Reading and editing the Pleurodeles waltl genome reveals novel features of tetrapod regeneration

Citation
Elewa, A., H. Wang, C. Talavera-López, A. Joven, G. Brito, A. Kumar, L. S. Hameed, et al. 2017. “Reading and editing the Pleurodeles waltl genome reveals novel features of tetrapod regeneration.” Nature Communications 8 (1): 2286. doi:10.1038/s41467-017-01964-9. http://dx.doi.org/10.1038/s41467-017-01964-9.

Published Version
doi:10.1038/s41467-017-01964-9

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:34652068

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Reading and editing the Pleurodeles waltl genome reveals novel features of tetrapod regeneration

Ahmed Elewa1, Heng Wang2, Carlos Talavera-López1,7, Alberto Joven1, Gonçalo Brito1, Anoop Kumar1, L. Shahul Hameed1, May Penrad-Mobayed3, Zeyu Yao1, Neda Zamani4, Yamen Abbas5, Ilgar Abdullayev1,6, Rickard Sandberg1,6, Manfred Grabherr4, Björn Andersson1 & András Simon1

Salamanders exhibit an extraordinary ability among vertebrates to regenerate complex body parts. However, scarce genomic resources have limited our understanding of regeneration in adult salamanders. Here, we present the ~20 Gb genome and transcriptome of the Iberian ribbed newt Pleurodeles waltl, a tractable species suitable for laboratory research. We find that embryonic stem cell-specific miRNAs mir-93b and mir-427/430/302, as well as Harbinger DNA transposons carrying the Myb-like proto-oncogene have expanded dramatically in the Pleurodeles waltl genome and are co-expressed during limb regeneration. Moreover, we find that a family of salamander methyltransferases is expressed specifically in adult appendages. Using CRISPR/Cas9 technology to perturb transcription factors, we demonstrate that, unlike the axolotl, Pax3 is present and necessary for development and that contrary to mammals, muscle regeneration is normal without functional Pax7 gene. Our data provide a foundation for comparative genomic studies that generate models for the uneven distribution of regenerative capacities among vertebrates.

1Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, SE-171 65, Sweden. 2College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China. 3Institut Jacques Monod, CNRS & University Paris-Diderot, Paris, 75205, France. 4Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, SE-751 23, Sweden. 5Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA. 6Ludwig Institute for Cancer Research, Stockholm, SE-171 65, Sweden. 7Present address: The Francis Crick Institute, NW1 1AT London, UK. Ahmed Elewa, Heng Wang and Carlos Talavera-López contributed equally to this work. Correspondence and requests for materials should be addressed to A.E. (email: ahmed.elewa@ki.se) or to A.S. (email: andras.simon@ki.se)
The random manifestation of extensive regeneration capacities in the animal kingdom implies a phylogenetically widespread regeneration potential, which is masked in most species\(^1\). Among tetrapods, salamanders, such as newts and axolotls, display the largest regenerative repertoire. A newt can rebuild entire limbs, tails, jaws, cardiac muscle, ocular tissues, and restore central nervous system tissues including brain structures\(^6\) (Fig. 1). However, it is important to note that major differences exist even among salamanders. In contrast to the paedomorphic axolotls, newts undergo metamorphosis, have a broader regeneration spectrum and mobilize additional cell sources for regeneration of the same body part\(^7\). Such interspecies differences among closely related species offer opportunities to reveal valuable information about the evolution of processes that allow or counteract regeneration. Although significant progress has been made to characterize salamander transcriptomes and proteomes\(^8\)\(^–\)\(^12\), features such as species-specific genes, expansion or contraction of gene families, and the underlying cause of their gigantic genome size remain largely unexplored. In addition, due to their complex and long life cycle, most newt species are cumbersome to breed under laboratory conditions, which has hampered the establishment of genetically modified lines. However, the Iberian ribbed newt *Pleurodeles waltl* (*P. waltl*) is easily bred in laboratories and retains the widest known spectrum of regeneration abilities among adult vertebrates\(^13\) (Fig. 1; Supplementary Fig. 1a). Here we describe the genome and transcriptome of *P. waltl* (Methods; Supplementary Methods, Supplementary Fig. 1b, and Supplementary Tables 1–5) as a resource to explore regeneration relevant novelties, and adapt CRISPR/Cas9 technology to perturb key transcription factors involved in regeneration.

