Correction: 13 February 2014

www.sciencemag.org/content/science.1248636/DC1

Supplementary Material for

Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors

Andrea Pauli,* Megan L. Norris, Eivind Valen, Guo-Liang Chew, James A. Gagnon, Steven Zimmerman, Andrew Mitchell, Jiao Ma, Julien Dubrulle, Deepak Reyon, Shengdar Q. Tsai, J. Keith Joung, Alan Saghatelian, Alexander F. Schier*

*Corresponding author. E-mail: pauli@fas.harvard.edu (A.P.); schier@fas.harvard.edu (A.F.S.)

Published 9 January 2014 on *Science* Express
DOI: 10.1126/science.1248636

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/science.1248636/DC1)

Movies S1 to S6
Table S1 (Excel file) and Data files S1 (.bed file) and S2 (.fa file) as zipped archives

Correction: On page 2, under "Genotyping of toddler mutants," the wrong size was listed for one of the bands; it has been corrected to 318 bp. Reference citations were also updated.

Materials and Methods

Identification of novel protein-coding genes and secreted peptides expressed during embryogenesis

To identify novel embryonically expressed protein-coding transcripts, RNA-Seq data (*8*), lncRNA annotations (*9*) and ribosome profiling data obtained during early zebrafish embryogenesis (from 2-4 cell to 5 days post fertilization) (*7*) was mined for the presence of novel protein-coding open reading frames. To filter out annotated protein-coding genes, transcripts overlapping in sense with known protein-coding transcripts in zebrafish (RefSeq, Ensembl) and other organisms (xenoRefSeq) were discarded. To identify novel protein-coding open reading frames (ORFs) in the remaining transcripts, all possible ATG-Stop ORFs were extracted and classified by TOC (Translated ORF Classifier; (*7*)), using the published ribosome profiling developmental time course (*7*). To improve de-novo identification of protein-coding ORFs, the feature "cover" was included as additional classification criteria. "Cover" represents the number of bases within an ORF that contain single-nucleotide mapped ribosome profiling reads (RPFs) (see (*7*) for a description of mapping RPF reads at single base-pair resolution).

To be classified as protein-coding, the following criteria had to be fulfilled: The ORF had to be classified as "coding" by TOC, and it had to contain at least 10 nucleotides hit by single-nucleotide resolution mapped RPFs ("cover" ≥ 10). To increase the stringency of the classification pipeline, any coding ORF additionally had to fulfill one of the following three criteria: a) at least 25% of the nucleotides in the ORF had to be hit by a single-nucleotide mapped RPF; or b) at least 30 nucleotides in the ORF had to be hit by a single-nucleotide mapped RPF ("cover" ≥ 30); or c) the PhyloCSF score for this transcript had to be greater than 100 (PhyloCSF ≥ 100). After manual curation of the resultant candidate list (e.g. to exclude artifacts (e.g. genome assembly errors), incomplete isoforms and ambiguous transcript structures), 399 loci (700 transcripts) were retained as confidently identified novel protein-coding genes. Transcript conservation was assessed by calculating the fraction of alignment between a given zebrafish transcript and UCSC Multiz Sequence Alignments (Danio rerio, Fugu, Medaka, Stickleback, Tetraodon, Xenopus, mouse and human).

To identify putative novel embryonic signaling proteins, the amino acid sequences of the resultant protein-coding genes were interrogated for the presence of a predicted signal peptide sequence (SignalP 4.0 (*57*)) and absence of predicted transmembrane domains (Phobius (*58*, *59*)). This analysis identified 28 putative novel embryonic signaling molecules, including Toddler. Note that the *toddler* locus had also been predicted to be protein-coding by our previous purely computational pipeline (*8*) based on its high level of amino acid sequence conservation (PhyloCSF score (*60*) of 98, which was above the applied cutoff for lncRNAs (PhyloCSF < 20)).

TALEN-mediated mutagenesis of *toddler*

TALE repeat arrays targeting the *toddler* gene as described in Figure S5 were constructed using the fast ligation-based automatable solid-phase high-throughput (FLASH) system and cloned into corresponding TALEN expression constructs harboring a wild-type FokI cleavage domain (http://www.addgene.org/talengineering/expressionvectors/) as previously described (*12*). Resulting plasmids were linearized with PmeI and transcribed, polyadenylated and purified with mMessage Machine T7 Ultra kit (Ambion) according to manufacturer's instructions. 300 pg of the TALEN mRNA pair was injected into the cell of 1-cell-stage zebrafish zygotes. Somatic mutation rate was measured from a pool of 10 injected embryos at 30 hpf using T7 Endonuclease I assay (*12*). To this end, a 510-bp *toddler*-specific PCR amplicon spanning the targeted *toddler* genomic location was generated with the *toddler* genotyping primers toddler_gt_F and toddler_gt_R (primer sequences are listed below). The remaining injected fish were raised to adulthood and outcrossed to TLAB wild-type fish. Clutches of embryos with potential heterozygous carriers of *toddler* mutations were used to identify founders with germline mutations in *toddler* (3-5 pools of 5-10 embryos each were tested by the T7 Endonuclease I assay). Offspring of confirmed founders were raised to adulthood, fin-clipped and genotyped by T7 Endonuclease I assay. The nature of the mutation was determined by sequencing cloned PCR amplicons.

Genotyping of *toddler* mutants

Since all seven TALEN-induced *toddler* mutations destroy an RsaI restriction site (see fig. S5), cleavage patterns resulting from RsaI-digested PCR amplicons were used as a readout for the *toddler* genotype. PCR products (standard PCR conditions, using the primer pair toddler_gt_F and toddler_gt_R (see primer sequences below) and annealing temperature of 55^{o}C) were digested for 2 hours at 37^{o}C with RsaI (NEB), and run on a 2% Agarose gel to determine the size of the resulting DNA fragments. Wild type: two bands of 192 bp and 318 bp; heterozygous *toddler* mutant: three bands of 510 bp (full-length, un-cut product from a *toddler* homozygous chromosome), 192 bp and 318 bp; homozygous *toddler* mutant: single band of 510 bp (full-length, un-cut product). Genotyping was performed on adult fin-clips as well as on individual embryos after *in situ* hybridization experiments to match phenotypes with genotypes of embryos derived from incrosses of *toddler* heterozygous adult fish.

Generation of *toddler* expression constructs

The *toddler* transcript sequence (CDS and UTRs) was PCR amplified from a shield-stage zebrafish cDNA library (primers toddler_F, toddler_R) and cloned into the pSC vector (Strataclone) to generate pSC-*toddler,* containing the full-length *toddler* transcript including 5' and 3'UTRs. The cloned sequence corresponds to the Ensembl *ENSDART00000129853* transcript isoform that is annotated as non-coding RNA in the *danRer7* genome assembly. To generate a Toddler expression construct containing only the *toddler* ORF but no endogenous UTRs, the *toddler* ORF was PCR-amplified from

pSC-*toddler* (primers toddler_ORF_F, toddler_ORF_R), digested with NheI/NcoI and cloned into a modified pCS2+ vector containing a customized multiple cloning site (MCS) upstream of the standard SV40UTR (pCS2+*toddler-SV40UTR*).
To generate heat-inducible Toddler expression vectors and transgenic fish, the NheI/NcoI digested *toddler* ORF PCR product was inserted into a pTol2-based vector downstream of a heat-shock promoter. Subsequently, a *cmlc2::eGFP-SV40UTR* fragment was inserted downstream of the *toddler* ORF-SV40-UTR to serve as visible marker for transgenesis.
To generate frame-shifted *toddler* expression constructs, one extra base was introduced six nucleotides downstream of the *toddler* ATG by site-directed mutagenesis (SDM), using pSC-*toddler* as template and primers toddler_FS_F and toddler_FS_R. *Toddler-eGFP* expression constructs were generated in a two-step process. First, SDM was used to convert the STOP codon of the *toddler* ORF into an AvrII restriction site (primers toddler_AvrII_F, toddler_AvrII_R). This restriction site was used to insert in frame PCR-amplified (primers GFP_noATG_AvrII_F, GFP_NheISpeI_R) and AvrII/SpeI-digested eGFP lacking its own ATG. In the resulting vector pSC-*toddler-eGFP*, *toddler-eGFP* is flanked by the endogenous *toddler* 5' and 3' UTRs. To generate signal peptide cleavage site mutant *toddler-eGFP* expression vectors (pSC-*toddler[A->W]-eGFP*), SDM (primers toddler_AtoW_F, toddler_AtoW_R) was used to mutate the signal peptide cleavage site preceding Alanine (A) to Tryptophan (W), a mutation that is predicted to abolish signal peptide cleavage. Mouse and human *toddler* transcript sequences were PCR amplified from mouse or human ESC cDNA libraries (primers mtoddler_F, mtoddler_R; htoddler_F, htoddler_R) and cloned into the pSC Strataclone vector according to standard protocols. All primer sequences are listed below.

