to identify significantly differentially expressed transcripts. Further details can be found in Supplemental Experimental Procedures.

Animal Experimentation

All animal experiments were performed in accordance with Harvard Medical School’s Institutional Animal Care and Use Committee regulations and in accordance with Animal Experimentation Protocol 04160. For all survival surgeries, animals were anesthetized in 0.1% tricaine and allowed to recover overnight in 0.5% sulfamethazine.

Library Preparation and Sequencing

Total RNA was purified using either Trizol or QIAGEN RNeasy columns. The illumina TruSeq v2 protocol was used throughout to generate bar-coded sequencing libraries. Paired-end, 100-bp sequencing was performed on the illumina HiSeq 2500 at Harvard Medical School’s Biopolymers Facility.
In Situ Hybridization
Sequences were amplified from cDNA and cloned into the pGEM-T Easy vector and sequenced. Depending upon orientation, T7 or Sp6 polymerase was used to transcribe the probe. In situ hybridization was performed as in Whitel et al. (2011).

Statistical Analyses
All results from morpholino knockdown and retroviral experiments were quantified by at least one blinded observer. Blasticoma length and chondrification area were measured using ImageJ (NIH). A two-tailed Student’s t test was used to assess statistical significance. The statistical significance of the phenotypic outcomes from retroviral overexpression of kazalid1 was assessed by performing a two-tailed Fisher’s exact test on a 2 × 2 contingency table.

For detailed methods, see Supplemental Experimental Procedures.

ACCESSION NUMBERS
The accession number for the raw sequencing reads reported in this paper is SRA: PRJNA300706 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA300706). The accession number for the transcriptome data reported in this paper is Transcriptome Shotgun Assembly Database: GFBM00000000. The accession number for all data reported in this paper is GEO: GSE92429.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, seven figures, five tables, and five data files and can be found with this article online at dx.doi.org/10.1016/j.celrep.2016.12.063.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS
This work was supported by the Harvard Stem Cell Institute (C.J.T.), Brigham & Women’s Hospital (J.L.W.), Richard and Susan Smith Family Foundation (J.L.W.), NIH/NICHD (R01 HD073104, to L.P.), and Howard Hughes Medical Institute (A.R.). This research was supported by the National Cancer Institute of the NIH under award 1U24CA180922-01 (to B.J.H. and A.R.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. D.M.B. was supported by an HHMI Gilliam Fellowship. We thank Borja Sese Ballesteros and Yick Fong for their helpful discussions on using CRISPR-Cas9; Jana Hersh for germ cell identification; Josh Gorham and Jim Pancoast for assistance in RNA extraction; Harvard Biopolymers for sequencing; Esther Pearl and Marko Horb for sequencing the embryo samples; Rich Lee for use of shared equipment; Leslie Gaffney for help with figures; and Benjamin Lewis for discussions. A subset of computations in this paper was run on the Orchestra cluster supported by the Research Information Technology Group at Harvard Medical School. We thank the Ambystoma Genetic Stock Center for providing some of the animals (Lexington, KY; NIH grant P40-OD019794).

Received: May 25, 2016
Revised: October 26, 2016
Accepted: December 20, 2016
Published: January 17, 2017

REFERENCES