**Results**

**Sequencing the genome and transcriptome of *P. waltl*.** The diploid genome of *P. waltl* is organized in 12 chromosome pairs, which have been the subject of classical lambrush chromosome studies\(^14\) (Supplementary Fig. 1c–h). The *P. waltl* haplotype genome size is \(~20\) Gb (Supplementary Table 1), making this one of the largest vertebrate genomes sequenced to date. Our genome annotation pipeline identified 14,805 complete protein-coding gene models and we estimate that this set represents 64.8% of the *P. waltl* protein-coding repertoire (Supplementary Methods; Supplementary Table 6). To provide a platform for comparative genomic studies, we identified 19,903 orthology groups involving *P. waltl* protein-coding genes and/or transcripts (Supplementary Methods; Supplementary Fig. 2a, Supplementary Table 1, Supplementary Data 1). Of these orthology groups, 1575 consisted of salamander members only (salamander groups) and 1130 consisted of salamander and *Xenopus* orthologs only (amphibian groups). The remaining 17,198 groups consisted of salamander and other vertebrate orthologs (human, mouse, chicken, lizard, or zebrafish). Importantly, we did not observe any expansion or loss of non-transposable protein-coding orthologs compared to other vertebrates (more than twofold increase or decrease) (Supplementary Methods, Supplementary Data 1–3).

An expansion of Harbinger elements in the *P. waltl* genome. A striking feature of the *P. waltl* genome is the extent and diversity of its repetitive elements. The genome is host to a diverse population of class I and class II transposable elements (Supplementary Table 1). We assembled a repeat library by majority vote k-mer extension\(^15\), followed by alignment to the repeat database RepBase\(^16\) using Satsuma\(^17\), yielding 428 distinct sequences (Methods). Gypsy retrotransposons are the most...
frequent repetitive elements in *P. walli*, followed by the Harbinger transposons, and together account for about two thirds of the genome repetitive content (Supplementary Table 1). A phylogeny of ~1200 Gypsy elements longer than one kilobase indicates continuous expansion of this family (Fig. 2a, b), while Harbinger elements have undergone two distinct evolutionary bursts, with one recent expansion, visible from the distribution of pairwise similarity (Fig. 2a, b; Supplementary Fig. 2b). Harbinger elements are distinct from other transposons in that they carry a Myb-like gene, a proto-oncogene that acts as a transcription factor. While Harbinger elements gave rise to the genes *Harb1* and *Nafl* in the vertebrate lineage, their contribution to vertebrate genome content is extremely rare with the leading example being the genome of coelacanth *Latimeria chalumnae* (~1 to 4% of the genome) [19]. Therefore, the Harbinger element expansion we describe in *P. walli* is hitherto unprecedented.
Expansion of embryonic stem cell-specific miRNAs in *P. waltl*. MicroRNAs (miRNAs) are short non-coding RNAs that regulate post-transcriptional gene expression. A number of miRNAs, including miR-302, are capable of reprogramming mammalian somatic cells into an induced pluripotent state. We identified 361 miRNA precursors in the *P. waltl* genome that produce 67 distinct mature miRNAs conserved in *Xenopus*. In addition, the de novo transcriptome included 202 transcripts bearing conserved miRNA precursors (primary miRNA transcripts) that produce 55 of the 67 genome predicted mature miRNAs (Supplementary Methods, Supplementary Fig. 3, Supplementary Tables 7–8, Supplementary Data 4). Surprisingly, most of the genome predicted precursors were copies of mir-93b and mir-427 (known as mir-430 in zebrafish and mir-302 in mammals) (66 and 155 precursors, respectively) (Fig. 2c). Both mir-93b and mir-427 bear the characteristic embryonic stem cell-specific cell cycle regulating (ESCC) seed (AAGUGC) and are of interest to regeneration studies (Fig. 2c). The expansion of mir-427 has been reported in *Xenopus* (~80 copies) and zebrafish (~100 copies) and thus predates the Devonian period. The mir-93b expansion, however, appears to be salamander-specific and we place its occurrence in the Jurassic period or later.