Generation of *apelin, aplnra, aplnrb* and eGFP-tagged *aplnra/b* constructs
Zebrafish Apelin, Aplnra and Aplnrb transcript sequences were PCR amplified from 28 hpf- (Apelin) or shield-stage (Aplnra/b) zebrafish cDNA libraries and cloned into the pSC Strataclone vector according to standard protocols. Primer sequences (apelin_F, apelin_R; aplnra_F, aplnra_R; aplnrb_F, aplnrb_R) are listed below. C-terminally GFP-tagged Aplnra/b expression constructs were generated in a two-step PCR protocol, in which a short linker sequence was inserted in frame between Aplnra/b and eGFP (amino acid sequence of the linker: LGPGLGSG; primers linker_aplnra_R, linker_aplnrb_R, linker_GFP_F). The resultant PCR products (Aplnra/b-linker-eGFP) were inserted as EcoRI/NheI digested fragments into a customized, EcoRI/NheI-digested pTol2-based plasmid.

Primer sequences

toddler_gt_F	ATTTTCTACAGTCCGTTACCTGCAC
toddler_gt_R	GTTCTTCCTCAAGGAGAATCAAGTG
toddler_F	GAATATCGCCCTCTCAAACTTTGAAAAAG
toddler_R	CTTTTAATAAATCAAGATAATAATAAAA

toddler_ORF_F	AAAGCTAGCATGAGATTCTTCCACCCGCT
toddler_ORF_R	AAACCATGGTCAAGGGAAAGGTACTCTGG
toddler_FS_F	GCATCACCATGAGAaTTCTTCCACCCGCTG
toddler_FS_R	CAGCGGGTGGAAGAAtTCTCATGGTGATGC
toddler_AvrII_F	CACTCCAGAGTACCTTTCCCTaGgGGTTTTATGATGCTCCGGGC
toddler_AvrII_R	GCCCGGAGCATCATAAAACCcCtAGGGAAAGGTACTCTGGAGTG
GFP_noATG_AvrII_F	AAACCTAGGGTGAGCAAGGGCGAGGAGCTGT
GFP_NheISpeI_R	AAAACTAGTGCTAGCTTACTTGTACAGCTCGTCCATG
toddler_AtoW_F	GTGCTGGTCCTCATCAGCtggGATAAACATGGTACAAAACACG
toddler_AtoW_R	CGTGTTTTGTACCATGTTTATCccaGCTGATGAGGACCAGCAC
mtoddler_F	TTAAGTTACTGCGTGCCTGAATGGAAA
mtoddler_R	ATGTGCAATAAAATCATTCATGTGTACAAACC
htoddler_F	TGAAGACAGCCACACAGATATTGCACAGAC
htoddler_R	AAAAATGATTCAGGACGGGACCGAAGA
apelin_F	CTCTCCTCTCCCTCCCATCCACAC
apelin_R	AAAGGCATAGAGTTCTGTCAGTCTAGGGGACT
aplnra_F	ACGCGCTTCAGCTTCCAGTGAG
aplnra_R	AAATAGACGTCAGCGAACATAAATTCACACTC
aplnrb_F	CTGTGATCACTTCAAAGATTTTCTTGAAGG
aplnrb_R	TCTTTAATCAAGCTGCAACACTTCTTCTCA
linker_GFP_F	GACCTGGACTCGGATCCGGAGTGAGCAAGGGCGAGGAGC
linker_aplnra_R	TCCGGATCCGAGTCCAGGTCCAAGCACTTTGGTGGCCAGCGACTG
linker_aplnrb_R	TCCGGATCCGAGTCCAGGTCCAAGCACCTTCGTAGCCAGAGAC

Zebrafish husbandry and fish lines

Zebrafish (*Danio rerio*) were maintained according to standard protocols. Embryos were staged according to (*61*).

Heterozygous and homozygous (mRNA-rescued) *toddler* mutants were maintained according to standard protocols. *toddler* mutant lines carrying the endodermal reporter tg:*sox17::eGFP* (*14*) were generated by crossing *toddler* heterozygous mutant and tg:*sox17::eGFP* fish. To generate *toddler* mutant lines expressing heat-inducible Toddler, wild-type TLAB embryos were injected with Tol2 transposase and the pTol2-*hs::toddler; cmlc2::eGFP* construct (see above). Embryos were raised to adulthood and screened for germline transmission of the transgene after mating putative founders with *toddler* heterozygous mutant adults and screening for green fluorescent hearts at 36 hpf. GFP-positive embryos were raised to adulthood and genotyped for the presence of the *toddler* mutation.

mRNA synthesis and microinjection

Capped mRNAs of *toddler, apelin, aplnrb, receptor-eGFP and human H2B-RFP* were synthesized with the T7, T3 or Sp6 mMessage Machine kit (Ambion) according to the manufacturer's protocol, using linearized plasmids as templates. Plasmids encoding

Cxcr4b-eGFP and *Cxcr7b-eGFP* had been described before (*62*). If not mentioned otherwise, all mRNAs were injected into embryos at the one-cell stage.

Rescue and overexpression of *toddler*

mRNAs derived from pSC-*toddler* (30 pg mRNA injection for rescue) and pSC2+*toddler* (2 pg mRNA injection for rescue) can both rescue *toddler* null mutants at the indicated concentration. All *toddler* rescue and overexpression experiments shown in this manuscript were performed with mRNA transcribed from the pCS2+*toddler* plasmid (*toddlerORF-SV40UTR*), and listed concentrations for rescue and overexpression of Toddler (10-50 pg mRNA injection for overexpression studies) refer to this mRNA. Exact amounts of mRNA required for rescue varied slightly between different batches of mRNAs (e.g. see the different effects of 2 pg *toddler* mRNA injections from different batches of mRNA into wild-type embryos in Fig. 2D and Fig. 5D).

Toddler and Apelin C-terminal peptide fragments were synthesized *in vitro* (Peptide2). The Toddler peptide fragment comprised the C-terminal 21 amino acids of zebrafish Toddler (YRRHNCPKKRCLPLHSRVPFP). The Apelin peptide fragment comprised the C-terminal 16 amino acids of zebrafish Apelin (RRRPRPRLSHKGPMPF). For overexpression and Aplnr-eGFP internalization studies, 1-5 ng of peptide was injected extracellularly after the mid-blastula transition.

Rescue of *toddler* mutants by heat-induced expression of Toddler

Heat-shock experiments were performed as described by (*25*). Briefly, adult *toddler* heterozygous mutants carrying the *hs::Toddler; cmlc2::eGFP* transgene were crossed to adult *toddler* homozygous mutants. Batches of embryos were heat-shocked at 37^{o}C for 45 minutes at different times during embryogenesis, starting from 4 hours post fertilization to 10 hours post fertilization. Only heat-shock treatments up to shield stage (~6 hours post fertilization) were able to rescue *toddler* homozygous mutant progeny.

