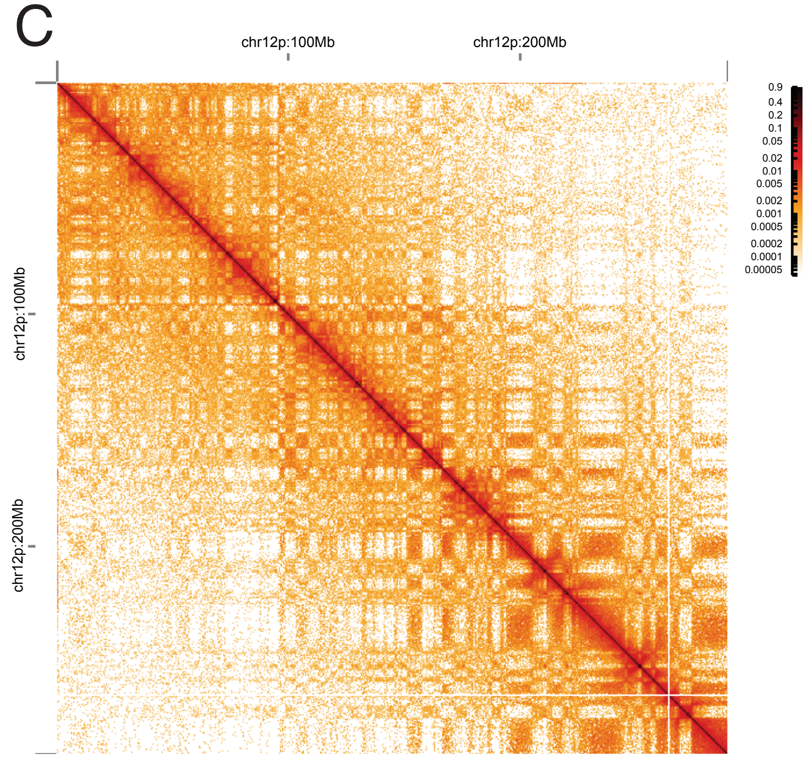
A

	Smith et al. 2019	AmexG_v6.0DD
Base assembly for scaffolding	BioNano scaffolds from Nowoshilow, Schloissnig et al 2018	Contigs from Nowoshilow, Schloissnig et al 2018 to eliminate misjoins from Bionano scaffolding
Bases in scaffolds	23.0	26.7
Assembly gaps	759,151	184,308
Orientation of scaffolded pieces	ambiguous	determined
Number of scaffolds	14 Working data set split into 20 assigned chromosomes or arms	14 Working data set split into 28 assigned arms

B

Chromosome	Chromosome length (bp)	Arm	Arm length (bp)
chr6	3090459613	p	1463442985
		q	1627016628
chr1	2890135120	p	1417697632
		q	1472437488
chr2	2797231127	p	1382406564
		q	1414824563
chr5	2566818623	p	1280495009
		q	1286323614
chr4	2421803826	p	1205706753
		q	1216097073
chr3	2395070288	p	868269716
		q	1526800572
chr7	1991948032	p	908020774
		q	1083927258
chr8	1632293852	p	745917890
		q	886375962
chr10	1608233191	p	525957290
		q	1082275901
chr9	1457894406	p	455256413
		q	1002637993
chr11	1385473386	p	305732125
		q	1079741261
chr12	1147023891	p	289110169
		q	857913722
chr13	725160601	p	241817284
		q	483343317
chr14	621144288	p	184706515
		q	436437773

C



Supplementary figure 1 referring to Figure 1. Comparison of d/d HiC scaffolding with meiotic scaffolding and with WT HiC contacts. **A.** Current HiC-based assembly (AmexG_v6.0DD) properties compared to meiotic linkage map based assembly (Smith et al, 2019). **B**. Lengths in basepairs of scaffolds and designation as chromosomes as in (Smith et al, 2019). **C**. HiC heat map of chromosome 12p made from WT erythrocyte nuclei. Large structural variants were not found between AmexG_v6.0DD and the WT erythrocyte data.

Supplementary figure 2 referring to Figure 2

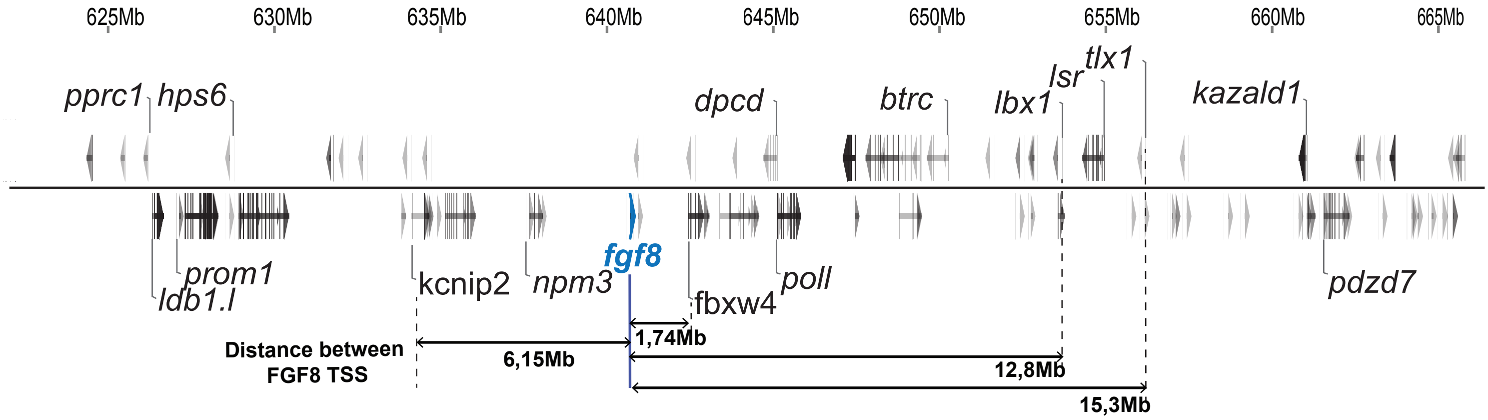
Supplementary figure 2A, 2B can be found at the end of the Supplement due to size of the image.