Harbinger and ESCC miRNA expression during limb regeneration. Given the compelling expansion of ESCC miRNAs and Harbinger transposable elements, we analyzed whether mir-93b, mir-427, Myb, and Harbidi domain containing genes are regulated during adult *P. waltl* limb regeneration (Supplementary Table 9). We mapped total RNAseq reads from three early developmental stages, nine adult (>2 years old) body parts and two limb regeneration stages; 3 and 7 days post amputation (Supplementary Table 3) to 14,805 gene models and 108,713 transcripts belonging to the *P. waltl* orthology groups in addition to 202 miRNA primary transcripts (Supplementary Methods, Supplementary Data 1,4). We found that mir-93b and mir-427 primary transcripts are upregulated in the regenerating limb 3 days post amputation (p < 0.001) and that their mature miRNAs are detected during limb regeneration (21 and 45 reads per million, respectively) (Fig. 2c; Supplementary Fig. 4, Supplementary Table 9 and Supplementary Data 4). In addition, we found that five orthology groups containing Myb or Harbidi domains were upregulated at 3 days post amputation (p < 0.001) (Supplementary Table 9) similar to mir-93b and mir-427 primary transcripts (Fig. 2e). On the contrary, the two vertebrate Harbinger derivative genes *Nafl1* and *Harb1l* were expressed at low levels in all our datasets and showed no regulation during regeneration (Supplementary Table 10). Four Gypsy orthology groups were also upregulated 3 days post amputation (p < 0.01), however, their expression levels were an order of magnitude lower than the Harbingers (Supplementary Table 9). Transposable elements are often domesticated to benefit their hosts and RNA transposable element expression has been described in mammalian pluripotent stem cells. Our data illustrate that the both RNA and DNA transposons, Gypsy and Harbinger, respectively, in addition to ESCC miRNAs respond to adult salamander injury. The extent to which they co-regulate regeneration awaits further studies.

Restricted expression of salamander methyltransferases. Genes found only in salamanders may offer insight into their unique regenerative abilities. Our annotation pipeline identified 1545 orthology groups that consisted of genes and transcripts in the *P. waltl* genome and at least one other salamander, but no sequences from the other seven model vertebrates analyzed. We consider these orthology groups to be putatively salamander-specific and refer to them by the protein domain(s) detected in their members. We calculated the tau score (an indicator of tissue specificity) for all orthology groups and found that groups with only salamander genes are more tissue specific than groups with orthologs in other vertebrates (Fig. 2d and Methods). Surprisingly, we found eleven salamander orthology groups encoding a NNNMT/PNMT/TEM1 methyltransferase domain and expressed specifically in adult limbs and tail, four of which were significantly downregulated on the seventh day of regeneration (*p* < 0.001; Fig. 2f; Supplementary Fig. 2b and Supplementary Table 9). No other salamander orthology groups showed this expression pattern. Together, our data show that *P. waltl* has evolved genes with tissue-restricted expression profiles and that target methylation may have evolved in a manner consistent with limb regeneration.

