The axolotl genome and the evolution of key tissue formation regulators

Sergej Nowoshilow1,2,3,*, Siegfried Schloissnig4,*, Ji-Feng Fei5,*, Andreas Dahl1,6, Andy W. C. Pang7, Martin Pippel4, Sylke Winkler1, Alex R. Hastie7, George Young8, Juliana G. Roscito1,9,10, Francisco Falcon11, Dunja Knapp3, Sean Powell4, Alfredo Cruz11, Han Cao1, Blanca Habermann12, Michael Hiller1,9,10, Elly M. Tanaka1,2,3,† & Eugene W. Myers1,10

Salamanders serve as important tetrapod models for developmental, regeneration and evolutionary studies. An extensive molecular toolkit makes the Mexican axolotl (*Ambystoma mexicanum*) a key representative salamander for molecular investigations. Here we report the sequencing and assembly of the 32-gigabase-pair axolotl genome using an approach that combined long-read sequencing, optical mapping and development of a new genome assembler (MARVEL). We observed a size expansion of introns and intergenic regions, largely attributable to multiplication of long terminal repeat retroelements. We provide evidence that intron size in developmental genes is under constraint and that species-restricted genes may contribute to limb regeneration. The axolotl genome assembly does not contain the essential developmental gene Pux3. However, mutation of the axolotl Pux3 paralogue Pux7 resulted in an axolotl phenotype that was similar to those seen in Pux3−/− and Pux7−/− mutant mice. The axolotl genome provides a rich biological resource for developmental and evolutionary studies.

Salamanders boast an illustrious history in biological research as the animal in which the Spemann organizer4 and Sperry’s chemoaffinity theory of axonal guidance2 were discovered. Since 1768, when Spallanzani discovered tail and limb regeneration, researchers have probed this animal’s remarkable regenerative capabilities with increasing molecular resolution. *A. mexicanum* (Fig. 1a) was first collected by von Humboldt, and has been cultivated in the laboratory since 1864 as a model for investigating phenomena such as nuclear reprogramming, the embryology of germ-cell induction, retinal neuron processing and regeneration5. Owing to the ease with which *A. mexicanum* can be bred in the laboratory, a sophisticated molecular toolkit has been developed for this species, including germline transgenesis and CRISPR-mediated gene mutation as well as viral and other transfection methods. These tools have enabled discoveries such as the identification of the source cells of regeneration and molecular pathways that control regeneration4,5. A full exploitation of the axolotl model, including understanding regeneration and why it is limited in other tetrapods, requires analysis of its genome regulation and evolution. However, efforts towards comprehensive assembly of salamander genomes have been challenging owing to their large genome sizes (14–120 Gb) and the large number of repetitive regions they contain; the 32-Gb axolotl genome is ten times the size of the human genome. Here we report the sequencing, assembly and analysis of the axolotl genome.

A long-read assembler for large genomes

Our aim was to generate a genome sequence assembly for the d/d axolotl strain (Fig. 1a), which is commonly used in laboratory regeneration studies owing to its compatibility with live imaging. Taking into consideration the expected challenge of assembling the complex 32-Gb genome6, we sequenced 110 million long reads (32× coverage, N50 read length 14.2 kb) using Pacific Biosciences (PacBio) instruments (Supplementary Information section 1) to avoid the read sampling bias that is often found when using other technologies and to span repeat-rich genomic regions that cause breaks in short-read assemblies (Fig. 1b, c).

We developed an assembly algorithm (MARVEL) that integrates a two-phase read-correction procedure that keeps long PacBio reads intact for assembly (Supplementary Information section 2). MARVEL produced a contig assembly with an N50 of 218 kb. Next, we used 7× Illumina-based sequencing to correct sequence errors in 1% of the contig bases (Fig. 1b), which yielded a sequence accuracy of more than 99.2%. On the basis of the Illumina data, we estimated a heterozygosity of 0.47% (Supplementary Information section 2.2).

To provide a scaffold for the contig assembly, we generated de novo optical maps using the Bionano Saphyr system (Supplementary Information section 2.3). The Bionano map resolved contig chimeras, which were found in 1.7% of contigs, slightly reducing N50 contig length to 216 kb (Fig. 1d). The final hybrid assembly yielded an N50 scaffold length of 3 Mb. Compared to the short-read assembly of the 20-Gb spruce genome7 or the 22-Gb loblolly pine genome8, which involved 12× long-read coverage, the axolotl assembly showed 56- and 29-fold improvements in contiguity, respectively (Table 1).