Morpholino knockdown of aplnra/b

Double Morpholino (MO) knockdowns of *aplnra/b* were performed as previously described (*25*, *26*). Briefly, MOs targeting *aplnra* (cggtgtattccggcgttggctccat; Genetools) and *aplnrb* (cagagaagttgtttgtcatgtgctc; Genetools) were injected at the one-cell stage at 1 ng or 0.5 ng, respectively.

Morphological analysis of the *toddler* mutant phenotype

Toddler mutants were allowed to develop at 28^{o}C to the indicated stage. For morphological analyses of stages past 24 hpf, embryos were anesthetized in MESAB and mounted in 2% methylcellulose. Embryos were imaged using an Axio Imager.Z1 microscope (Zeiss). Images were processed in ImageJ (*63*).

Toddler-eGFP secretion assay

To test for Toddler-eGFP secretion, 50 pg of mRNA encoding either Toddler-eGFP or the predicted signal peptide cleavage site mutant Toddler[A->W]-eGFP was injected into

wild-type embryos at the one-cell stage. Embryos were allowed to develop at 28°C. At 3 hours post fertilization, embryos were mounted in 1% low-melting point agarose and imaged on an inverted Pascal confocal microscope (Zeiss) (20x air objective). Note that overexpression of GFP-tagged Toddler does not cause Toddler overexpression phenotypes, suggesting that fusing GFP to the C-terminus of Toddler maintains secretion but renders Toddler non-functional.

***In situ* hybridization and quantification of mesendodermal phenotypes**

In situ hybridization experiments were performed according to standard protocols (*64*), using DIG-labelled antisense RNA probes against *toddler, sox17, fn1, ntla, gsc, drl, tbx16, foxA2, cmlc2, scl/tal, cp, hgg1, aplnra,* and *aplnrb.* BCIP/NBT/Alkaline-Phosphatase-stained embryos were dehydrated in methanol and imaged in BB/BA, using an Axio Imager.Z1 microscope (Zeiss).

For quantification of lateral mesendodermal phenotypes at 70% epiboly, embryos were imaged either laterally (measuring one lateral side) or as cross-sections (dorsal side up, measuring both lateral sides). The height of lateral *sox17* or *fn1*-expression domains was measured in ImageJ (*63*). Each value was normalized to the average height in untreated wild-type embryos (or *toddler* mutant embryos for Fig. 4) that were processed in parallel. To assess the ratio between dorsal versus lateral spread of endoderm at 70%, the dorsal height of the *sox17* expression domain was measured as outlined in fig. S9C and divided by the lateral band height in the same embryo. Endodermal cells were counted in a fixed-size circular area on one lateral side. Boxplots and standard pairwise Student's t-test calculations of two samples with unequal variance were generated in R using RStudio.

Transplantation and YSL injection

To generate localized sources of Toddler peptide, *toddler* mRNA (synthesized from the pCS2+ *toddlerORF-SV40UTR* vector) was expressed either vegetally (yolk syncytial layer (YSL) injections) or animally (transplantation of *toddler* overexpressing donor cells to the animal pole) in *toddler* mutant embryos. For YSL injections, 1000-cell stage *toddler* mutant embryos were injected into the YSL with 2 pg *toddler* mRNA and Dextran-Red as tracer. For transplantation experiments, one-cell stage *toddler* mutant embryos were injected with 200-400 pg *toddler* mRNA (donor embryos). Approximately 20 to 50 cells from donor embryos were transplanted to the animal pole of stage-matched *toddler* mutant host embryos.

Tracking of endodermal cell migration

1) Confocal microscopy

Endodermal cell migration was analyzed in wild-type and *toddler* mutant embryos carrying the *sox17::eGFP* transgene (movies S1 and S2) (*14*). Nuclei were labeled by injection of 50 pg of *H2B-RFP* mRNA at the one-cell stage. Wild-type and *toddler* mutant embryos were grown to 50% epiboly (the stage when fluorescence in endodermal cells becomes first detectable) at 28°C and mounted laterally in a single dish in 1% low-

melting point agarose in 1x Danieau's medium. Time-lapse imaging was performed at 26^{o}C (using a heated stage) on an inverted LSM5 Pascal confocal miscroscope (Zeiss). Movies S1 and S2 were obtained with a 10x air objective. Manual tracking of endodermal cell movements was performed in wild-type and *toddler* mutant embryos for visualization purposes (movie S2 and fig. S11B), using the manual tracking plugin in Fiji.

2) Endodermal cell tracking in MATLAB

For quantification of endodermal cell movement (Fig. 3B), endodermal cells were imaged with the following settings: 25x/0.8 water immersion objective, 1.2x zoom; 3 z-slices, each 10.95 µm thick; 188-second intervals between frames. A multi-time imaging macro (ZEN software) was used to image each embryo in the dish sequentially over 4 consecutive 45-60 minute time-periods. Consecutive 45-60 minute time periods were chosen instead of a continuous long movie to adjust the imaging field for each embryo. Cell tracking was performed with a custom MATLAB script, using the nearest-neighbor strategy. After image segmentation implementing the Moore-Neighbor tracing algorithm modified by Jacob's stopping criteria to identify GFP-labeled endodermal cells, the centroid position and area of each object was determined in every stack for all time frames. Because the same cell can appear in different z-slices, objects whose pixels overlap in consecutive slices were compared according to their area, and the centroid of the object with the smallest area was discarded. Next, the distance between centroids in the xyz space between consecutive frames was calculated. The nearest centroid in the next frame was selected as part of the cell track if it was less than 10 pixels from the centroid in the previous frame. This process was reiterated through all the frames to generate complete cell tracks. When new cells appeared between two frames (e.g., due to cell intercalation), a new track was started for this given cell. False-positive tracks (due to 'jumps' in the tracks or loss of cells from the imaging field) were subsequently removed in MATLAB. Based on manual inspection of the resulting tracks, ~90% of the tracks are estimated to be accurate. Velocity, displacement and directionality (angles) of cell tracks were calculated and plotted in MATLAB.

Tracking of gastrulation movements

1) Lightsheet microscopy

Gastrulation movements were analyzed by lightsheet microscopy in wild-type and *toddler* mutant embryos injected with 50 pg of *H2B-RFP* mRNA (movies S3-S6). Embryos were allowed to develop at 28^{o}C until early dome stage. Within a single experiment, 2-5 embryos (containing at least one embryo of each genotype) were mounted in 1% low-melting point agarose in 1x Danieau's in glass capillaries (Zeiss) according to published protocols (*65*, *66*). After placing the glass capillary in the pre-heated (26.5^{o}C), fish-water-filled imaging chamber of a Lightsheet Z.1 microscope (Zeiss), the agarose cylinder containing the embryos was pushed out of the glass capillary to allow for equilibration before image acquisition. The multiview and z-stack options in the ZEN Black software (Zeiss) were used to record time-lapse movies of up

to 5 embryos at time-intervals between frames no larger than 90 seconds. Although internalizing cells move rapidly, a temporal resolution of 90 seconds was sufficient to allow for automated nuclei tracking (see below) without compromising the ability to image multiple embryos at the same time. The latter was crucial in order to obtain movies from wild-type and mutant embryos of comparable views, imaged within the same experiment. To minimize the imaging time per embryo, cross-sections of 40-50 slices (1.5-2 µm interval per slice) were recorded instead of z-stacks through full (equivalent to >200 z-slices) embryos.