**Pax3 and Pax7 are present and functional in *P. waltl***. We were intrigued by the absence of Pax3 in the axolotl (*Ambystoma mexicanum*) transcriptome, genome and previous gene expression studies. Pax3 is a paired-end homeodomain transcription factor required for early development and its paralog Pax7 regulates skeletal muscle regeneration. Manual curation of the Pax gene family and in situ hybridization confirmed the presence of Pax3 and Pax7 in *P. waltl* (Supplementary Fig. 5a–c). We adapted CRISPR/Cas9 technology to *P. waltl* and mutated both genes (Methods; Supplementary Fig. 5a,d and Supplementary Table 11). Whereas Pax7/−/− compound heterozygote F1 animals exhibited no apparent deficiencies in muscle development, Pax3 mosaic-mutants died or developed animals with several developmental anomalies, exemplified by muscle-less limbs (Fig. 3a, b; Supplementary Fig. 5e–f and Supplementary Video 1). Next we asked how these mutations affect limb regeneration. Pax7/−/− F1 animals regenerated limbs indistinguishable from wildtype controls, including skeletal muscle and cells in satellite cell position (Fig. 3c, d and Supplementary Fig. 6; Methods). Pax3 mosaic F0 mutants lacking limb skeletal...
muscle regenerated morphologically normal muscle-less limbs without any other overt phenotypes (Fig. 3e).

**Discussion**

Salamander genomes are gigantic, ranging in size between 14 and 120 Gb. While this feature has facilitated genome structure studies using lampbrush chromosomes, it has delayed sequencing projects and impeded regeneration studies. Lampbrush chromosomes epitomize genome structural plasticity, while salamander regeneration is a prime example of cellular plasticity. Whether these two forms of plasticity are mechanistically related is now a pursuable question.

Reading and editing the *P. waltl* genome reveals several new features of limb regeneration and genome evolution. Our loss of function data on Pax3 and Pax7 paralogues show that limb regeneration is not dependent on skeletal muscle and that no other tissues give rise to muscle during limb regeneration in *P. waltl*. Furthermore, the expression and developmental functions of Pax3 are conserved between *P. waltl* and mammals, which also raises the question whether the lack of Pax3 in the axolotl confers additional functions for Pax7. Mammalian skeletal muscle regeneration is impaired in the absence of Pax7 but we have found that Pax7 loss of function does not impair muscle regeneration in *P. waltl*. In this context, it is important to note that a distinctive process during newt limb regeneration is the reversal of terminal differentiation of skeletal muscle fibers. Myogenic dedifferentiation during limb regeneration produces progenitor cells that build up the new muscle and thus has the potential to compensate for the potential loss of contribution from satellite cells. However cell tracking studies of satellite cells and their progeny are necessary to resolve this question.

The genomic expansion and expression of ESCC miRNAs and Myb domain-bearing Harbingers, along with limb-specific expression of putatively salamander-specific methyltransferases present new possibilities to uncover mechanisms of adult...
vertebrate regeneration using *P. waltl* as a model organism. For example, the multitude of mir-427 copies in fish and amphibian genomes in addition to that of mir-93b in newts may provide an opportunity for individual copies to evolve and acquire expression rights and co-regulate a regenerative response after injury. In *X. laevis* and zebrafish, miR-427/430 helps clear of cache of maternally supplied RNA when embryos transition to rely on their own gene expression.11, 13. One intriguing possibility is that miR-427 performs a similar function in *P. waltl* during limb regeneration. In this model, miR-427 clears the cytoplasmic cache of mRNAs, thereby erasing a cell’s “working memory” and allowing de differentiation to occur, while the “deep memory” of the cell’s identity and function would not be lost by a permanent inactivation of gene expression.

**Methods**

**Genomic DNA isolation.** High molecular weight genomic DNA was purified from a single larva after discarding the digestive system. The larva represents the first generation of our *P. waltl* colony, established from fertilized eggs produced in a laboratory colony located in Madrid, Spain. The larva is therefore the fourth/fifth generation of laboratory-bred *P. waltl*, originally obtained with permission from a wild population in Doñana National Park (Spain) for research purposes by Agustin Gonzalez. The body was segmented into five pieces and each piece was placed in 1.5 mL centrifuge tube. A total of 700 µL digestion buffer (50 mM Tris-HCl pH 8.0, 100 mM EDTA, 1% SDS) was added to each sample. 50 µL wild population in Doñana National Park (Spain) for research purposes by Agustin Gonzalez.