To assess the completeness of the assembly (Supplementary Information section 4.1), we first determined the number of aligning non-exonic ultraconserved elements9 (UCEs). We found that 194 (98.5%) of 197 non-exonic UCEs that are conserved across vertebrates align to the axolotl assembly. By comparison, 189 and 192 UCEs align to the Tibetan frog and *Xenopus* genomes, respectively, and 195 UCEs align to the coelacanth genome, indicating that the completeness of the axolotl genome assembly is comparable to or better than the two other amphibian genomes, which are smaller than 2.3 Gb10.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
To further assess the completeness of the assembly, we generated a comprehensive gene catalogue by sequencing mRNA from 22 tissues (Supplementary Information section 3). Tissue-specific transcriptome assemblies and a composite assembly of all 1.5 billion transcript reads resulted in 180,649 transcript contigs (Supplementary Table 6) that contained 99% of the conserved core eukaryotic genes\(^1\) and achieved the highest BUSCO score (http://busco.ezlab.org/) of an axolotl transcriptome reported to date (Supplementary Information section 3.4). More than 85% of the transcripts aligned to the genome along at least 95% of their length (Supplementary Information section 3.5), confirming the high completeness of the assembly. Furthermore, 71% of transcript contigs in which more than 95% of the sequence aligned with the genome were located on single scaffolds, demonstrating the high contiguity of the assembly. Using this comprehensive transcript set, we annotated a total of 23,251 protein-coding genes in the axolotl genome, a similar number to those found in other vertebrate genomes (Supplementary Information section 4.2).

### Expansion of long terminal repeat retroelements

Given the similar number of genes in the *A. mexicanum* genome in comparison to other smaller vertebrate genomes, we analysed repetitive sequences (Supplementary Information sections 4.2.2, 4.2.3). Repetitive sequences made up 65.6% of the contig assembly, representing a total of 18.6 Gb. Distinct long terminal repeat (LTR) retroelement classes and endogenous retroviruses made up the largest portion of the repetitive sequences (Fig. 2a, b, Supplementary Table 13) and included elements of more than 10 kb in length (Fig. 2c, Extended Data Fig. 1). Such long elements pose challenges for assembly, and indeed 97% of contigs ended in LTR elements. The number of substitutions to the repeat consensus, which is an estimate of the relative age of the LTR retroelement, indicates that the axolotl genome has undergone a long period of transposon activity followed by a recent and apparently continuing burst of expansion (Fig. 2d). This profile is consistent with previous small-scale characterizations of other salamander genomes\(^2\).

The presence of many repeated elements contributes to a median intron size (22,759 bp) 13, 16 and 25 times that observed in human (1,750 bp), mouse (1,469 bp) and frog (906 bp), respectively (Fig. 3a, Supplementary Information section 4.3), a trend that was previously observed in five genes obtained from selective bacterial artificial chromosome sequencing of the axolotl genome\(^3\). Figure 3b shows a typical gene organization in axolotl compared to its human orthologue. Consistent with intron expansion, a distance comparison of pairs of highly conserved non-exonic elements shows that intergenic regions in the axolotl genome are 12 to 17 times larger than those in human, mouse and frog (Supplementary Information section 4.4).

### HoxA cluster and intron size constraints

To examine gene cluster organization within this large genomic context, we focused on the HoxA locus, which has an important role in proximal-to-distal limb development and is reactivated during limb regeneration\(^4,5\). The entire HoxA locus is contained on a single contig (Fig. 3c), and the conserved neighbouring gene *Evx1* is contained on the same 3.34-Mb scaffold. Compared to the orthologous human and frog clusters, the *A. mexicanum* HoxA cluster has a substantially increased repeat content and is 3.5 times larger, mostly owing to a 170-kb expansion between *HoxA3* and *HoxA4* (Fig. 3c). Notably, highly conserved non-exonic elements shows that intergenic regions in the axolotl genome are 12 to 17 times larger than those in human, mouse and frog (Supplementary Information section 4.4).
On the basis of these observations, we examined the intron size distribution among a larger set of orthologous genes involved in developmental processes. While introns of non-developmental genes in axolotl show a median size expansion of 13- to 25-fold compared to human, mouse and frog, the expansion of introns of developmental genes is significantly lower (6- to 11-fold, \( P < 10^{-11} \)) (Fig. 3a, Supplementary Information section 4.3). In contrast to human, mouse and frog, introns of developmental genes in axolotl are shorter than introns of non-developmental genes. Furthermore, axolotl multi-exon genes that contain only short introns exhibit gene ontology enrichments related to developmental patterning that are not enriched in multi-exon genes with larger introns (Supplementary Table 16). These results suggest that intron size in developmental genes is under constraint in the axolotl, possibly because smaller gene sizes facilitate rapid transcription and thus upregulation of these genes in specific developmental contexts.