Supplementary figure 2A: Maximum likelihood phylogenetic trees constructed for HLA codon and protein alignments. The Bootstrap values are shown next to each node of the tree. Branches were collapsed into triangles base on their position in the tree for easier interpretation. Colored in red are the sequences that belong to axolotl, while in blue the sequences that belong to human

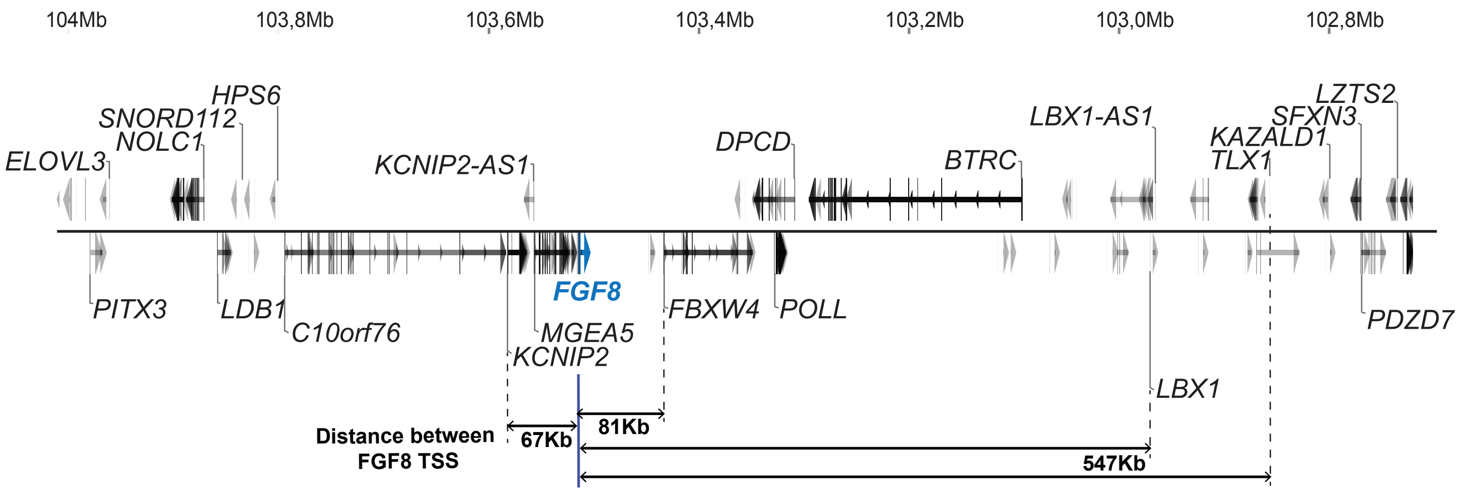
Supplementary figure 2B is available as a separate file:

Supplementary figure 2B: Maximum likelihood phylogenetic trees constructed for the TRIM protein family alignments. The Bootstrap values are shown next to each node of the tree. Branches were collapsed into triangles base on their position in the tree for easier interpretation. Colored in red are the proteins that come from the axolotl MHC, while in blue those proteins that belonged to the TRIM39 clade.

A. Axolotl *fgf8* and surrounding genes



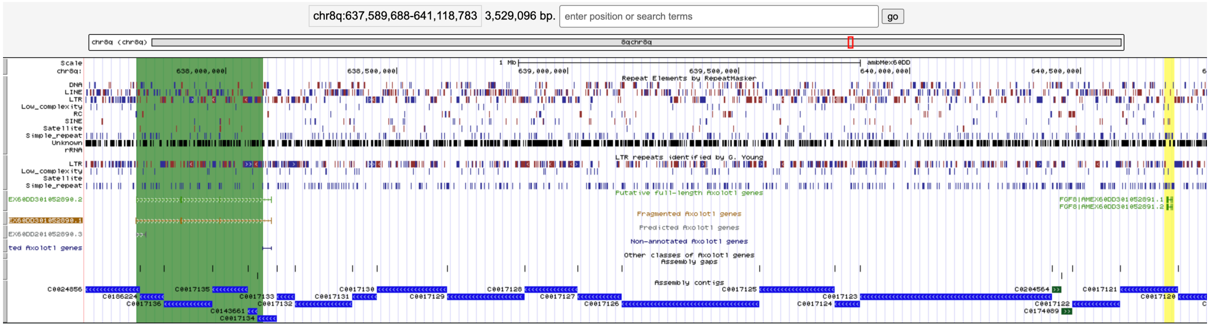
B. Human *FGF8* and surrounding genes



Supplementary figure 3 referring to Figure 3:
Gene synteny of the axolotl *fgf8* regulatory locus.
A. The genomic region surrounding the axolotl *fgf8* gene. *Fgf8* is marked with a blue arrow and the TSS is denoted in blue. Conserved gene alignment between ldb1.L to pdzd7 with the human *FGF8* genomic region is observed. The estimated genomic distances between the *Axfgf8* TSS and each gene are denoted.

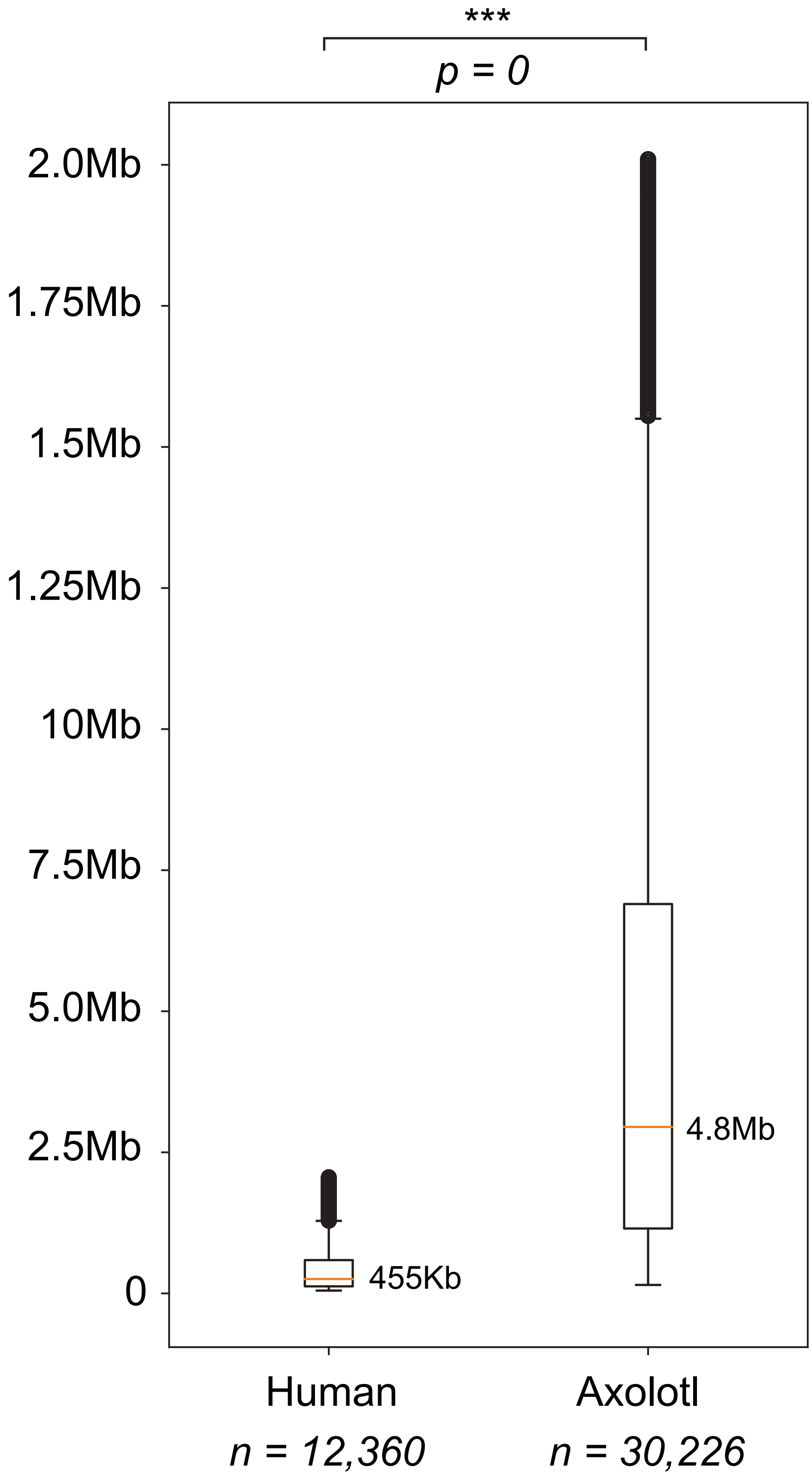
B. The genomic region surrounding the human *FGF8* gene. *FGF8* is marked with a blue arrow and the TSS is denoted in blue. The gene annotation was exported from GRCh37/hg19. The estimated genomic distances between the *hsFGF8* TSS and each gene are denoted.