Methods inactivation of gene expression.

**strategies**

**Phylogeny of Gypsy and Harbinger elements.** We computed pairwise genomic distances with Satsuma (https://sourceforge.net/projects/satsuma/ last update 2016-12-14) and selecting the longest alignments respectively if multiple overlapping hits were reported.

**transcriptome RNA preparation and sequencing.** 25 indexed Illumina Truseq Stranded total RNA Zero-Gold libraries (mean insert size 150 bp) from 20 *P. waltl* body parts and regeneration stages were combined into four pools and sequenced on 4 lanes (one pool per lane) on Illumina HiSeq Hight Output mode v4, PE 2 × 125 bp at the Science for Life Laboratory, Stockholm (Supplementary Table 3). In all cases, RNA was isolated using Trizol after grinding the sample with a mortar and pestle in liquid nitrogen and precipitated overnight in isopropanol. RNA integrity was assessed using Agilent Bioanalyzer. Samples with RIN ≥ 8.0 were used for library preparation.

**Genome annotation generation of gene models with MAKER.** MAKER was used in a two-step approach to identify gene models in the *P. waltl* genome. MAKER Tier 1: The clustered assembled transcripts, the Augustus hinted gene models and the protein sequences from uniref90 were assembled into contigs using Abyss (28) using a k = 71. Illumina paired-end RNAseq reads from representative body sections of *P. waltl* were mapped to *Trans-Abyss* (29) (L = 20) with the aim to reconstruct only longer, complete transcripts rather than isoforms or short RNAs. Trans-Abyss transcripts were then used to scaffold the genomic Abyss-contigs in a third assembly aimed at improving the reconstruction of gene content. The final assembly—denominated *Pw_v2.0*—was 19.66 Gb in size and included ~66 million contigs of ~109,895 scaffolds (Table 1). Approximately 45% of the genome assembly was captured in contigs ranging between 1 kb and 10 kb in size and ~6% (1.2 Gb) were assembled in contigs and scaffolds longer than 10 kb, the longest being 438.8 kb in size (Supplementary Fig. 1b).

**Repeat library assembly.** We assembled a repeat library by majority vote k-mer (k = 25) extension algorithm15 to the *P. waltl* genome assembly *Pw_v2.0*, resulting in a set of 428 consensus repeat sequences larger than 1000 nt. Out of these, 349 were identified as known repeats through nucleotide alignments against RepBase (version 21.10 downloaded 10/23/2016) using Satsuma (https://sourceforge.net/projects/satsuma/ last update 2016-12-14) and selecting the longest alignments respectively if multiple overlapping hits were reported.

**Lampbrush chromosome analyses.** Ovarian biopsies were performed on adult females of *P. waltl* that were anesthetized in 0.15% MS222 (Amino-benzoic Acid Ethyl, Fluka). Stage IV-VI oocytes were selected and maintained in MBR buffer (Modified Barth’s solution) at 18°C. Germinal vesicles were manually isolated from the oocytes and dissected in 75 mmol L⁻¹ KCl, 25 mmol L⁻¹ NaCl, 0.01 mmol L⁻¹ MgCl₂ and 0.01 mmol L⁻¹ CaCl₂, pH 7.2, and LBCs were prepared as previously described1. Nuclear spread preparations were centrifuged at 300 g for 10 min, and at 3000 × g for 30 min at 4°C. Standard transmitted light were performed at the ImagoSeine core facility (member of the IBISA and the France-Biolimaging (ANR-10-INBS-04) infrastructure) using a wide field Leica microscope with phase contrast plan Apo objectives: ×40 (NA = 1.3), ×63, ×100 (NA = 1.4). Images were captured using a CoolSnap HQ. Photometrics camera driven by the software MetaMorph (Universal imaging). Series of 11 confocal Z-planes (0.7 μm distance) were collected. Pixel size of the images were 0.1625 µm, 0.103 µm, and 0.065 µm for the ×40, ×63, and ×100 objectives, respectively.