A reduced Pax-family complement

Next, we interrogated the genome for families of canonical developmental signalling molecules (Supplementary Information section 5). All three hedgehog paralogues as well as a full set of vertebrate Wnt genes were present (Extended Data Fig. 2a, b). However, we noted that certain members of the paired box family of transcription factors, which have diverse roles in tissue formation, were not found in the assembly. Consistent with the absence of Pax4 in amphibians and other vertebrate lineages\(^1\), the axolotl genome does not contain Pax4 but does contain Pax10. Notably, despite the presence of the Pax3 and Pax7 paralogues in all other known vertebrate lineages, we were able to identify Pax7 but not Pax3 in the axolotl genome assembly (Extended Data Fig. 2c). No Pax3 sequence was found in either the raw PacBio sequencing reads or the transcriptome. To confirm the loss of Pax3, we further examined the genomic region that would be syntenic for Pax3 in the presence of neighbouring genes and highly-conserved non-exonic elements (CNEs). The orthologues of genes surrounding mouse Pax3 (Sgpp2 and Epha4) were present in the *A. mexicanum* genome assembly; however, neither the Pax3 gene nor any of the Pax3-associated CNEs were found (Fig. 3d). By contrast, several CNEs that overlap the Pax7 gene were identified in the assembly. Together, this evidence strongly suggests that Pax3 and several of its cis-regulatory elements are absent in the axolotl genome, possibly owing to a deletion.

**Axlolotl Pax7 has similar functions to Pax3**

To functionally assess the consequence of the absence of Pax3 in the axolotl, we used TALEN- and CRISPR-mediated gene editing\(^1\) to mutate Pax7. In other vertebrates, Pax3 and Pax7 play key roles in muscle, neural tube and neural crest-derived tissue development\(^1\). Although these two genes share some common functions, deletion of Pax3 or Pax7 causes distinct phenotypes in mice\(^20-22\). We investigated whether frameshift deletions introduced into the AmPax7 gene would yield a comparable Pax7 phenotype, or whether AmPax7 may have taken on functions that are carried out by Pax3 in other vertebrates. Two different TALEN-mutant alleles (7-nt and 20-nt deletions) of AmPax7 were bred through two generations (Fig. 4a, Supplementary Information section 6). In the F2 generation, the developmental phenotype described below was observed in 83 out of 345 (24%) progeny from the *Pax7\(^{Δ20nt/Δ20nt} \)* intercrossing and 57 of 232 (24.6%) progeny from the *Pax7\(^{Δ7nt/Δ7nt} \)* intercrossing (Fig. 4b, Extended Data Fig. 3). The phenotype was evident in homozygous mutants, as analysed by PCR and loss of protein (Supplementary Information sections 6.1, 6.3). This information, combined with the CRISPR-mediated gene mutation results (Supplementary Information sections 6.2), shows that the homozygous *Pax7\(^{Δ20nt/Δ20nt} \)* and *Pax7\(^{Δ7nt/Δ7nt} \)* mutants represent recessive, complete or partial loss of Pax7 function.

The *Pax7\(^{Δ20nt/Δ20nt} \)* and strong F0 *Pax7*-CRISPR mutants exhibited a curved body, were unable to maintain an erect posture and exhibited a delay in growth compared to controls. Immunohistochemical analysis of trunk or tail cross-sections of early stage, 20-day-old *Pax7\(^{Δ20nt/Δ20nt} \)* and 17-day-old F0 *Pax7*-CRISPR axolotls showed normal muscle mass. However, at later stages, consistent with the mouse *Pax7* deletion phenotype, tail and trunk muscles were greatly decreased (Fig. 4c, Extended Data Figs 4–6). Remarkably, the *Pax7* mutant axolotls also completely lacked limb muscle (Figs 4d, Extended Data Fig. 7). In mice, *Pax3*, but not *Pax7*, is required for limb muscle formation\(^21-23\) (Supplementary Table 18). These results demonstrate that AmPax7 has comparable functions to MmPax3 in the control of limb muscle genesis.
In mice, Pax7 deletion affects craniofacial neural crest derivatives, including the facial bones, whereas in zebradishes, Pax7 mutants show loss of xanthophores and reduction of melanophores, but no loss of iridophores. The AmPax7 mutants lacked a prefrontal bone, had a reduced number of melanophores, and were deficient in xanthophores and iridophores except in the eyes (Fig. 4e–g, Extended Data Fig. 8). Pax3 deletion in mice is associated with neural tube closure defects (Supplementary Table 18). Similarly, Pax7Δ20ntΔ21nt TALEN and Pax7-CRISPR axolotls displayed failed closure of the neural tube in the midbrain (Fig. 4h, Extended Data Fig. 9). In summary, mutation of AmPax7 yields a combination of the Pax3 and Pax7 mutant phenotypes that are observed in other vertebrates (Supplementary Information section 6). It will be interesting to understand how the regulation of Pax7 has changed in axolotl to enable the loss of Pax3, which is essential in other vertebrates.