Timelapse acquisition settings for imaging on the Lightsheet Z.1 (Zeiss) (Single Plane Illumination System):

- 20x/1.0 water immersion objective, 0.5x zoom
- Incubation temperature: 26.5^{o}C
- 561 nm laser: 2% power; 24.97ms exposure time
- Detection: Dual PCO.Edge 2nd generation sCMOS cameras, liquid cooled, 1920x1920 pixels, ~70fps
- 16 bit images (1920x1920 pixels; 878.8 µm x 878.8 µm; pixel size 0.46 µm)
- Dual-side illumination with online dual fusion and pivot scan
- Z-stacks: 40-100 slices per embryo; 1.5 - 2 µm intervals
- Time-series: 90 second intervals (Fig. 3C-G, movies S4 and S5);
  60 second intervals (fig. S13)
  70 second intervals (fig. S12 and S14; movies S3 and S6)

Initial post-acquisition processing of the imaging data was performed with the ZEN Black (Zeiss) software to reduce the file sizes from ~1TB to the 100-200GB range. To this end, stacks were cropped and down-sampled by a factor of 2 in x and y. Individual positions (i.e., embryos) were extracted, and maximal intensity projections were obtained. For each embryo, a subset of 20-40 z-slices was selected from the larger z-stack based on equivalent positioning/depth within embryos of comparable views and different genotypes. These 20 z-slice-stacks were further cropped to only one lateral side of the embryo and size-reduced to two-thirds of the size in ImageJ/Fiji (*63*).

2) Nuclei tracking using TrackMate v2.0.4

Cell-tracking was performed in Fiji (Life-Line version for OS X) (*63*), using the inbuilt plugin TrackMate v2.0.4 (*67*) with the following settings: 1) Detection of nuclei: LoG detector; 10 µm 'blob' (nuclear) diameter; median filter, no sub-pixel localization; 2) Thresholding values (by Quality and Total Intensity) were set by visual inspection to discard false-positive spots while retaining the maximal number of correctly identified spots. 3) Tracking of paths: simple LAP tracker; linking max distance = 15 µm; gap-closing max distance = 15 µm; gap-closing max frame gap = 0. Tracks with less than 5 nuclei were discarded. Track statistics for tracks, links and spots were saved for custom analysis in R.

3) Analysis of gastrulation movements

Embryos (tracks) were first rotated to match a standard positioning (animal up, lateral margin right). The onset of internalization (time-frame when the first mesendodermal cells were observed to move inside) was annotated manually based on inspection of the movies. Marginal areas were annotated individually for each embryo at the time of internalization and defined as the region extending from the margin to 100 µm above (~ 10 marginal cell tiers). Movies of wild-type and *toddler* mutant embryos from matching lateral views were aligned in time to the onset of internalization (t(Int)). Velocities and angles were calculated using the measurements (velocities and coordinates) provided in the TrackMate results files. A cell/track displacement was categorized as "moving upwards" if the displacement was within a +/- 45^{o} angle of the straight upwards (animal) movement. "Internalizing tracks" were defined as cells/tracks that had visited the marginal area at any time during the movie and were "moving upward" at a rate of at least 20% of total movement.

Annotated tracking data was visualized in three different ways: 1) Plot of cell (nucleus) position, speed (dot size) and directionality (color-code from blue (vegetal movement) to red (animal movement)), at the beginning, during, and after internalization (Fig. 3E; fig. S13C; fig. S14B). 2) Visualization of tracks before (time < -5 min), during (-5 min < t < 1 h), and after (t > 1 hour) the onset of internalization (t(Int)), color-coded by the total number of animal pole- (red shades) or vegetal pole- (blue shades) directed movements, normalized to the total number of frames per track (Fig. 3F; fig. S13D). 3) Visualization of tracks before (t < -5 min), during (-5 < t < 1 h), and after (t > 1h) the onset of internalization, color-coded based on their relative position and directionality with respect to the margin at the time of internalization (Fig. 3G; fig. S13E, fig. S14C).

GPCR internalization assays

Three strategies were used to test whether Toddler can internalize Apelin receptors *in vivo*. To co-express signaling peptides and eGFP-tagged GPCRs, 50 pg of mRNA encoding eGFP-tagged GPCRs (Aplnra, Aplnrb, Cxcr4a, Cxcr4b or Cxcr7b) was injected into one-cell stage *toddler* mutant embryos, either individually or together with 50 pg of signaling peptide encoding mRNA (Toddler, Sdf1a or Sdf1b). To generate a localized source of signaling peptide-expressing cells, 50 pg of mRNA encoding eGFP-tagged Aplnrb was injected at the one-cell stage into *toddler* mutant embryos. At the 128-cell stage, 200 pg of mRNA encoding either Toddler or Sdf1a was injected into a single cell together with Dextran-Red as fluorescent tracer. Embryos were allowed to develop at 28^{o}C until ~3 hours post fertilization. For peptide-induced internalization studies, 1-5 ng of peptide (C-terminal Toddler or Apelin peptide fragments) was injected extracellularly together with Alexa543-Dextran into *toddler* mutant embryos that had been injected at the one-cell stage with 50 pg of mRNA encoding a fluorescently tagged GPCR. Embryos were mounted in 1% low-melting point agarose and imaged on an inverted Pascal confocal microscope (Zeiss) (20x objective). Images were processed in ImageJ (*63*).

Preparation of protein extracts and LC-MS/MS analysis

Wild-type embryos were injected with 100 pg *toddler* mRNA at the one-cell stage and grown at 28^{o}C. At ~6 hpf (early shield stage), embryonic cells including the extracellular fraction were collected using a wide transplantation needle to aspirate all cells and extracellular material. Batches of 10 embryos each were processed at a time (20 batches in total), and the resultant protein fraction was flash-frozen in liquid nitrogen. Boiling water (500 µl) was added directly to the frozen embryo pellets, and samples were boiled for 20 minutes to eliminate proteolytic activity. After cooling to room temperature, samples were sonicated on ice for 20 bursts at output level 4 with a 40% duty cycle (Branson Sonifier 250; Ultrasonic Convertor). The cell lysate was then brought to 0.25% acetic acid by volume and centrifuged at 20,000 x g for 20 minutes at 4°C. The supernatant was evaluated for protein content by BSA assay and then evaporated to dryness at low temperature in a SpeedVac. Pellets were re-suspended in 50 µl of 25mM TCEP in 50mM NH_4HCO_3 (pH=8) and incubated at 37°C for 1 hour. The reaction was cooled to room temperature before addition of 50 µl of a 50 mM iodoacetamide solution in 50 mM NH_4HCO_3. Following incubation in the dark for 1 hour, the mix was bound to a C18 Sep Pak cartridge (HLB, 1cm^3; 30mg, Oasis), washed thoroughly with water and eluted with 1:1 acetonitrile/water. The eluate was evaporated to dryness at low temperature on a SpeedVac. The sample was then resuspended in 25 ul of 0.1% formic acid (aqueous).

10 ul of the sample was injected onto a Waters nanoAcquity HPLC configured with in-house packed 75-µm reverse-phase capillary trapping and analytical columns (New Objective) coupled to an LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific). The liquid chromatography gradient proceeded from 3 to 33% acetonitrile/water (0.1% formic acid) over 180 min. Mass spectra from 395 to 1,600 *m/z* were acquired in the Orbitrap with a resolution of 60,000, and the 20 most abundant ions were targeted for concurrent MS/MS in the linear ion trap with relative collision energy of 30%, 2.5-Da isolation width, and recurring ions dynamically excluded for 60 s. Sequencing was done using the SEQUEST algorithm. Peptides were accepted within 3 ppm of the expected mass if they also met a series of custom filters on ScoreFinal (Sf), –10 Log P and charge state that attained an average peptide false discovery rate (FDR) of <2% across datasets. Manual inspection of spectra, FDR calculation, and protein inference were performed in Proteomics Browser Suite 2.23 (Thermo Fisher Scientific).

Supplementary Figure Legends

Fig. S1: Identification of novel embryonic putative signaling proteins in zebrafish. Sequence alignments and genomic regions of nine newly identified genes encoding proteins with predicted signal peptide sequences. Shown are single-nucleotide mapped ribosome profiling reads (Ribosome Protected Fragments – RPFs (*7*); blue), RNA-Seq based transcripts ((*8*) or (*9*); black) and – if available - Ensembl (red) and RefSeq (blue) gene annotations. Within secreted peptide-encoding transcripts, the region encompassing the predicted protein-coding ORF is highlighted as a thick bar. Note that for all predicted secreted peptide-encoding transcripts, the Ensembl gene annotation is non-coding (absence of a thick bar, indicating the absence of a predicted coding ORF). Genomic sequence alignments from eight species (Medaka, Stickleback, Tetraodon, Fugu, Xenopus tropicalis, Mouse, Human) are shown underneath the gene tracks. The full-length amino acid sequence of each peptide is given below the tracks (predicted signal peptide sequence, blue; mature (secreted) peptide, black). The *toddler* genomic locus is shown on the top left.