**Karyotyping.** Cells were arrested in metaphase using classical procedures, standardized for *P. waltl*:12 stage 20 embryos (n = 6) were placed in saline solution (30 mM NaCl, 0.34 mM KCl, 0.9 mM CaCl₂ containing 0.5% colchicine. At 18°C, after 15 min, the eggs were fixed with a mixture of 50% acetic acid/ 50% EtOH for 3–5 min, washed in distilled water and transferred to 50% acetic acid, were they were dissociated mechanically with a Pasteur glass pipette. Cells were then dropped on Superfrost slides, let dry, and stained with DAPI for confocal analysis. For Supplementary Figure 1, a Z-stack confocal projection was made with an interval of 0.57 µm between planes, and the twelve pairs of chromosomes were identified and distributed in three groups according to ref.50. For an easy interpretation of the results, colors were given applying color filters to individually selected chromosomes using Adobe Photoshop. The full extent of the chromosomes in the Z-projection was then selected, filled with color, and organized to provide the karyotype.
The PCR products were cloned into TOPO cloning vector (Invitrogen). Individual clones were sequenced with T7 primer (Supplementary Table 9). Single-cell fertilized eggs were obtained by either natural or induced breeding and injected according to previously published protocols with modifications60, 61. Briefly, 500 pg Ca9 RNA and 100 pg gRNA were mixed into 5 nL and injected into freshly laid single-cell-stage embryos. The animals were raised according to60. The screening of F0 animals was done by both genotyping (see above) and phenotype characterization. The apoptosis analysis in immunohistochemistry of limbs/tails, according to60, 61. The Pax5−/− and Tnr−/− F1 animals were produced by crossing between adult F1 animals. The larvae or post-metamorphic news were anaesthetized with 0.01% or 0.1% ethyl-p-aminobenzoate (benzocaine; Sigma) prior to imaging and tissue collection. News utilized for this study were processed according to Swedish Board of Agriculture animal ethical regulations (NIV2912).

In situ hybridization and immunohistochemistry. The PCR fragments of Pax3 and Pax7 were cloned into TOPO cloning vectors. The digoxigenin-labeled anti-sense RNA probes were synthesized using T7 RNA polymerase (Roche). Whole-mount in situ hybridizations were performed by using alkaline Tnr−/− F1 embryos according to60. In situ hybridizations on 10 µm transverse sections were carried out by fixing tissues in 4% formaldehyde for 15 min, treating with 0.2 N HCl for 12 min, washing, and then incubating in acetylation buffer (0.1 M triethanolamine and 0.25% acetic anhydride) for 10 min. Next, the slides were rinsed in RNase free water and permeabilized with a solution of 1 µL mL−1 Proteinase K (Roche) and 2 µM CaCl2 for 15 min at 37°C. The sections were incubated in prehybridization buffer (50% deionized form of amide, 5 × saline sodium citrate (SSC), 5 × Denhardt’s solution, 250 µg mL−1 yeast RNA, 500 µg mL−1 herring sperm single-stranded DNA (ssDNA; Sigma) for 2 h at room temperature before being incubated in hybridization solution (50 ng probe, 50% denized form of amide, 5 × SSC, 5 × Denhardt’s solution, 250 µg mL−1 yeast RNA, 500 µg mL−1 herring sperm ssDNA; Sigma). The slides were then washed with 0.2 × SSC buffer containing 0.05% tween 20 at 70°C for 3 × 10 min. Finally, BM purple was used to visualize the signal. For immunostainings, sections were blocked with 10% donkey serum in 0.1% Triton-X for 30 min at room temperature. Sections were incubated with a relevant primary antibody overnight at 4°C and with secondary antibodies for 1 h at room temperature. Antibodies were diluted in blocking buffer and sections were mounted in mounting medium (DakoCytomation) containing 5 mg mL−1 1,4-diamidino-2-phenylindole (Sigma)34.