**Species-restricted genes in regeneration**

Previous searches for mRNA and microRNA (miRNA) transcripts associated with limb regeneration relied on mapping to de novo transcriptome assemblies. We sought to re-examine these datasets using our newly acquired genomic data. Recent functional work has highlighted the role of diverged gene or protein function during regeneration25–27. Analysis of published tissue-enriched datasets28, combined with regeneration time courses29,30 and our own transcriptional profiling of 22 tissues, identified five transcripts that are upregulated in the limb blastema (the mass of proliferating cells involved in regenerating the limb) with orthology limited to non-amniote vertebrates (Supplementary Information section 7). One of these protein sequences shows a weak similarity to tectorin, a basement membrane component normally found in the inner ear, consistent with studies that implicate extracellular matrix components with having an important role in limb regeneration31,32. Notably, another of these transcripts encodes a Ly6 family member Prod1 as a key factor involved in salamander limb development and regeneration25,33. Our results suggest that Ly6 family members have a broader role in limb regeneration. Finally, we also investigated the role of non-coding RNAs by mapping a dataset of small RNA sequences expressed in the limb and limb blastema34 to our genome assembly. This analysis classified 93 small RNAs as orthologue, resulting in a gene that is 4.3 times larger. **c**, Comparison of genes and repetitive elements in the HoxA cluster. CNEs that align to the axolotl HoxA cluster are shown in red. d, Axolotl lacks the Pax3 locus. Analysis of tetrapod-conserved genes and CNEs associated with Pax7 and Pax3 genomic loci. Red, gene sequences and CNEs that are absent in axolotl; *CNEs that overlap well-characterized mouse Pax3 enhancers38,39.  

**Figure 3 | Genome organization and loss of Pax3.** a, Intron size of developmental genes appears to be under constraint in A. mexicanum. Violin plots represent the full distribution of intron sizes (thick bar, first to third quartile; white dot, median). * * * * P < 10–11; two-sided Wilcoxon rank-sum tests. Quantitative data are shown in Source Data. b, Organization of the Agr2 exon–intron structure shows a consistent expansion of axolotl intron sizes compared to those in the human genome. Red, gene sequences and CNEs that are absent in axolotl; *CNEs that overlap well-characterized mouse Pax3 enhancers38,39.  

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Figure 4 | Pax7 mutation in A. mexicanum yields a phenotype similar to that of Pax3−/−Pax7−/− mice. a. Deletion of AmPax7 coding sequences using TALEN and CRISPR. Deletions were made in exon 1 or exon 2. The first three AmPax7 exons (Ex) are shown. Red rectangles, TALEN targets; arrows, CRISPR-guide RNA (gRNA)-binding sites. b. Images of 6-month-old AmPax7Δ20nt/Δ20nt mutants compared to controls show loss of body elongation. Scale bars, 1 cm. c. Reduced body wall muscle in Pax7 mutants. Immunofluorescence images of myosin heavy chain (MHC, red) and DAPI (blue) in trunk cross-sections from a 6-month-old AmPax7Δ20nt/Δ20nt mutant (Pax7-TALEN#2) and a control animal. Scale bar, 500 μm. d. Limbs of Pax7 mutants lack muscle. Forelimb (left; scale bars, 500 μm) and immunofluorescence images of MHC (red) and DAPI (blue) in limb cross-sections (right; scale bars, 100 μm) of a 3-month-old AmPax7Δ20nt/Δ20nt mutant and a control animal. e. Loss of prefrontal bone in Pax7 mutants. Dorsal and lateral views of Alcian blue and Alizarin red-stained AmPax7−Ex1-CRISPR#3 and AmPax7−Ex1-CRISPR#3 mutants and controls (right). Red arrows, prefrontal bone; blue arrows, maxillary bone. The yellow arrowhead points to a small remnant of the prefrontal bone. Scale bar, 1 mm. f, g. Reduced melanophores, xanthophores (f) and iridophores (g) in Pax7 mutants. Images of 17-day-old AmPax7−Ex2-CRISPR#1 (f), 2-month-old AmPax7Δ20nt/Δ20nt (g) mutants and controls (Crt). Right panels in f show a magnified view of the outlined area; red arrows in g point to the eyes that are magnified in the insets; black arrows indicate the belly iridophores in the control. Scale bars, 1 mm. h. Neural tube closure defect in Pax7 mutants. Images of a 31-day-old AmPax7−Ex2-CRISPR#1 mutant and control (Crt). Right, magnified view of the outlined area. Scale bar, 1 mm. Quantitative data and sample sizes are provided in the Life Sciences Reporting Summary and Source Data.

in amniotes. Future investigations of such sequences are likely to be a fruitful avenue for understanding the evolution of regeneration capabilities.

Discussion

We have generated a comprehensive whole-genome assembly for the salamander A. mexicanum, and analysis of this assembly has allowed us to draw conclusions about the structure of the expanded genome. Our data, together with data from plants and partial data from several other salamander species, show that LTR expansion is a major contributor to giant genome size across animals and plants. Our assembly is sufficiently complete to reliably detect the absence of Pax3, which is present in fish and other amphibians. This analysis was confirmed using gene editing, which showed that AmPax7 has assumed functions that are carried out by Pax3 in other animals.