Fig. S2: Genomics of *toddler*. Genomic location of *toddler* in zebrafish (**A**), human (**B**) and mouse (**C**). *Toddler* (highlighted by a red arrow) is annotated as a non-coding gene in all three species according to the current genome assemblies (Zv9/danRer7 (zebrafish), GRcg37/hg19 (human), and GRCm38/mm10 (mouse)).

Fig. S3: Toddler is conserved in vertebrates. (**A**) ClustalW2 multiple protein sequence alignment of Toddler peptide sequences from 16 vertebrates. Two color-codes are shown – clustalW2 (top) and percentage identity (bottom; darker shading indicates higher percentage identity of the amino acid). A Toddler consensus sequence is shown on top. The predicted signal peptide cleavage site and the highly conserved C-terminal 11 amino acid (aa) peptide fragment that was detected by mass-spectrometry are indicated. (**B**) Phylogenetic tree of Toddler in 16 vertebrates. The tree is based on the protein sequence alignment shown in (**A**).

Fig. S4: Mass-spectrometric identification of Toddler. LC-MS/MS analysis identified the C-terminal 11 amino acid Toddler peptide fragment in *toddler* mRNA overexpressing shield-stage embryos (6 hours post fertilization). The spectra is from the detected cleavage fragment CLPLHSRVPFP. For details see Materials and Methods.

Fig. S5: TALEN-mediated mutagenesis of *toddler*. (**A**) Genomic sequence encompassing the 5' end of the *toddler* ORF (ATG highlighted in bold) in exon1 up to the start of intron1. Binding sites of the *toddler* TALEN pair ($TALEN^{left}$ and $TALEN^{right}$, green) are highlighted in red. (**B**) Nature of the TALEN-induced *toddler* mutations. The predicted TALEN cleavage site is flanked by the $TALEN^{left}$ and $TALEN^{right}$ binding sites; TALEN binding sites are indicated by the red lines above the sequence alignment.

Sequences were aligned with ClustalW2 to the wild-type *toddler* DNA sequence. (**C**) Putative Toddler translation products from the mutated *toddler* alleles shown in (**B**). All 7 alleles maintain the signal peptide sequence (blue), but abolish the production of the mature, secreted portion of Toddler (propeptide sequence). Amino acids that differ between the wild-type Toddler protein (top) and the mutant alleles are highlighted in red.

Fig. S6: *Toddler* mutants lack a functional heart. *cmlc2 in situ* hybridization reveals different degrees of loss of heart tissue in *toddler* mutants and *toddler* overexpressing larvae at 36 hpf.

Fig. S7: Variability in expressivity of *toddler* mutant phenotypes. Example images of lateral (left) and ventral (right) views of wild-type and *toddler* mutant embryos at 28 hours post fertilization. Phenotypic severity of *toddler* mutants varies both within and between clutches. The predominant and most consistently observed phenotype is of Class I (50-100%). The more severe Class II phenotypes are often observed in a sub-fraction (10-50%) of embryos of clutches of predominantly Class I phenotypes. Class III phenotypes are observed at low frequencies (0-10%).

Fig. S8: Toddler is required during late blastula/early gastrula stages. (**A**) Overview of the crossing scheme used to determine the time-requirement of Toddler during early embryogenesis. (**B**) Phenotypic analysis of *cmlc2::eGFP*-positive animals at 60 hpf. Embryos were heat-shocked at the given time for 45 minutes at 37^{o}C. Note that all *cmlc2::GFP*-positive animals also carry the *hs:toddler-SV40UTR* transgene. Rescue of *toddler* homozygous mutants was only observed for heat shock treatments during late blastula/early gastrula stages.

Fig. S9: Toddler is required for ventrolateral movement of endoderm and mesoderm towards the animal pole. (**A**) Endodermal marker gene expression (*sox17*) in wild-type and *toddler* mutant embryos (*toddler*) at 70% epiboly. Embryos of the indicated genotypes were injected at the one-cell stage with either 2 pg (rescuing concentration for *toddler* mutants) or 10 pg (overexpression) of *toddler* mRNA. Doses of 2 pg *toddler* mRNA in wild types and of 10 pg *toddler* mRNA in both wild types and *toddler* mutants cause endodermal defects resembling *toddler* loss-of-function phenotypes (decreased spread of ventral and lateral endoderm towards the animal pole). The embryos shown are from the experiment quantified in Fig. 2D. Shown are lateral and dorsal views for each treatment. (**B**) Mesodermal marker gene expression (*fascin*, *tbx16*, *draculin/drl*) in wild-type and *toddler* mutant embryos at the indicated stages. Note the ragged appearance and decreased width of *fascin*- and *tbx16*- marked mesoderm in *toddler* mutant embryos. *In situ* hybridization with *drl* (labeling blood progenitor cells) at 70% epiboly reveals a distinct subpopulation of *drl*-positive cells in *toddler* mutant embryos that are positioned more vegetally than the margin. The arrow indicates the position of the margin. (**C**) *Toddler* loss- and gain-of function interfere predominantly with lateral and ventral animal pole-directed movement of endodermal

cells, while dorsal animal pole-directed movement is largely unaffected. Quantification of the ratio between dorsal and lateral height of the band of *sox17*-expressing cells at 70% epiboly for embryos of the experiment shown in (**A**) and quantified in Fig. 2D. (**D**) Quantification of the defect in lateral mesoderm spread observed in *fn1 in situ* hybridization experiments of 70% epiboly embryos (example images are shown in Fig. 2C). p-values in (**C**) and (**D**) for pairwise comparisons with wild type (black, top) or *toddler* mutant (magenta, bottom) were calculated based on a standard Welch's t-test (* p-value < 0.01; ** p-value < 0.00001).

Fig. S10: Analysis of germ layer specification and morphogenesis in *toddler* mutant embryos. (**A**) Formation and morphogenesis of ectoderm (*sox3*), dorsal mesendoderm (*gsc, hgg1*) and tail mesoderm (*ntla*) is largely normal in *toddler* mutant embryos. (**B**) Initial specification of endoderm (*sox17*), dorsal mesendoderm (*gsc*) and mesoderm (*fn1, ntla*) is largely normal in *toddler* mutant embryos.

Fig. S11: *Toddler* mutants are defective in animal pole-directed endodermal cell movement. Analysis of endodermal cell migration in *sox17::eGFP* transgenic wild-type (top) and *toddler* mutant (bottom) embryos by confocal microscopy. Endodermal cells (marked by *sox17*), green; nuclei (*human Histone2B-RFP* (*H2B-RFP*) mRNA injection), red. (**A**) Still images of maximum intensity projections of a timelapse movie from 60% to 90% epiboly (movie S1). The leading edge of animal pole-directed endodermal cells (yellow circles) and the position of the margin (white crosses) are overlayed and plotted on the right. Note that endodermal cells in *toddler* mutant embryos show reduced animal pole-directed endodermal cell displacement, both in respect to the position of the margin at 80% epiboly (red arrow) as well as in respect to their starting position in the embryo (green arrow – maximal displacement towards the animal pole). (**B**) Still images of maximum intensity projections of the green channel (*sox17*-positive cells) of the movie of panel **A**, overlayed with representative example cell tracks obtained by manual tracking in Fiji (see movie S2 for the complete movie). Arrows indicate the direction of migration. In contrast to the animal pole-directed movements in wild-type embryos (up), endodermal cells migrate vegetally (down) in *toddler* mutant embryos.