Axolotl



Human (hg38)

Supplementary Figure 4 referring to Figure 4
Expanded genomic region between axolotl *Fgf8* and *Npm3* contains abundant repeats. Genome browser visualization of axolotl and human hg38 assemblies with repetitive element annotation tracks. The axolotl region (upper) contains dense arrays of diverse repetitive elements while the human region (lower) has sparse repeats.



Supplementary figure 5 referring to Figure 5: TAD size distribution.
The average TAD size in the human genome is 455 kb, while the average size in the axolotl genome is 10 times longer 4.8 Mb. The difference is significant (p=0, t-test, number of TADs -12,360 and 30,226 for human and axolotl, respectively).

Materials and Methods

Animals and cultured cell lines

Axolotls were maintained in individual aquaria and all animal breedings were undertaken by the IMP animal facility. All animal handling and surgical procedures were carried out in accordance with local ethics committee guidelines. Animal experiments were performed as approved by the Magistrate of Vienna. Animals which have white skin are called *d/d* strain. All animal surgeries and tissue amputations were carried out under anesthesia in 0.03% benzocaine (SIGMA).

Axolotl cells (AL1 cell line) were obtained from Dr. David Gardiner at UCI(1, 2). AL1 cells were maintained under optimized culture conditions. Briefly, AL1 cells were grown at a constant temperature of 25℃ with 2% of CO_2 in 10%FCS/AMEM (70% volume of DMEM, 10% volume of Fetal Cow Serum, 100 U of Penicillin-Streptomycin, Glutamin, Insulin) in gelatin-coated flasks.

In situ Hi-C library preparation

Cell preparations from d/d animal embryo

3cm (nose to tail) *d/d* embryos were kept under starving conditions for 3 days to avoid genomic DNA contamination from food. After starving, intestines and guts were removed, the remaining parts were dissected from 10 embryos and washed with 80% diluted PBS (APBS) under the anesthetic conditions. The dissected pieces were treated with Liberase TM (1:100 dilution Unit) for 30 min with rotation. Cells were spun down, resuspended into 10%FCS/AMEM. 1x 10^6 cells were resuspended into 36ml of 10%FCS/AMEM.

Cell preparations from AL1 culture

Culture cells were treated with 1% Trypsin at room temperature for 3 minutes, then Trypsin was quenched by addition of 10%FCS/AMEM. Cells were spun down and counted. 1x 10^6 cells were resuspended into 36ml of 10%FCS/AMEM.

Cell preparations from AL1 cell line during mitotic phase

For arresting AL1 cells in mitotic phase, AL1 cells were treated with Thymidine-nocodazole under culture conditions (3). The condition of two-step drug treatment were tested by analysis of Propidium iodide staining intensity. In brief, AL1 cells were passaged into flasks at 30% confluency and cultured for 5 days in normal culture conditions. To arrest in mitotic phase, the culture medium was replaced by 10%FCS/AMEM with supplemental Thymidine (2 mM f. c. (SIGMA)), and the medium was changed every day during Thymidine treatment. After 72 hours of Thymidine treatment, the medium was replaced by 10%FCS/AMEM supplemental Nocodazole (4 nM f. c. (SIGMA)) and cultured for 18 hours. After release from nocodazole treatment, the cell cycle arrested cells were detached by shaking and the supernatant was collected. The cells in the supernatant were spun down (200 x g for 10 min at room temperature) and resuspend in 10%FCS/AMEM. 5x 10^5 cells were resuspended into 36ml of 10%FCS/AMEM for Hi-C library preparation.