Functional analysis of axolotl development, physiology and regeneration is facilitated by the availability of tissue- and time-dependent gene expression profiles. The axolotl genome provides a foundation for applying methods such as chromatin immunoprecipitation with sequencing (ChIP–seq) or assay for transposable-accessible chromatin using sequencing (ATAC-seq) to investigate the genomic basis of gene regulation during regeneration. Together with methods such as CRISPR-mediated gene editing, viral expression methods, transplantation and transgenesis, the axolotl is a powerful system for studying questions such as the evolutionary basis of its remarkable regeneration ability. Our approach of long-read sequencing, optical mapping and genome assembly using MARVEL also demonstrates that it is now feasible to assemble very large repeat-rich genomes.

Online Content

Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April; accepted 13 December 2017.

Published online 24 January 2018.


Acknowledgements E.W.M. was supported by the Max Planck Society President’s fund and BMBF grants 01IS14014C and 0311L0102. E.M.T., S.N. and J.-F.F. were supported by MPI-CBG, DFG FI111 and IMP. E.M.T. and J.-F.F. were supported by DFG TA 2744/3-1. J.-F.F. was supported by NSFC grant 31711611 and a Research Starting Grant S821111-E2 from SCNU. F.F. was supported by a CONACYT Masters Fellowship, A.C. was supported by The Swedish IRLG GRANT 2014–9040–114152–32, CONACYT FOINS–301 and Ciencia Básica IO017–CB-2018–01–0000000000292126. G.Y. was supported by the Crick institute award FC0011662 (to J. P. Stoye). The Crick receives its core funding from CRUK, the MRC and the Wellcome Trust. B.H. was supported by BMBF grant 01H111003C. We thank W. Bonacci, E. Gromberg, A. Tajaki and I. Stützer for sample preparation; S. Clauing, N. Gacheidei, Y. Duport, A. Sommer and the VBCF NGS facility for sequencing; A. Kavirayani and T. Engelmaier at VBCF for histology; E. Heude for confirmation of Pax7 mutant immunofluorescence; T. Anantharaman for optimizing assembly and scaffolding parameters for Bionano data; and G. Papoutsoglou for advice.

Author Contributions S.N. assembled and analysed the transcriptome, and performed genome analysis. S.S. created, engineered and implemented the assembly algorithm and analysed the genome assembly. J.-F.F. performed DNA extraction and all biological experiments and analysis. M.P. and S.P. contributed to the assembly, assembled the genome and contributed to the genome analysis. A.D. and S.W. performed Pacific Biosciences sequencing. G.Y. analysed the TLR elements. J.G.R. performed the conserved element and intron analysis. F.F. and A.C. performed the developmental orthologue and miRNA analysis. A.W.C.P., A.R.H. and H.C. performed Bionano optical mapping, generated the hybrid and scaffolded the assembly. D.K. performed tectorin intron analysis. B.H. supervised transcriptome assembly and annotation and acquired funding. E.M.T. and M.H. wrote the manuscript. E.W.M. conceived and implemented the assembly strategy, acquired major funding and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.H. (hillcr@mpl-cbg.de), E.M.T. (elly.tanaka@imp.ac.at) and S.S. (siegfried.schloissnig@h-its.org).

Reviewer Information Nature thanks C. Crews and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Axolotl genomic DNA was prepared from freshly isolated liver and spleen of an individual three year old adult d/d male using DNAzol followed by phenol/chloroform extraction and ethanol precipitation.

A total of 50 size-selected SMRTbell libraries were prepared with a minimum fragment length cutoff between 10 kb and 20 kb. We sequenced medium and large insert libraries on the PacBio RSII instrument, making use of three different sequencing polymerases (P4, P5 and P6) and the corresponding sequencing chemistries (C2, C3 and C4). Movie times ranged from 180 min to 360 min with the majority of SMRT cells (1,414 of 1,933) at 240 min.

Sequences were assembled using the MARVEL assembler.

Optical mapping was performed using the Saphyr System (Bionano) based on NanoChannel array Technology. DNA was labelled with Nt.BspQI and Nb.BssSI enzymes in separate labelling reactions. Each enzyme reaction was run on the Saphyr System. 2.813 Tb of data were collected on three Saphyr Chips for Nt.BspQI and 2.0 Tb of data were collected on two Saphyr Chips for Nb.BssSI samples; single molecule N50 lengths were 240 kb and 184 kb, respectively. Each dataset was de novo assembled using Bionano Solve 2.1 software.

RNA was isolated from 22 tissue types using TRIzol or RNeasy reagents and sequenced using Illumina technology. The Trinity software package was used for transcriptome assembly.