Fig. S12: Ventral cell tracks in wild-type and *toddler* mutant embryos during internalization. Shown are a series of still images of maximum intensity projections of 40 µm ventral cross-sections in a wild-type and a *toddler* mutant embryo during the time of internalization, obtained by Single Plane Illumination Microscopy (SPIM). Movies were aligned to the time of internalization at the dorsal side (not depicted in this figure). Example cell tracks were obtained by manual tracking in Fiji and are representative for the distinct internalization movements observed in wild-type and *toddler* mutant embryos. The stills are derived from movie S3. The movie S6 contains the full embryo views (dorsal-ventral cross-section).

Fig. S13: Abnormal gastrulation movements in *toddler* mutant embryos. Analysis of early gastrulation movements in a single pair of *H2B-RFP* mRNA injected wild-type and *toddler* mutant embryos by lightsheet microscopy (Single Plane Illumination Microscopy (SPIM)). Note that this embryo pair is different from the one analyzed in Fig. 3C-H and was imaged in an independent experiment. Time-interval between frames is 60 seconds; timelapse imaging was performed up to ~70% epiboly. Analyses were performed on lateral cross-sections of time-lapse movies of z-stacks (20 z-slices of 2 µm each). (**A** and **B**) Still images of maximum intensity projections of 40 µm cross-sections during the time of internalization. Movies were aligned at 50% epiboly (shown here as 00:00) (**A**) or at the onset of internalization (**B**). Leading edges of internalizing mesendodermal cells (yellow dots) and vegetally moving cells (green dots) highlight the opposing paths of cells during gastrulation. Animal pole- and vegetal pole-directed migratory paths of the wild-type and mutant embryo are plotted in a single graph underneath. Frame-to-frame displacements (left) were used to derive the actual net animal pole-directed cell movement per genotype. *toddler* mutants show delayed onset of internalization and reduced step-to-step and net animal pole-directed movement. (**C-E**) Cell tracking and digital analysis of gastrulation movements. (**C**) Position, speed (dot size) and directionality (color-code from blue (vegetal movement) to red (animal movement)) of tracked cells during and after the time of internalization. Movies were aligned to the onset of internalization (t(Int) = 00:00, time in hours:minutes). (**D** and **E**) Cell tracks before (t < -5 min), during (-5 min < t < 1h), and after (t > 1h) internalization in wild-type and *toddler* mutant embryos. In (**D**), tracks were color-coded based on the total number of animal pole- (red) or vegetal pole- (blue) directed movements, normalized to the total number of frames per track. In (**E**), tracks were color-coded based on their relative position and directionality with respect to the margin at the time of internalization (margin cells: cells located within 100 µm above the margin at the time of internalization). Non-margin cells, grey; margin cells, black; internalizing and upwards-moving margin cells, red.

Fig. S14: *Toddler* mutant embryos are defective in ventral but not dorsal internalization movements. (**A**) Still images of maximum intensity projections of 40 µm (30 slices) dorsal-ventral cross-sections of a wild-type (left) and *toddler* mutant (right) embryo pair that was quantified in Fig. 3I (movie S6). Two time-points during internalization are shown. Arrows highlight the paths that the leading internalizing cells took on dorsal (D, dashed white line) and ventral (V, solid yellow line) sides of the embryo. Ventral movement towards the animal pole is severely reduced in the *toddler* mutant embryo, while dorsal internalization occurs normally. (**B-C**) Cell tracking and digital analysis of dorsal versus ventral gastrulation movements of the embryo pair shown in panel (**A**). (**B**) Position, speed (dot size) and directionality (color-code from blue (vegetal movement) to red (animal movement)) of tracked cells during and after the time of internalization (t(Int)). Movies were aligned to the onset of internalization (t(Int) = 00:00; time in hours:minutes). (**C**) Cell tracks before (t < -5 min) and during (-5 min < t < 1h) internalization. Tracks are color-coded based on their relative position and

directionality with respect to the margin at the time of internalization (margin cells: cells located within 100 µm above the margin at the time of internalization). Non-margin cells, grey; margin cells, black; internalizing and upwards-moving margin cells, red.

Fig. S15: *Toddler* mutants have defects in cell internalization. *In situ* hybridzation experiments using *fibronectin* (*fn1*) as a probe reveal pile-up of internalizing mesodermal cells at the margin in *toddler* mutant embryos. In contrast, *fn1*-expressing cells spread out evenly and migrate animally in wild-type embryos. Shown are dorsal views of shield-stage embryos. The marginal, lateral region is enlarged at the bottom.

Fig. S16: Toddler peptide is biologically active. (**A**) Comparison of full-length zebrafish Toddler and Apelin peptide sequences. Signal peptide sequence, light blue; mature prepropeptide sequence, dark blue. The *in vitro* synthesized C-terminal 21 (Toddler) or 16 (Apelin) amino acid peptide fragments used for overexpression and Apelin receptor internalization studies are highlighted in red. (**B**) Extracellular injection of 5 ng of Toddler or Apelin peptide fragments into wild-type embryos induces gastrulation defects reminiscent of *toddler* mRNA overexpression (reduced width of endoderm (*sox17*) and mesoderm (*fn1*) at 70% epiboly). (**C**) Quantification of the defect in endodermal cell movement based on *sox17 in situ* hybridization. p-values for pairwise comparisons with wild type were calculated based on a standard Welch's t-test (** p-value < 0.00001). (**D**) Wild-type embryos injected extracellularly with 5 ng of Toddler or Apelin peptide show characteristic *toddler* mRNA overexpression phenotypes at 30 hpf (posterior accumulation of blood cells; lack of a functional heart).

Fig. S17: *Toddler, aplnra* and *aplnrb* expression during embryogenesis. (**A**) Expression analysis of *toddler*, *aplnra* and *aplnrb* in wild-type embryos at the indicated stages by *in situ* hybridization (see also (*24*, *25*, *39*)). (**B**) The marginal expression of *aplnra* and *aplnrb* is dependent on Nodal signaling (see also (*39*)). Nodal signaling was inhibited in wild-type embryos by drug-treatment (20 µM SB-505124). Shown are animal views of wild-type or drug-treated embryos at germring stage. Note that the animal pole expression of *aplnrb* and the EVL expression of *aplnra* are unchanged in embryos lacking Nodal signaling.

Fig. S18: Toddler drives internalization of Apelin Receptors. Coinjection of *signal* and *receptor-eGFP* mRNA into one-cell stage *toddler* mutant embryos. This Figure contains the complete set of experiments and controls shown in part in Fig. 6B (additional panels are Cxcr4b-eGFP and *sdf1b* mRNA overexpression). In the absence of signal overexpression, ectopically expressed receptors localize to the plasma membrane in pre-gastrulation *toddler* mutant embryos. Coinjection of *toddler* and *aplnra-eGFP* or *aplnrb-eGFP* mRNA prevents membrane-localization of Aplnra-eGFP or Aplnrb-eGFP, while chemokine receptors Cxcr4a-, Cxcr4b- and Cxcr7b-eGFP remain unaffected. Conversely, Sdf1a or Sdf1b overexpression results in internalization of Cxcr7b-eGFP while Aplnra-eGFP or Aplnrb-eGFP remain membrane-localized.

Fig. S19: Overexpression of mouse and human Toddler in zebrafish leads to similar phenotypes as overexpression of zebrafish Toddler. Endodermal gastrulation phenotypes were quantified by measuring the lateral height of *sox17*-stained 70% epiboly embryos (n = number of embryos per category). Mouse and human *toddler* mRNA overexpression are shown in green shading on the right. The remaining treatments are from Fig. 5C. p-values for pairwise comparisons with wild type (black, top) or *toddler* mutant (magenta, bottom) were calculated based on a standard Welch's t-test (* p-value < 0.01; ** p-value < 0.00001).