Hi-C Library preparation

All Hi-C libraries from this paper were generated as described in(4) (5), with the following modifications. After harvesting the cells in 10% FCS/AMEM they were fixed with formaldehyde (1% f. c. (Merck)) for 10 minutes at room temperature. The fixation was stopped by adding ice-cold glycine (125 mM f. c. (SIGMA)). Fixed cells from embryonic samples were strained with a 100 µm cell strainer to generate single cells, collected by centrifugation (200 x g for 10 min at 4℃), and washed with ice-cold PBS twice (200 x g for 10 min at 4℃). 1 × 10^6 cells were collected and incubated in ice-cold lysis buffer (10 mM Tris–HCl pH 8, 10 mM NaCl, 0.2% Igepal CA-630, 1% Triton-X100, protease inhibitor cocktail EDTA free) with occasional agitation for 30 min on ice. After centrifugation to pellet the cell nuclei (250 x g for 10 min at 4°C), nuclei were washed once with 1.25x NEBuffer2 (with HindIII library) or NEBbuffer3 (with DpnII library) (NEB). The nuclei were resuspended in 1.25x NEBuffer2 or NEBbuffer3, SDS (0.6% f. c.) was added, and the mixture was incubated with agitation (950 rpm for 2 h at 37℃). Triton X-100 (3.3 % f. c.) was added to quench SDS, and the nuclei were incubated with agitation (950 rpm for 2 h at 37℃). Restriction digest with HindIII or DnpII (in 1x NEB2.1 or 1x DnpII buffer from NEB; 2,000 U per 0.25 million cells) was performed overnight with agitation (950 rpm at 37℃). Using biotin-14-dATP (Thermo Fisher), dCTP, dGTP, and dTTP, the restriction sites were filled with Klenow (50 U per 0.5 million cells) for 1 h at 37°C with repeated agitation (700rpm 10 sec and rest 30 sec for 1 h in a thermal cycler). Ligation was performed overnight at 18℃ (2000 U of T4 DNA ligase). After ligation, crosslinking was reversed by incubation with proteinase K in SDS buffer (50 mM Tris-HCl pH8.0, 1% SDS, 10 mM EDTA) overnight at 65°C. An additional proteinase K incubation at 65°C for 2 h was followed with RNase A treatment and two sequential phenol/chloroform extractions. After DNA precipitation, the DNA was spun down (centrifugation with max-speed for 30 min at 4°C). The pellets were resuspended in 20 µl TE and the DNA concentration was determined using Qubit 2 device (Thermo Fisher). 20 µg of Biotinylated DNA was used for the library preparation. Biotin from non-ligated fragment ends was removed with T4 DNA polymerase (NEB) for 30 min at 37°C and EDTA was added to stop the reaction (10 mM f. c.). DNA was sonicated using the Covaris system to generate DNA fragments with a size peak around 400 bp (Covaris S2 settings: duty factor: 10%; peak incident power: 5 ; cycles per burst: 200; time: 60 sec). After end repair with T4 DNA polymerase and Klenow Large fragment (both from NEB) and T4 DNA polynucleotide kinase (in-home) in the presence of dNTPs in T4 DNA ligation buffer (for 30 min at room temperature), the DNA was purified (QIAGEN mini purification kit). A double-size selection using DNA purification beads (in-home) was performed: First, the ratio of the DNA purification bead (in-home) volume to DNA sample volume was adjusted to 0.6:1. After incubation for 15 min at room temperature, the sample was transferred to a magnetic stand, the supernatant was transferred to a new Eppendorf tube, while the beads were discarded. The ratio of the DNA purification beads solution volume to DNA sample volume was then adjusted to 0.9:1 final. After incubation for 15 min at room temperature, the sample was transferred to a magnetic stand. Following two washes with 80% ethanol, the DNA was eluted in Elution buffer (QIAGEN). Biotinylated ligation products were isolated using pre-washed MyOne Streptavidin C1 Dynabeads (Life Technologies) on a magnet stand in binding buffer (5 mM Tris pH8, 0.5 mM EDTA, 1 M NaCl) for 30 min at room temperature. dA-tailing was carried on beads: dATP was added with Klenow exo- (for 1h at 37°C, NEB), then the enzyme was heat-inactivated (20 min at 65°C). After two washes in binding buffer and one wash in T4 DNA ligation buffer, Illumina adapters were ligated onto Hi-C ligation products bound to streptavidin beads in T4 DNA ligase with slowly rotation (for 2 h at room temperature). After washing twice with wash buffer (5 mM Tris pH 8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) and then once with binding buffer, the DNA-bound beads were resuspended in a final volume of 20 µl 1x NEBuffer2. Captured biotinylated Hi-C DNA was amplified with PCR amplification cycles (with NEBNext® High-Fidelity 2X PCR Master Mix (NEB)). After PCR amplification, the Hi-C libraries were purified with DNA purification beads. The concentration of the Hi-C library was determined using fragment analyzer and qPCR, and the Hi-C libraries were paired-end sequenced at VBCF NGS.

<u>Hi-C data pre-processing</u>

For scaffolding, we generated 16 Hi-C datasets (SAMN15470432- SAMN15470436) from AL1 cells and embryo. The quality of the Hi-C reads was checked using FastQC(6) (v0.11.8). Low quality bases at the beginning and the end of the reads as well as the retained Illumina adapter sequences were clipped using Trimmomatic(7) (v0.39) with the following settings:

```
ILLUMINACLIP:Illumina.fa:2:20:10 LEADING:20 TRAILING:15
SLIDINGWINDOW:4:20 MINLEN:100
```

The processed read pairs were mapped and filtered using the HiCUP pipeline(8)(v0.7.1). In order to increase the number of valid pairs, mapped reads with MAPQ>=30 and the difference between AS and XS tags more than 10 were considered unique mappers rather than multimappers.

Genome Scaffolding

The high level of repetitiveness of large genomes poses a challenge for Hi-C based scaffolding. Contact depletion in repetitive regions coupled with sequence biases disturbing the assumed uniform spread of digest sites throughout the genome result in biases that lead to suboptimal scaffolding. We developed an agglomerative hierarchical clustering-based scaffolding approach utilizing various normalization techniques to overcome these hurdles while also achieving high performance.

The scaffolder can be used in two ways. Fully automated scaffolding of a list of contigs and their associated contacts. Guided scaffolding, which allows forcing a global structure onto a scaffold, e.g. reusing an existing scaffolding, but still allowing for local reorderings of the contigs. This can be used for a final polishing pass on the scaffolds or after contigs have been assigned to an existing scaffold and placed close to their assumed position in the scaffold.

Normalization encompasses contact repositioning, length normalization and midpoint computation. Contact repositioning entails the adjustment of the contact positions to the nearest digest site. For length normalization we employ two strategies resulting in two measures for the effective sequence length. One dividing the sequence in equally sized bins, with the counts of the bins corresponding to the number of contacts contained in them. The effective sequence length corresponds to the number of non-zero bins multiplied by the bin size. The other measure uses the number of active digest sites (i.e. sites present in the sequence with an actual contact mapping to it) as a proxy for sequence length. The midpoint, i.e. the middle of the contig, is used to determine the orientation (reverse complement or not) in relation to other contigs. Here, contact depletion can effectively result in the breakdown of the assumption that the midpoint corresponds to the length divided by two. To illustrate, imagine a contig, in which the left half is depleted of repetitive elements, whereas the right half is not. Most contacts will therefore be located in the left half. If the midpoint is not adjusted, this would result in the contig preferentially being attached to others at its left side.

The high-level structure of the scaffolding algorithm is as follows:

Pre-processing

a. Reading of contig sequence lengths from provided fasta or fasta index
b. Reading of the location of the digest sites
c. Reading of the contacts and remapping to the closest digest site
d. Effective sequence length computation based on binning
e. Midpoint computation
f. Effective length computation based on digest sites
g. Build initial scaffold graph (node = contigs, edges = contacts)

Scaffolding (repeated until no more nodes can be joined)

a. Node scoring – for each node compute the most likely node to the left and right
b. Nodes having each other as most likely neighboring nodes are joined into paths. Paths effectively represent a node in the graph and subsume their contained nodes.
c. Compute effective lengths and midpoints for new nodes (ie. paths)

Post-processing

a. Write final set of paths (i.e. scaffolds)

We created initial clusters based on the linkage map from Smith et al(9). We then added contigs based on unique assignability to the clusters. This was followed by scaffolding the cluster separately, visual inspection of an approximate contact map and return of wrongly assigned contigs to the set of unassigned contigs. We created contact maps for all clusters (effectively a whole genome contact map) and merged or split clusters based on the signal within those. The process of assigning contigs, scaffolding, merging and splitting clusters was repeated until no more useful changes could be made to the clusters.

Genome Error Correction

Due to relatively high residual error rate in the assembled contigs, it was necessary to perform error correction prior to scaffolding. For this purpose, we generated 5 datasets (2,866,891,596 read pairs, 965Gb, 30x coverage of the genome) using the DNA from the same animal that was used to generate the data for the genome assembly(10).