Cryosectioning and in situ hybridization for miRNAs. Limb samples were collected at the indicated timepoints and fixed overnight at 4°C, in freshly made 4% PFA in PBS. After fixation, samples were washed three times in PBS and PBST, equilibrated in 30% sucrose in PBS, embedded in Tissue Tek (Sakura, Torrance, CA, USA) and frozen for cryosectioning. Frozen sections were sectioned into 20 µm-thick longitudinal sections. Tissue sections were dehydrated through a MeOH/PBST series (25, 50, and 75%), followed by MeOH for 10 min and rehydrated through the opposite series of MeOH. The sections were then permeabilized with 2 µg mL−1 proteinase K (Thermo Scientific), inactivated with glycin solution (2 mg mL−1 in PBST) and washed in PBST. Before hybridization, the slides were incubated with hybridization buffer (50% formamide, 5 × SSC, 50 µg mL−1 heparin, 0.2% Tween-20, and 100 µg mL−1 yeast tRNA (Sigma)) for 1 h at RT. Digoxigenin-labeled locked nucleic acid (LNA) probes against the mature form of miRs-93b and 427 were custom-made with a relevant primary antibody overnight at 4°C and with secondary antibodies for 1 h at room temperature. Antibodies were diluted in blocking buffer and sections were mounted in mounting medium (DakoCytomation) containing 5 mg mL−1 1,4-diamidino-2-phenylindole (Sigma)34.

Small RNA sequencing. Libraries from small RNAs were prepared using TruSeq Small RNA preparation kit. A total of 1 µg of total RNA from 10 samples were processed according to manufacturer instructions. The libraries were sequenced on Illumina MiSeq (Rapid mode), single read 1 × 50 bp.

microRNA quantification. microRNA expression levels were quantified using miR-Deep2 (quantifier.pl). To identify miRNAs associated with regeneration. Since our small-RNA dataset was intended for discovery, not differential expression analysis we did not have replicates for each source. To mitigate the lack of replicates in pursuit of miRNAs associated with regeneration, we divided the miRNA expression data into three groups: adult tissue (brain, eyes, heart, liver, lung), regenerating limb (limb 3 dpf, limb 7 dpfa); and embryo (larvae, late embryo). Then we tested whether there was significant miRNA up or downregulation between these groups based on a Gaussian-Laplace distribution, as implemented in moose26 (http://grabherr.github.io/moose2/).

The top 10 miRNAs upregulated in the regenerating group compared to the adult group (ranked by Bonferroni-corrected p-value) were also the top 10 upregulated in the embryonic group compared to the adult group.

Gene expression tissue specificity. We used the taur score to quantify tissue specificity of salamander-specific orthology groups compared to orthology groups with members from other vertebrates. A taur score represents the average ratio of a gene’s expression between each tissue and the tissue with the highest expression of that gene. Note that “tissue” often refers to a body part that is a composite of tissues (i.e., heart, brain, etc). A gene with equal expression across tissues will yield an average ratio of 1. In order to make the taur score intuitive (a higher score corresponds to higher tissue specificity), the average ratio is subtracted from 1. Therefore, a gene with equal expression across tissues will have a taur score of 1 = 0. Conversely, if a gene is expressed in one tissue and not expressed in the others, the average ratio will approach zero and after subtracting from 1, the taur score will approach 1. Therefore, a gene is more tissue specific the closer its taur score is to 1.

Data availability. Datasets are deposited under BioProject PRJNA353981 and BioSample accession SAMN07571895, Runs SR6001098-SRR6001140. Transcriptome and genome assemblies are available upon request.