Supplementary Movie Legends

Movie S1: Endodermal cell migration in wild-type and *toddler* mutant embryos (embryo pair shown in Fig. 3A and fig. S11). Analysis of endodermal cell migration in *sox17::eGFP* transgenic wild-type (left) and *toddler* mutant (right) embryos by confocal microscopy. Endodermal cells (marked by *sox17* expression), green; nuclei (*H2B-RFP* mRNA injection), red. Maximum intensity projections of a 6 hour and 40 minute timelapse movie, imaged at 4 min time-intervals on an inverted Pascal confocal miscroscope (Zeiss) (10x objective). D, dorsal side; time in hours:minutes.

Movie S2: *Toddler* mutants are defective in animal pole-directed endodermal cell movement (same embryo pair as shown in movie S1; this movie is complementary to fig. S11B). Analysis of endodermal cell migration in *sox17::eGFP* transgenic wild-type (left) and *toddler* mutant (right) embryos by confocal microscopy. Imaged endodermal cell movements (marked by *sox17* expression; white) were overlayed with example cell tracks derived from manual cell tracking in Fiji (coloured tracks) (top). The current position of the tracked cell is indicated by a sphere. The bottom two panels show only the cell tracks. Note the severe reduction in animal pole-directed cell movement in *toddler* mutant embryos.

Movie S3: Cell tracking of internalization movements in wild-type and *toddler* mutant embryos (this movie is complementary to fig. S12). Maximum intensity projections of 40 µm ventral cross-sections (30 z-slices) in wild-type (left) and *toddler* mutant (right) embryos, in which all nuclei had been labeled by *H2B-RFP* mRNA injection. Example cell tracks following the movement of individual internalizing cells were obtained from manual cell tracking in Fiji. Cell tracks illustrate the defect in ventral internalization and animal pole-directed cell movement in *toddler* mutant embryos. Intervals between frames are 70 seconds; time in minutes:seconds; light-sheet microscopy.

Movie S4: Gastrulation movements in wild-type and *toddler* mutant embryos (embryo pair analyzed in Fig. 3C-G). Maximum intensity projections of 40 µm lateral cross-sections (20x 2µm z-slices) of the time-lapse movie of the wild-type (left) and *toddler* mutant (right) embryos analyzed in Fig. 3C-G. Movies were aligned at 50% epiboly (reached at 48 minutes). Intervals between frames are 90 seconds; time in minutes:seconds; light-sheet microscopy.
The movie is played twice: in the first half without annotations, and in the second half with annotations. Colored dots were overlayed every 12th frame at the position of the

leading edge of animal pole-directed (yellow) and vegetal pole-directed (blue) cells (see Fig. 3C for still images of the dot-annotations). The onset of internalization is highlighted in red.

Movie S5: Gastrulation movements in a wild-type embryo and a *toddler* mutant embryo that shows no animal pole-directed cell movement. Maximum intensity projections of 40 µm lateral cross-sections (20x 2µm z-slices) of a time-lapse movie of a wild-type embryo (left) and both lateral sides of a *toddler* mutant embryo (mid, lateral side 1 (*toddler L*); right, lateral side 2 (*toddler R*)). All nuclei had been labeled by *H2B-RFP* mRNA injection. Note that there is no animal pole-directed cell movement on both lateral sides in this *toddler* mutant embryo. Marginal cells in the *toddler* mutant embryo move instead vegetally. Intervals between frames are 90 seconds; time in minutes:seconds; light-sheet microscopy.

Movie S6: *toddler* mutant embryos are defective in ventrolateral but not dorsal internalization movements (this movie is complementary to fig. S14 and was used for the quantification shown in Fig. 3I). Maximum intensity projections of 40 µm dorsal-ventral cross-sections (30 z-slices) of a time-lapse movie of a wild-type (left) and a *toddler* mutant embryo pair in which all nuclei had been labeled by *H2B-RFP* mRNA injection. While internalization occurs normally on the dorsal side in *toddler* mutants, it is severely impaired on the ventral side. Intervals between frames are 70 seconds; time in minutes:seconds; light-sheet microscopy.

Supplementary Table

Table S1 (Excel file)
Table listing the 700 newly identified protein-coding transcripts contained within 399 loci (genes). The 40 novel putative secreted protein-encoding transcripts (28 loci, highlighted in light blue) are shown on top. Each transcript is characterized by a series of features:
column1: index
column2: locus ID (XLOC_xxx)
column3: isoform ID (derived either from the RNA-Seq based embryonic transcriptome described in (8) (Zv7_xxx IDs) or from (9) (linc-xxx IDs))
column4: length of the predicted coding ORF in amino acids
column5: PhyloCSF score (8)
column6: fraction of nucleotides hit by single-base mapped ribosome profiling reads (RPF coverage)
column7-column9: predicted features of the encoded protein (secreted protein; transmembrane protein; predicted signal peptide and transmembrane annotations (see online Phobius documentation for details (http://phobius.sbc.su.se/instructions.html))
column10-column16: percentage alignment between the transcript in zebrafish and other vertebrates (Fugu, Medaka, Tetraodon, Stickleback, Xenopus, Mouse and Human).
column17: additional comments.

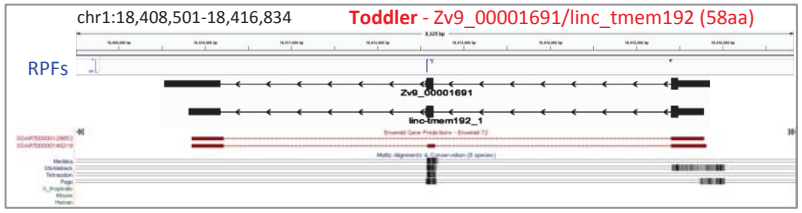
Note that an alignment score of 0 does not necessarily mean that a transcript is not conserved. This is highlighted by the example of *toddler*: The alignment scores of 0 between zebrafish *toddler* and mammalian species is due to the absence of sequence alignments encompassing the *toddler* genomic region. As demonstrated in Fig 1D and fig. S3A, the *toddler* transcript encodes a highly conserved short peptide present from fish to human.

Supplementary Data Files

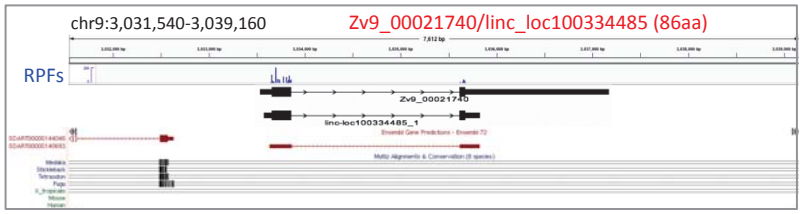
Data file S1: Novel protein-coding transcripts (.bed file)
Bed file containing the transcript structures (including genomic positions and ORF annotations) of all 700 identified putative protein-coding transcripts. Each transcript is identified by its unique isoform ID (Zv9_xxxx or linc-xxx).

Data file S2: Novel proteins (.fa file)
Fasta file containing the predicted amino acid sequences of all 700 identified putative novel proteins. Each protein is identified by its associated unique transcript isoform ID (Zv9_xxxx or linc-xxx).