Library preparation for Illumina sequencing

The library preparation was performed using the Westburg NGS DNA library kit (Westburg). The final library was excised with the Pippin prep with 400bp DNA size. The DNA library was paired-end sequenced at VBCF NGS.

The quality of the reads was checked using FastQC, low quality bases at either end of each read were clipped as well as retained Illumina sequencing adapters were clipped with Trimmomatic (v0.39).

```
ILLUMINACLIP:Illumina.fa:2:30:10 LEADING:30 TRAILING:20
SLIDINGWINDOW:4:30 MINLEN:50
```

The DNA reads were aligned to the genome using bowtie2(11) (v.2.3.4.2) with the parameter '--very-sensitive'. Afterwards, the mapped reads were used to correct polymorphisms, indels, small gaps and local mis-assemblies using Pilon(12) (v1.23).

In the second round of correction, RNA-seq reads (SAMN06564480-SAMN06564500 and SAMN10869061- SAMN10869072) were mapped to the corrected reference using hisat2(13) (v2.1.0) run with the options '--fr --rna-strandness RF'. Since the reads originated from multiple individuals, only indels and small gaps were fixed, but not polymorphisms.

Annotation

The scaffolded genome was annotated using the bottom-up approach, in which the RNA-seq data was mapped to the genome to generate gene models. Two different sources of RNA-seq data were used: stranded (SAMN06564480-SAMN06564500 and SAMN10869061 - SAMN10869072) and unstranded short Illumina reads (PRJNA300706), and long IsoSeq reads generated using the Lexogen Teloprime kit to ensure that only full-length transcripts were sequenced. The samples were selected to cover as many different tissues and developmental or regeneration stages as possible.

Illumina reads were mapped to the genome using hisat2(13). Afterwards, StringTie(14) (v2.1.1) was used to reconstruct the gene models. Next, both assembled transcripts and long PacBio reads were aligned to the genome using GMAP(15) (v2019-09-12) to generate a combined set of gene models. The resulting gene models were further merged using StringTie --merge.

The IsoSeq reads are more likely to represent correct full-length transcripts than the transcripts assembled from short Illumina reads. In turn, stranded RNA-seq data are more likely to generate correct gene models than unstranded RNA-seq data in case of overlapping transcripts. In order to determine, which sequence a particular isoform was based on, we additionally introduced a prefix system for isoform names (but not gene names). Thus, a gene AMEX60DDxxxxx can have isoforms AMEX60DDpppxxxxx.1, AMEX60DDpppxxxxx.2 and AMEX60DDpppxxxxx.3, where ppp is a three-digit prefix and xxxxx is the gene ID. Possible values of ppp are listed in Supplementary table 1. Note that the prefix only appears in the isoform names and all isoforms have the same xxxxx part.

Prefix	Meaning
101 102 103	The isoform is based on one or more IsoSeq reads
201	The isoform was reconstructed from stranded RNA-seq data
301	The isoform was reconstructed from unstranded RNA-seq data

Supplementary table 1: Isoform prefixes and their meaning

The final gene models were annotated using a custom annotation pipeline. In brief, the transcripts were BLASTed(16) (blastx, e-value cut-off 1e-10) against the nr database limited to vertebrate GIs and against the nr database limited to human GIs only. Best homologs from both searches were kept and used to annotate the transcripts. In case gene symbols from both runs matched, the gene symbol was kept. Otherwise, the gene was annotated using the following schema: 'Symbol [nr]|Symbol [hs]'.
Additionally, open reading frames (ORFs) were predicted by extending the homologous alignment to the left until the left most start codon (if possible) and to the right until the first in-frame terminal codon (if present). The resulting ORFs were split into 4 broad categories:

1. Putative – full-length ORFs based on homologous alignment.
2. C-terminal, N-terminal, partial – ORFs based on homologous alignment, but lacking the left-most start codon, the in-frame terminal codon or both, respectively.
3. Enforced – the homologous alignment is present but interrupted by terminal codons. Those transcripts are likely to represent pseudogenes.
4. Predicted – the transcripts that did not have a homolog but had an ORF longer than 150 amino acids.

Synteny analysis

We downloaded the set of gar proteins (from ftp://ftp.ensembl.org/pub/release-102/fasta/lepisosteus_oculatus/pep/Lepisosteus_oculatus.LepOcu1.pep.all.fa.gz) and extracted those that were annotated as protein-coding. Then we downloaded the list of annotated genes (ftp://ftp.ensembl.org/pub/release-102/gff3/lepisosteus_oculatus/Lepisosteus_oculatus.LepOcu1.102.gff3.gz). Then we extracted protein-coding genes from the axolotl genome and ran a two-way blastp with default settings – axolotl against gar and vice versa. Finally, we identified a subset of 11,677 mutually best hits in both runs and plotted those genes such that the coordinates in the axolotl are on the X axis, while the coordinates in the gar are on the Y axis.

The same principle was used to visualize the synteny between the axolotl and the chordate linkage groups (CLG, Simakov et al, NEE, 2020). The CLG proteins (obtained from Oleg Simakov) were blasted against the axolotl proteins and vice versa. Finally, 4,938 mutually best hits were plotted, while different CLGs were colored according to the colors in Simakov et al, 2020. The Jupyter notebook as well as any additional scripts that were used for this analysis are available at https://github.com/labtanaka/schloissnig_axolotl .

TAD prediction

The same Hi-C reads that were previously used for scaffolding were mapped to the scaffolds using HiCUP pipeline(8). Due to the repetitiveness of the axolotl genome, the default exclusion criteria of the pipeline were slightly relaxed as described in **Hi-C data pre-processing**.

The TADs were predicted using Homer(17) (v4.10). In brief, forward and reverse reads were extracted from the SAM output file of the HiCUP pipeline and makeTagDirectory was used to generate the input for findTADsAndLoops.pl. The script was run for 4 different resolutions as follows:

```
for RES in 5 20 50 100; do RES_KB=$((RES * 1000)); WND=$((RES *
3 * 1000)); findTADsAndLoops.pl find . -cpu 24 -res ${RES_KB} -
window $WND -keepOverlappingTADs -minTADscore 20000000; done
```

Due to the size of the TADs it is important to keep overlapping TADs. In order to generate the final set of TADs, the TADs generated at different resolutions were merged using merge2Dbed.pl, which is shipped with Homer.

TAD size comparisons

In order to assess the size difference between the TADs in axolotl and human, TADs were predicted in both organisms using the same set of parameters described in **TAD prediction**. The human gene models were downloaded from Ensebl (ftp://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/Homo_sapiens.GRCh37.87.chr.gtf.gz).

The axolotl gene models were annotated as described in **Genome annotation**. In order to identify strongly homologous TADs, only putative Axolotl gene models were used. For each human gene, we first found the smallest TAD that contained that gene. Next, we identified all other genes contained in the same TAD. Finally, we found homologous axolotl genes based on the gene symbol and identified the smallest axolotl TAD that contained all or most of those genes.