Received: 6 September 2017 Accepted: 26 October 2017
Published online: 22 December 2017

References

1. Tanaka, E. M. The molecular and cellular choreography of appendage regeneration. Cell 165, 1598–1608 (2016).
2. Gemberling, M., Bailey, T. J., Hyde, D. R. & Posp, K. D. The zebrafish as a model for complex tissue regeneration. Trends Genet. 29, 611–620 (2013).
3. Galliot, B. & Chera, S. The Hydra model: disclosing an apoptosis-driven generator of Wnt-based regeneration. Trends Cell Biol. 20, 514–523 (2010).
4. Roberts-Galbraith, R. H. & Newmark, P. A. On the organ trail: insights into organ regeneration in the planarian. Cont. Opin. Genet. Dev. 32, 37–46 (2015).
5. Brookes, J. P. Amphibian limb regeneration: rebuilding a complex structure. Science 276, 81–87 (1997).
6. Simon, A. & Tanaka, E. M. Limb regeneration. Wiley interdisciplinary Rev. Dev. Biol. 2, 291–300 (2013).
7. Sandoval-Guzman, T. et al. Fundamental differences in dedifferentiation and stem cell recruitment during skeletal muscle regeneration in two salamander species. Cell Stem Cell 14, 174–187 (2014).
8. Looso, M. et al. A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. Genome Biol. 14, R16 (2013).
9. Abdullaev, J., Kirkham, M., Björklund, A. K., Simon, A. & Sandberg, R. A reference transcriptome and inferred proteome for the salamander Notophthalmus viridescens. Exp. Cell. Res. 319, 1187–1197 (2013).
10. Bryant, D. M. et al. A tissue-mapped axolotl de novo transcriptome enables model for complex tissue regeneration. Curr. Opin. Genet. Dev. 29, 1608 (2016).
11. Che, R., Sun, Y., Wang, R. & Xu, T. Transcriptomic analysis of endangered Notophthalmus viridescens. Exp. Cell. Res. 291, 219–229 (2013).
12. Nakamura, K. et al. Transcriptome and inferred proteome for the salamander. PLoS ONE 9, e109831 (2014).
13. Connolley, T. et al. Molecular genetic system for regenerative studies using newts. Dev. Growth Differ. 55, 229–236 (2013).
14. Lacroix, J. Etude descriptive des chromosomes en écouvillon dans le genre pleurodeles (Amphibien, Urodèle). Ann. Embryol. Morphogen. 1, 179–202 (1968).
15. Li, R. et al. ReAS: recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. PLoS Comput. Biol. 1, e13 (2005).
16. Bao, W., Kojima, K. K. & Kohnan, O. RePhase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6, 11 (2015).
hybridization experiments. We thank Marc Friedlander for advice regarding miRNA prediction. A.E. is supported by NIH Ruth Kirschstein postdoctoral fellowship F32GM117806. This work was supported by grants from the National Natural Science Foundation of China (31771617) to H.W., the European Research Council, the Swedish Research Council, Wenner-Gren Foundation and Cancerfonden to A.S.

Author contributions
A.E. conceived the study, assembled and annotated the *P. waltl* transcriptome, predicted miRNAs in the genome and transcriptome, performed orthology group analysis and wrote the manuscript. H.W. performed the *Pax3* and *Pax7* in situ hybridization, gene editing, genotype and phenotype analysis and wrote the corresponding part of the manuscript. A.J. established the *P. waltl* colony and breeding program, helped in the design of the sampling strategy, performed karyotyping from primary cell cultures, CRISPR/Cas9 egg injections, phenotype characterization, breeding, and assisted with figure preparation for the manuscript. M.P.-M. prepared, acquired and analyzed lampbrush chromosome images. Z.Y. preformed nuclear DNA quantification. Y.A. assisted with gene expression analysis. N.Z. performed parts of the transposable element analysis. G.B. performed miRNA in situ hybridization. A.K. and L.S.H. performed in situ hybridization. I.A. and R.S. contributed to *P. waltl* transcriptome assembly. M.G. designed the transposable element analysis, assembled repeat libraries, and identified repeat content, transposable element distribution and wrote the manuscript. B.A. participated in the planning of the study and supervised the genome sequencing and analysis and edited the manuscript. A.S. conceived the study, supervised all aspects of the project and wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01964-9.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

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