**Code availability.** The MARVEL assembler with documentation is available at https://github.com/schloi/MARVEL.

**Data availability.** A browser of the axolotl genome is available at https://genome.axolotl-omics.org. The transcriptome assembly and the genome and transcriptome BLAST database can be accessed at https://www.axolotl-omics.org with no restrictions. The sequence data and both assemblies have been deposited in the NCBI BioProject database with accession numbers PRJNA378970 (genome data) and PRJNA378982 (transcriptome data). Both genome data and transcriptome data were deposited to the NCBI Nucleotide Database (nuccore) with accession numbers PGSH00000000 and GFZP00000000, respectively.
Extended Data Figure 1 | Analysis of LTR retroelement frequencies according to their lengths. The line shows a moving average (period 25) to highlight clusters of elements of similar lengths.
Extended Data Figure 2 | Phylogenetic trees. 

a, Phylogenetic tree of vertebrate hedgehog proteins show the presence of axolotl orthologues. 
b, Phylogenetic tree of vertebrate Wnt proteins show the presence of axolotl orthologues in all Wnt classes. 
c, Phylogenetic tree of vertebrate PAX proteins. Pax4 and Pax3 are absent in axolotl.
Extended Data Figure 3 | Developmental phenotype of Pax7 mutants. 

a, b, Images of live Pax7-Δ20nt/Δ20nt mutants compared to controls show no obvious phenotype at early stages (a, 1 month), but an obvious phenotype at later stages (b, 3.5 months). c, d, Images of live F0 Pax7-Ex1-CRISPR#3 (c, 4 months) and Pax7-Ex2-CRISPR#1 (d, 6 months) mutants show the curved body phenotype. Scale bars, 1 cm. Numbers of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 4 | Progressive depletion of the trunk muscle in Pax7 mutants. 

**a**, Images of live Pax7Δ20nt/Δ20nt animals at different ages compared to corresponding controls show the progressive loss of the trunk muscle in mutant animals. Black arrows indicate trunk muscles; blue arrows highlight the visibility of the spine after reduction and/or depletion of trunk muscle. Scale bars, 2 mm.

**b**, **c**, Images of live 7-month-old F0 Pax7-Ex1-CRISPR#3 (**b**) and 6-month-old F0 Pax7-Ex2-CRISPR#1 (**c**) mutants compared to controls, showing loss of trunk muscle. Black arrows indicate trunk muscles; blue arrows indicate the visibility of the spine after depletion of trunk muscle. Scale bars, 5 mm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 5  Progressive depletion of the tail muscle in Pax7\textsuperscript{Δ20nt/Δ20nt} mutants. a, b, Images of live 75-day (a) and 6-month-old (b) Pax7\textsuperscript{Δ20nt/Δ20nt} homozygous mutants compared to controls show the progressive depletion of tail muscle. White arrows indicate tail muscle fibres; right, magnified view of the outlined area. Note the decrease in myotome length in 75-day-old Pax7\textsuperscript{Δ20nt/Δ20nt} homozygous mutants (b). Scale bars, 500 μm.