MHC cluster annotation and synteny analysis

To annotate the axolotl MHC, we first aligned the human MHC(18) (which is delimited from GABBR1 to KIFC1) obtained from the ENSEMBL(19) (Release 100) to the axolotl transcriptome using BLAST and the e-value threshold 1e-10. Once we identified the region where the axolotl MHC was likely to be, we aligned all the transcripts obtained from the locus to the human proteome obtained from ENSEMBL using BLAST and the e-value threshold 1e-10.
To classify some of the proteins whose families have expanded over large distances (MR1 and TRIM related genes), we obtained a set of orthologous protein sequences for *Anolis carolinensis*, *Mus musculus*, *Gallus gallus*, *Homo sapiens*, *Lepisosteus oculatus* and *Danio rerio* from ENSEMBL (Release 100) through the BioMart tool. The sequences for *Xenopus tropicalis(20)*, *Nanorana parkeri*(21) and *Lithobates catesbeianus*(22) were obtained from Xenbase (assembly JGI-10.0), and NCBI assemblies GCF_000935625.1 and GCA_002284835.2, respectively. We then used the human proteins as query for BLAST searches with the e-value threshold 1e-10. After this, we aligned the HLA nucleotide sequences using MACSE (v2.03) aligner(23), while using MAFFT(24) (v7.427) to align the TRIM39 amino acid sequences.

Phylogenetic trees were generated using IQTree(25) (v1.6.12). For the HLA protein family tree, we annotated the protein domains using HMMER(26) (v3.2.1) and Pfam(27) (v33.1) as the database, with the e-value threshold 1e-5. After we identified the Ig and the MHC-I domains for all the proteins, we partitioned the analysis for the protein alignment. Finally, we constructed the phylogenetic tree using a fast-bootstrap of 1000 replicates and the ModelFinder Plus mode to find the most appropriate substitution model.

For the tripartite motif (TRIM) family, we used the protein alignment produced by MAFFT, however, we did not make a search for protein domains due to the TRIM protein family being very structurally diverse. Thus, we created a phylogenetic tree using a fast-bootstrap of 1000 replicates and the ModelFinder Plus mode to find the most appropriate substitution model.

Identification of axolotl *Fgf8* conserved non-coding elements

multiVISTA software (http://www-gsd.lbl.gov/vista/) was used to test the axolotl *Fgf8* genomic sequence. All CNEs of axolotl *Fgf8* were tested based upon their conservation score with the reported human and mouse highly conserved sequences(28). mVISTA was required to have at least 70% identity over a 100-bp window. The tested genomic sequences surrounding *Axfgf8* were in the range of chr8q: 630,000,000-660,000,000. The identified genomic coordinates of axolotl *Fgf8* CNEs are listed in Dataset S5.

Generation of transgenic axolotls

Plasmid construction

The plasmid IS-axCNE80-bEGFP was generated by standard PCR procedures and cloning into the IS-bEGFP vector, which harbors the chicken beta-actin minimal promoter upstream of the coding sequence. The genomic sequence of axCNE80 was tested by mVISTA program (As explained in 9. "Identification of axolotl *Fgf8* conserved non-cording elements") and located at chr8q:641,235,321-641,235,453 (v.6.0-DD assembly). The axCNE80 DNA fragment was amplified by 2xQ5 PCR mix

(NEB) with the following PCR primers: forward 5’-GGGgcggccgcATAATAGCGGGTAACGAGAAGTTCTCTC and reverse 5’-GGGtctagaTTAACGGTTCTTAAATGTCAAGACGGACC), and cloned into the IS-bEGFP vector between the NotI and XbaI sites.

I-SceI-mediated transgenesis

IS-axCNE80-bEGFP (AxCNE80-GFP) axolotls were generated by I-SceI meganuclease-mediated transgenesis, according to previously described methods (29). Briefly: de-jellied one-cell stage d/d axolotl eggs were each injected with ~5nl of injection mix containing 0.5ng IS-axCNE80-bEGFP and 0.005U I-SceI meganuclease (NEB) diluted in 1X CutSmart buffer (NEB). Injected axolotl eggs were maintained in 0.1X MMR/tap water at room temperature until initiation of limb bud outgrowth (~15 days after egg laying). Transgenic animals were identified by screening for EGFP fluorescence in the early limb bud using a Zeiss AXIOzoom V16 stereo microscope. Axolotls were anaesthetized in benzocaine (Sigma) diluted in tap water prior to imaging. F0 transgenic animals were used for IS-axCNE80-bEGFP experiments. A total of 7 animals showed similar EGFP expression patterns to that shown in Fig. 4G.

Image acquisition

Images were acquired using a Zeiss AXIOzoom V16 stereo microscope and ZEN software (blue edition). Images were adjusted for brightness and contrast using Fiji (30) and Adobe Photoshop software.

HCR *in situ* hybridization for *Axfgf8* mRNA

Whole mount HCR fluorescent *in situ* hybridization was performed for axolotl limb buds according to previously published methods (31). The following reagents were purchased from Molecular Instruments: *Fgf8* probes, hybridization buffer, wash buffer, amplification buffer, HCR hairpins. *Fgf8* probes were coupled with B1 initiator sequences. B1-Alexa_546 h1 and h2 hairpins were used for signal amplification.

Larvae were fixed overnight (14 to 16 hours) in 4% PFA at 4 degrees, then washed in PBS. Limb buds were harvested from the fixed larvae, dehydrated in 100% Methanol and stored in 100% Methanol at -20 degrees until use (2 to 8 weeks).
Prior to staining, samples were progressively rehydrated in 75% Methanol 25% PBST (0.1% v/v Tween 20), 50% Methanol 50% PBST, 25% Methanol 75% PBST and 100% PBST. Samples were pre-hybridized in hybridization buffer for 30 minutes at 37 degrees and subsequently hybridized for 14 to 16 hours at 37 degrees in 500ul of hybridization buffer containing 2 pmol of *Fgf8* probes.

Four stringent washes were performed at 37 degrees for one hour each in wash buffer, samples were then washed twice for 30 minutes in 5x SCCT (0.1% v/v Triton X-100) at room temperature. Samples where then incubated in amplification buffer for 30 minutes at room temperature. H1 and h2 hairpins were snap cooled separately (90 seconds at 95 degrees and 30 minutes at room temperature) before adding them to a new aliquot of amplification buffer. Samples were incubated overnight (14 to16 hours) at room temperature in 500ul of amplification buffer containing 30 pmol of h1 and h2 hairpins. Samples were washed twice for 10 minutes in 5x SCCT, incubated with DAPI (Sigma, #D9542) in 5x SCCT for 40 minutes and washed again 2 times 30 minutes in 5x SCCT.
Samples were mounted for imaging in 35 mm imaging dishes (Ibidi, #81156) in 5x SCCT. Imaging was performed using a Zeiss LSM 880 Axio Observer (inverted) with Airyscan (Airyscan fast mode, excitation wavelength 561nm). Single plane images were adjusted for brightness and contrast using Fiji (30) and Adobe Photoshop software.