c, Immunofluorescence images of MHC (red) and DAPI (blue) in tail cross-sections show reduction in tail muscle in 3-month-old Pax7\textsuperscript{Δ20nt/Δ20nt} mutants compared to controls. Scale bar, 100 μm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 6 | Depletion of the abdominal muscle in Pax7 mutants. a, b, Bright-field image and GFP fluorescence of live 34-day (a) and 64-day-old (b) F0 Pax7-Ex1-CRISPR#3 mutants obtained by injecting Pax7-gRNA#3–CAS9 protein complex into eggs of CarAct:EGFP transgenic axolotls, compared to un-injected CarAct:EGFP controls. Mutant animals show a reduction in the EGFP-labelled abdominal muscles. Right, magnified view of GFP fluorescence in the outlined area; white arrows indicate forelimbs that either contain or lack EGFP-labelled muscles; green arrows indicate the GFP-labelled abdominal muscle; red arrows indicate regions that lack GFP-labelled abdominal muscle. Scale bars, 1 mm. c, Images of live 6-month-old Pax7Δ20nt/Δ20nt mutants compared to controls show the loss of abdominal muscle. Scale bars, 1 mm. d, Immunofluorescence images of MHC (red) and DAPI (blue) in cross-sections show the presence of ventral body-wall muscle in the chest position in 6-month-old Pax7Δ20nt/Δ20nt homozygous mutants compared to controls, and the gradual depletion of the abdominal muscle along the anterior–posterior axis. Arrows, skin; yellow stars, liver; green stars, intestine. Scale bar, 100 μm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 7 | Loss of limb muscle in Pax7 mutants.
a, Images of live 54-day-old Pax7Δ20nt/Δ20nt mutants compared to controls show loss of limb muscle. Right, magnified view of the outlined area. Scale bars, 1 mm. 
b, c, Non-muscle tissues are normal in Pax7Δ20nt/Δ20nt mutant limbs. Immunofluorescence images for TU-J1 (b, green) MBP (b, red), CO1A2 (c, green) and DAPI (blue) in forelimb cross-sections of 56-day-old Pax7Δ20nt/Δ20nt mutants and controls. Scale bars, 100 μm. 
d, Images of live 80-day-old F0 Pax7-Ex1-CRISPR#3 heterozygotes compared to controls show loss of forelimb (FL) and hindlimb (HL) muscle on one side of the body (green arrows) but not on the other side (red arrows). Scale bar, 1 mm. 
e, Images of live 54-day-old F0 Pax7-Ex2-CRISPR#1 and Pax7-Ex2-CRISPR#3 mutants compared to a control (bottom) showing loss of forelimb muscle in CRISPR animals. Right, magnified view of the outlined area. Scale bars, 1 mm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 8 | Reduced melanophores, loss of xanthophores and iridophores in Pax7 mutants. a, Images of a live 25-day-old Pax7Δ20nt/Δ20nt homozygous mutant compared to a control animal showing loss of xanthophores and reduction of melanophores in the head and neck region of mutant animals. Right, magnified view of the outlined area. Scale bar, 1 mm. b, Images of a live 54-day-old Pax7Δ20nt/Δ20nt homozygous mutant compared to a control animal showing a reduction in melanophores along the body. Arrows, melanophores. Scale bar, 500 μm. c, Images of a live 17-day-old F0 Pax7-Ex2-CRISPR#3 mutant compared to a control animal showing loss of xanthophores and reduction in melanophores in the head and neck region. Right, magnified view of the outlined area. Scale bar, 1 mm. d, Images of a live 2-month-old Pax7Δ20nt/Δ20nt homozygous mutant compared to a control animal showing loss of iridophores on the belly. Red arrows point to the eye, which is displayed at higher magnification showing eye pigmentation defects; green arrows indicate the presence of iridophores in the control animal (with silver eyes), but not in the mutant (with black eyes). Iridophores are absent in the Pax7-TALEN#2 mutants, irrespective of the eye colour. Scale bars, 1 mm. e, Images of live 6-month-old Pax7-Ex2-CRISPR#1 and Pax7-Ex2-CRISPR#3 mutants compared to a control animal, showing the reduction or loss of belly iridophores (right) in axolotls with silver eyes (left). Red arrows point to the eye, which is displayed at a higher magnification on the right; green arrows indicate remaining iridophores in F0 mosaic Pax7-CRISPR mutants or iridophores in the control animal. Scale bars, 1 mm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Extended Data Figure 9 | Neural tube closure defects in Pax7 mutants.

a, Images of a live 55-day-old Pax7\Delta20nt/\Delta20nt mutant compared to a control animal show an open brain phenotype. Right, magnified view of the outlined area. Scale bar, 1 mm.

b–d, Images of live 15-day (b), 31-day (c) and 6-month-old (d) Pax7-Ex2-CRISPR#1 and Pax7-Ex2-CRISPR#3 mutants compared to controls, showing an open brain phenotype. Right, magnified view of the outlined area. Scale bar, 1 mm.

e, Haematoxylin and eosin-stained paraffin cross-sections show the open neural tube of a 17-day-old F0 Pax7-Ex2-CRISPR#1 mutant compared to a control. Red arrows indicate the boundaries of the opened neural tube (NT). Scale bar, 200 μm. Number of replicate matings and experiments are shown in the Life Sciences Reporting Summary and Source Data.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

1. Sample size

Describe how sample size was determined.

For genome sequencing no sample size was calculated as we sequenced one individual axolotl. For phenotype analysis, no a priori sample size was determined. The axolotl lay hundreds of eggs per breeding, and with the very strong phenotypes observed, we could discern the phenotype in homozygous mutants with 100% correspondence of phenotype with genotype, as described in Supplement section 6.1.2. Breeding of a pair of F1 Pax7-TALEN#2 mutants produced more than 100 progeny for phenotype analysis. We analyzed phenotypes from three breedings. For Pax7-CRISPR animals, we injected at least 30 eggs for each gRNA. We repeated the injections at least twice and analyzed the Pax7-CRISPR animals derived from those injections. The precise number of animals for morphological phenotype analysis is listed in Supplement section 6.1.3 and section 6.2 and file SD Fig4.xlsx. For immunofluorescence analyses of trunk muscle and limb muscle in Pax7 TALEN#2 mutants, (Fig 5c, 6.4.8) we performed two independent experiments with the same result, (total n=4). For other immunofluorescence and histochemical datasets in figures 5, S6.4.2, S6.4.3, S6.4.5, S6.4.7, S6.4.10 we analyzed three individual animals (n=3) for each of the mutant and control groups. The experiment was performed once and gave a consistent result in all three sets of animals.

2. Data exclusions

Describe any data exclusions.

We did not exclude data from the manuscript. In the final transcriptome assembly we did not include the oocyte dataset due to the presence of numerous short RNAs in that dataset that did not assemble with other contigs.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts to replicate the mutant analysis were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

In mutant analysis, progeny were allocated as having a phenotype or no phenotype based on limb and body muscle mass. Then genotyping of both types of individuals was performed in the same experiment and the data examined for the genotype.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The investigators were not blinded.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Policy information about availability of computer code

Describe the software used to analyze the data in this study.

A new assembler, MARVEL, was generated and fully described in the manuscript. The MARVEL documentation and source code are available at: https://github.com/schloi/MARVEL

BUSCO (v3)
Canu assembler (1.3)
Blasr (4.x/5.x)
Bowtie2 (2.2.9)
Pilon (1.22)
SAMtools (1.3.1)
Bionano Solve/Access™ (2.1)
fastx_toolkit (0.0.13)
Trinity (r20140717)
Mira (4.9.3)
NCBI Blast+ (2.6.0)
CD-HIT (V4.6.6)
R/R-Studio (3.2)
Blat (36x2)
RepeatModeler (1.0.8)
RepeatMasker (4.0.6)
LTR-FINDER (1.0.6)
LTRharvest (GenomeTools) (1.5.9)
MGEScan-LTR (0.1)
BEDOPS (2.4.20)
HMMER (3.1b2)
GyDB (2.0)
MAFFT (7.271)
FastTree (2.1.9)
Ensembl Biomart (89)
lastz (1.02.00)
SPALN (2.3.0)
MUSCLE (3.8.31)
RAxML (8.2.9)
Trimmomatic (0.33)
ShortStack (3.8.1)
Adobe Photoshop (CS6)
Adobe Illustrator (CS6)
Zeiss ZEN 2 (blue edition)
Olympus CellSens Standard (1.5)
Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All unique materials are openly available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

monoclonal anti-CO1A2 (clone SP1.D8, DSHB), Validated on limb sections as staining the extracellular, region between dermis and epidermis, and surrounding muscle fascia

monoclonal anti-MBP (GTX76114, GeneTex), MBP: Myelin Basic Protein. Validated by staining on CNP:GFP transgenic animal, which expresses in glial cells. Co-localization was observed.

monoclonal anti-muscle MHC (clone 4.102538, DSHB). Validated on cultured salamander myogenic cell line - shows positive staining in myotubes and negative in myoblasts. Validated in limb sections as co-staining with muscle as defined by morphology in by DIC microscopy, which shows muscle striations. 

monoclonal anti-PAX7 (catalog number: PAX7, DSHB), validation on axolotl tissue was performed by Schnapp et al (2005) Development, 132:3243-53

monoclonal anti-TUJ-1 (MAB1195, R&D), validated by staining on betaIII-tubulin: GFP transgenic animal, which expresses in neurons. Colocalization was observed. Alexafluor 488- (A21202) and 555-(A31570) conjugated donkey anti-mouse secondary antibodies (Invitrogen)

Cy3-conjugated donkey anti-rat IgG (H+L)secondary antibody (712-165-153, Jackson ImmunoResearch)

Secondary antibodies were validated by staining sections with secondary antibody only to determine that there was no signal in the cell types to be analyzed.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used
c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Ambystoma mexicanum, strain: white (d/d). Male was used for sequencing genome.

We analysed the phenotypes at diverse stages. Pax7-TALEN animals: 20, 25, 30, 50, 54, 55, 56, 75, 80-day old axolotls, and 1, 2, 3, 3.5, 6-month old axolotls; Pax7-CRISPR animals: 15, 17, 30, 31, 54, 64, 80-day old axolotls, and 4, 6 and 7 month old axolotls.

Up to juvenile stages, it is not possible to determine the sex of the axolotls by morphology and no cytochemical or molecular assay is available. The phenotype is very likely independent of the gender.
12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not include human research participants.
CORRECTION

https://doi.org/10.1038/s41586-018-0141-z

Author Correction: The axolotl genome and the evolution of key tissue formation regulators

Sergej Nowoshilow, Siegfried Schloissnig, Ji-Feng Fei, Andreas Dahl, Andy W. C. Pang, Martin Pippel, Sylke Winkler, Alex R. Hastie, George Young, Juliana G. Roscito, Francisco Falcon, Dunja Knapp, Sean Powell, Alfredo Cruz, Han Cao, Bianca Habermann, Michael Hiller, Elly M. Tanaka & Eugene W. Myers

Correction to: Nature https://doi.org/10.1038/nature25458, published online 24 January 2018.

In the originally published version of this Article, the sequenced axolotl strain (the homozygous white mutant) was denoted as 'D/D' rather than 'd/d' in Fig. 1a and the accompanying legend, the main text and the Methods section. The original Article has been corrected online.