References

1. A. L. Groell, D. M. Gardiner, S. V. Bryant, Stability of positional identity of axolotl blastema cells in vitro. *Roux Arch Dev Biol* **202**, 170-175 (1993).
2. S. Roy, D. M. Gardiner, S. V. Bryant, Vaccinia as a tool for functional analysis in regenerating limbs: ectopic expression of Shh. *Dev Biol* **218**, 199-205 (2000).
3. M. L. Whitfield *et al.*, Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell* **13**, 1977-2000 (2002).
4. T. Nagano *et al.*, Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome biology* **16**, 175 (2015).
5. G. Wutz *et al.*, ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. *Elife* **9** (2020).
6. S. Andrews (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
7. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
8. S. Wingett *et al.*, HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310 (2015).
9. J. J. Smith, D. K. Kump, J. A. Walker, D. M. Parichy, S. R. Voss, A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in Ambystoma. *Genetics* **171**, 1161-1171 (2005).
10. S. Nowoshilow *et al.*, The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50-55 (2018).
11. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
12. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
13. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
14. M. Pertea *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290-295 (2015).
15. T. D. Wu, J. Reeder, M. Lawrence, G. Becker, M. J. Brauer, GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods Mol Biol* **1418**, 283-334 (2016).
16. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
17. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
18. T. Shiina, A. Blancher, H. Inoko, J. K. Kulski, Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology* **150**, 127-138 (2017).
19. A. D. Yates *et al.*, Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688 (2020).
20. K. Karimi *et al.*, Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Research* **46**, D861-D868 (2017).

21. Y.-B. Sun *et al.*, Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences* **112**, E1257-E1262 (2015).
22. S. A. Hammond *et al.*, The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat Commun* **8**, 1433 (2017).
23. V. Ranwez, E. J. P. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol Biol Evol* **35**, 2582-2584 (2018).
24. T. Nakamura, K. D. Yamada, K. Tomii, K. Katoh, Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490-2492 (2018).
25. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).
26. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
27. R. D. Finn *et al.*, Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-230 (2014).
28. M. Marinic, T. Aktas, S. Ruf, F. Spitz, An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* **24**, 530-542 (2013).
29. L. Sobkow, H. H. Epperlein, S. Herklotz, W. L. Straube, E. M. Tanaka, A germline GFP transgenic axolotl and its use to track cell fate: dual origin of the fin mesenchyme during development and the fate of blood cells during regeneration. *Developmental biology* **290**, 386-397 (2006).
30. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).
31. H. M. T. Choi *et al.*, Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145** (2018).


Zebrafish_ENSDARG00000016056_mhc1laa_1080
Zebrafish_ENSDARG00000016227_mhc1lba_936
Zebrafish_ENSDARG00000055813_si_dkey-225f5.5_1053
Lithobates_AB205_0068200_mRNA
Nanorana_XM_018555842.1
Lithobates_AB205_0205950_mRNA
Lithobates_AB205_0091950_mRNA
Nanorana_XM_018572510.1
Lithobates_AB205_0144440A_mRNA
Lithobates_AB205_0144440B_mRNA
Lithobates_AB205_0088690_mRNA
Nanorana_XM_018555837.1
Lithobates_AB205_0016830_mRNA
Lithobates_AB205_0062710A_mRNA
Lithobates_AB205_0062710B_mRNA
Nanorana_XM_018555874.1
Nanorana_XM_018558100.1
Lithobates_AB205_0157660_mRNA
Nanorana_XM_018558127.1
Nanorana_XM_018572511.1
Nanorana_XM_018558101.1
Nanorana_XM_018555843.1
Nanorana_XM_018572505.1
Nanorana_XM_018574102.1
Xenopus_rna-NM_001113065.1
Xenopus_rna-XM_002940070.5
Xenopus_rna-XM_012953253.3
Xenopus_rna-XM_012953246.2
Xenopus_rna-NM_001130199.1
Xenopus_rna-XM_018096341.2
Xenopus_rna-XM_031891825.1
Xenopus_rna-XM_031890278.1
Xenopus_rna-XM_002940060.4
Xenopus_rna-NM_001247995.1
Xenopus_rna-NM_001037273.1
Xenopus_rna-XM_004918940.4
Xenopus_rna-XM_004918939.4
Xenopus_rna-XM_002940062.4
Xenopus_rna-XM_031890829.1
Xenopus_rna-XM_002940065.5
Xenopus_rna-XM_031890827.1
Xenopus_rna-XM_031890828.1
Xenopus_rna-XM_031890830.1
Xenopus_rna-XM_031890831.1
Xenopus_rna-NM_001366672.1
Xenopus_rna-XM_002940066.5
Xenopus_rna-XM_031891823.1
Xenopus_rna-XM_031890985.1
Xenopus_rna-XM_031891894.1
Xenopus_rna-XM_031891895.1
Xenopus_rna-XM_018096988.2
Xenopus_rna-XM_002938952.5
Xenopus_rna-XM_031891502.1
Xenopus_rna-XM_018096955.2
Xenopus_rna-XM_031891824.1
Xenopus_rna-XM_004918936.4
Xenopus_rna-XM_018096512.2
Xenopus_rna-XM_031892155.1
Xenopus_rna-XM_002935343.5
Axolotl|MR1-like|chr13q_476316837-476379472
Axolotl|MR1-like|chr13q_476403178-476413061
Axolotl|MR1-like|chr13q_478587715-478600592
Axolotl|MR1-like|chr13q_465011001-465054647
Axolotl|MR1-like|chr13q_465659830-465705170
Axolotl|MR1-like|chr7q_75718428-75718655
Axolotl|MR1-like|chr13q_467822214-467822662
Axolotl|MR1-like|chr13q_467821826-468054267
Axolotl|MR1-like|chr13q_467680271-467680924
Axolotl|MR1-like|chr13q_466443858-466584460
Axolotl|MR1-like|chr13q_464566543-464567456
Axolotl|MR1-like|chr13q_463231620-463407721
Mouse_ENSMUSG00000073411_H2-D1_1143
Mouse_ENSMUSG00000061232_H2-K1_1110
Mouse_ENSMUSG00000035929_H2-Q4_1065
Mouse_ENSMUSG00000067235_H2-Q10_978
Mouse_ENSMUSG00000073406_H2-Bl_1050
Mouse_ENSMUSG00000073402_Gm8909_1191
Mouse_ENSMUSG00000079507_H2-Q1_1107
Mouse_ENSMUSG00000091705_H2-Q2_1083
Mouse_ENSMUSG00000073409_H2-Q6_981
Mouse_ENSMUSG00000060550_H2-Q7_1005
Mouse_ENSMUSG00000067212_H2-T23_1074
Mouse_ENSMUSG00000079492_Gm11127_1125
Mouse_ENSMUSG00000092243_Gm7030_858
Mouse_ENSMUSG00000054128_H2-T3_1155
Mouse_ENSMUSG00000079491_H2-T10_1140
Mouse_ENSMUSG00000056116_H2-T22_1140
Mouse_ENSMUSG00000067201_H2-M9_1029
Mouse_ENSMUSG00000037334_H2-M1_1035
Mouse_ENSMUSG00000037537_H2-M11_1006
Mouse_ENSMUSG00000048231_H2-M10.4_996
Mouse_ENSMUSG00000037130_H2-M10.6_993
Mouse_ENSMUSG00000023083_H2-M10.2_990
Mouse_ENSMUSG00000024448_H2-M10.1_996
Mouse_ENSMUSG00000037246_H2-M10.5_993
Mouse_ENSMUSG00000058124_H2-M10.3_993
Mouse_ENSMUSG00000016206_H2-M3_1011
Mouse_ENSMUSG00000016283_H2-M2_1032
Human_ENSG00000204642_HLA-F_1329
Human_ENSG00000204632_HLA-G_1032
Human_ENSG00000204525_HLA-C_1119
Human_ENSG00000234745_HLA-B_1089
Human_ENSG00000206503_HLA-A_1116
Human_ENSG00000204592_HLA-E_1077
Mouse_ENSMUSG00000024459_H2-M5_1047
Mouse_ENSMUSG00000064308_2410137M14Rik_489
Mouse_ENSMUSG00000053835_H2-T24_1092
Mouse_ENSMUSG00000054005_Mill1_1188
Mouse_ENSMUSG00000040987_Mill2_1068
Human_ENSG00000204516_MICB_1152
Human_ENSG00000204520_MICA_999
Mouse_ENSMUSG00000079685_Ulbp1_1005
Human_ENSG00000131015_ULBP2_741
Human_ENSG00000155918_RAET1L_741
Human_ENSG00000203722_RAET1G_1005
Human_ENSG00000131019_ULBP3_735
Mouse_ENSMUSG00000053219_Raet1e_756
Mouse_ENSMUSG00000078452_Raet1d_750
Human_ENSG00000164520_RAET1E_792
Human_ENSG00000285991_AL355312.6_765
Mouse_ENSMUSG00000006611_Hfe_1080
Human_ENSG00000010704_HFE_1047
Human_ENSG00000104870_FCGRT_1098
Mouse_ENSMUSG00000003420_Fcgrt_1110
Mouse_ENSMUSG00000037053_Azgp1_924
Human_ENSG00000160862_AZGP1_897
Anolis_ENSACAG00000011070_NoName_1092
Anolis_ENSACAG00000006478_NoName_1074
Anolis_ENSACAG00000024528_NoName_957
Anolis_ENSACAG00000026915_NoName_318
Anolis_ENSACAG00000006674_NoName_1086
Anolis_ENSACAG00000013947_NoName_465
Anolis_ENSACAG00000017226_NoName_1077
Anolis_ENSACAG00000024939_NoName_345
Anolis_ENSACAG00000000562_NoName_963
Anolis_ENSACAG00000022634_NoName_1080
Anolis_ENSACAG00000027871_NoName_1068
Anolis_ENSACAG00000023667_NoName_1080
Anolis_ENSACAG00000025539_NoName_342
Anolis_ENSACAG00000024337_NoName_411
Anolis_ENSACAG00000022657_NoName_1017
Anolis_ENSACAG00000003088_NoName_783
Anolis_ENSACAG00000024309_NoName_609
Anolis_ENSACAG00000024081_NoName_567
Anolis_ENSACAG00000025947_NoName_567
Anolis_ENSACAG00000000468_NoName_906
Anolis_ENSACAG00000006471_NoName_936
Anolis_ENSACAG00000024638_NoName_1041
Anolis_ENSACAG00000011968_NoName_963
Anolis_ENSACAG00000006498_NoName_1026
Anolis_ENSACAG00000029516_NoName_1092
Anolis_ENSACAG00000003922_NoName_933
Mouse_ENSMUSG00000026471_Mr1_1026
Human_ENSG00000153029_MR1_1026
Nanorana_XM_018555836.1
Nanorana_XM_018555841.1
Nanorana_XM_018569386.1
Nanorana_XM_018555840.1
Lithobates_AB205_0118370_mRNA
Lithobates_AB205_0192570_mRNA
Lithobates_AB205_0082500A_mRNA
Nanorana_XM_018555838.1
Human_ENSG00000158481_CD1C_1002
Human_ENSG00000158485_CD1B_1002
Human_ENSG00000158477_CD1A_984
Human_ENSG00000158488_CD1E_1167
Human_ENSG00000158473_CD1D_1008
Mouse_ENSMUSG00000028076_Cd1d1_1011
Mouse_ENSMUSG00000041750_Cd1d2_1010
Human_ENSGALG00000043052_CD1B_1047
Human_ENSGALG00000030270_CD1C_1125
Anolis_ENSACAG00000011395_NoName_1062
Nanorana_XM_018558121.1
Zebrafish_ENSDARG00000092731_mhc1uka_1191
Zebrafish_ENSDARG00000075963_mhc1uba_1047
Zebrafish_ENSDARG00000059039_FO904903.1_1035
Zebrafish_ENSDARG00000039164_mhc1uma_1029
Gar_ENSLOCG00000000215_NoName_1059
Gar_ENSLOCG00000000760_NoName_414
Gar_ENSLOCG00000007238_NoName_981
Gar_ENSLOCG00000000807_NoName_837
Nanorana_XM_018558103.1
Gar_ENSLOCG00000000734_NoName_810
Gar_ENSLOCG00000008666_NoName_978
Zebrafish_ENSDARG00000074765_mhc1zja_1152
Zebrafish_ENSDARG00000104293_mhc1zka_1149
Zebrafish_ENSDARG00000092162_mhc1zaa_1076
Zebrafish_ENSDARG00000088022_mhc1zfa_1116
Zebrafish_ENSDARG00000001470_mhc1zea_1053
Zebrafish_ENSDARG00000069471_mhc1zca_1095
Zebrafish_ENSDARG00000036588_mhc1zba_1197
Gar_ENSLOCG00000000416_NoName_1113
Gar_ENSLOCG00000000528_NoName_603
Gar_ENSLOCG00000000335_NoName_1035
Zebrafish_ENSDARG00000051710_CR339041.1_1098
Zebrafish_ENSDARG00000097275_si_ch211-147g22.4_651
Zebrafish_ENSDARG00000051712_mhc1lfa_1122
Zebrafish_ENSDARG00000051713_mhc1lga_1089
Zebrafish_ENSDARG00000097766_mhc1lia_657
Zebrafish_ENSDARG00000051711_CR339041.2_942
Zebrafish_ENSDARG00000023203_mhc1lda_999
Zebrafish_ENSDARG00000096940_si_dkey-52p2.5_1041
Zebrafish_ENSDARG00000096977_mhc1lja_1041
Zebrafish_ENSDARG00000096830_mhc1lka_1074
Zebrafish_ENSDARG00000115781_CR339041.3_765
0.9