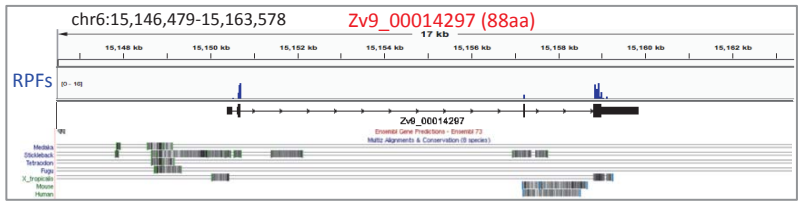
fig. S1



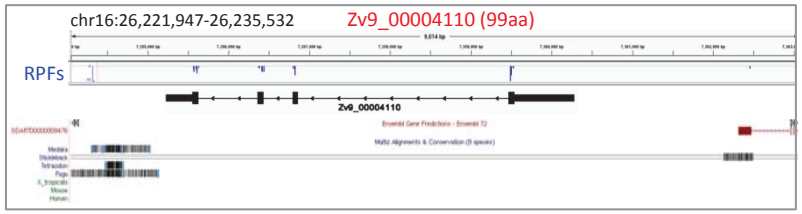
MRFFHPLYLLLLLLTVLVLISADKHGTKHDFLNLRRKYRRHNCPKKRCLPLHSRVPFP*



METTEAVLLLLLVCCWWRARQRHLDEEIMARRRFRRKLILAARRKRAIERRERRLERQR
KRQEWMNMMMHYYYTNILCESDESDDE*

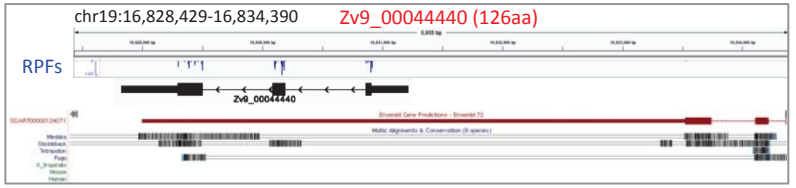


MFRLGFSVLLCVVWLGLLCLTGAVENQDESREKRSVPNWAMTSSDFFGWIEELRSHAGYD
KMEELARTFWAHFPSACRLGYDPPAPEE*

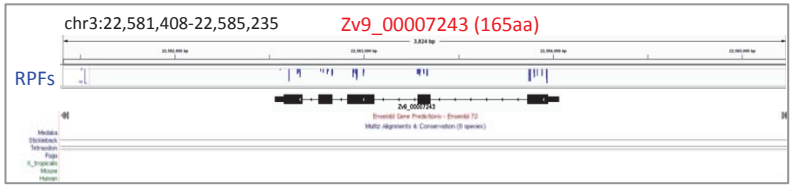


MEKMVHLVLLVGFLGLSAGQLIPTGNSDVTPSNMEATANGTQPPVVIGMRMTVKSVVD
LLRNSAEIELLVRQVESLIPYADKGKTSVTVRSIRSSSSIA*

MGCVLLFLLLVCVPVVLPQGLRCLFCPVTSLNSSCAPVVTECPVQELCYTADGRFGRSSVLF
RKGCMLRADCSRSRHQMIRGNNISFSFSCCGGHYCNSQPRAEPGGRLLLLLLPAAALTAAG
AL*

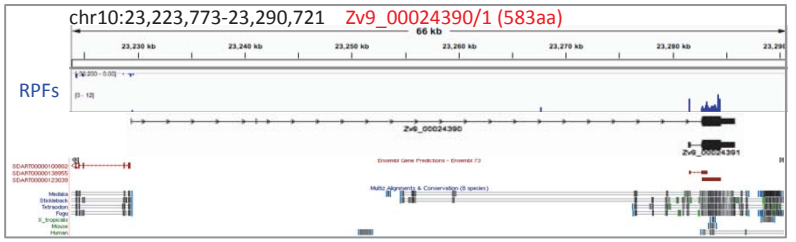


MMNRILLGVITVIGFFTLAEALTCNSCKVGILGKCLYSSQVACASSEINCYTASAVFNVTGFLS
LSSSGCTSNCNNTVGTILGAGYSITKTCCITDLCNGASAVQLSTTAAFSTALLASIWSSYML*



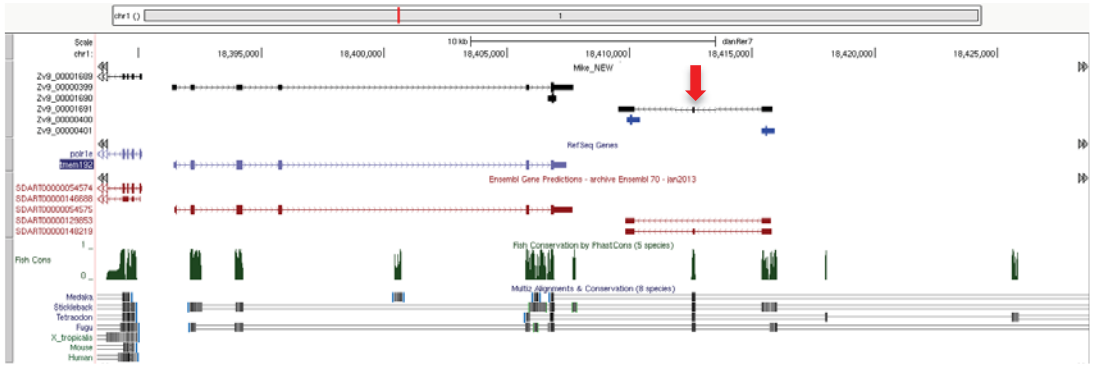
MRAVLCFLCLIFFPSLFVSSNQERRTLITTIKLLNALGPKACINEVENANSSRFRNLIKKRQEIP
EALKISYNSTFEIFKKKVNNCHWHNNAFRDLMEHLNRQIISNTAPVISSELETMVNDFKLLD
DFLTTQKFSCCAWEIARHDVLVVMDHFRKRSAQRKIHY*

MNSSGVCVLLSLLFLSVSADQCDNDNTADNCNKWTETSHNQPRNTGIAVTYEQLKHMM
STGRVQLIDVREPDELTAGFIPGATNIPLGELEEALTLDPDQFRQRYGVSKPHPDDSDIVLY
CQRGVRSLNALEIAIRLEYSRARHYVGGYSEWAERERL*



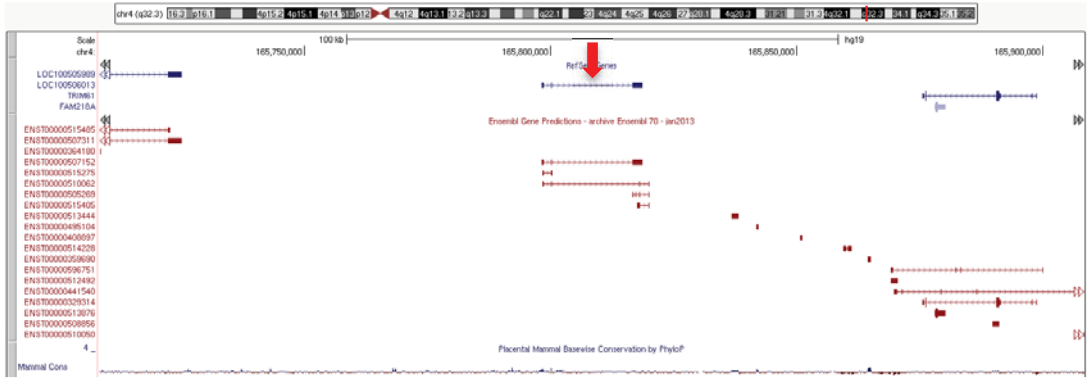
MFCCCYPSSTSALLLVLLSLTPLPALAAPALQQRSSVSSIDEALHMAKEDPGATEGDTWTH
RKVDSLLQKPLGQELVSDMDSASALASVLLEALDHPVREKMAKGNSEEALLIEATRNHAV
MEQDMEKSSESVHQGEENGEQEKPKDGSENSVNTVEAVDGELQSQSDVASSLQRKAR
GYFQNIDLGLQDNEILPPLKGYKAYNQQLAQATKKLKWQEERTKSPSQLERNFMDDFDF
VEEEEEEILTRKEEEARARAEQEEVRRQEAEAQEAREEEQRLADIASDMLLEYMGRKQKS
SSYMRDMNEVALLNNVAEDKRSNEVEDSDDDDEMDPQTIDKLIEISSKLHLPADDVVEII
SDVQEKKKKRKDLPQNNTPRFRPLVPLKAKPRPQAYQYPKSPKKSSYTVDPYKKWYKEKN
KSKFSKQDYWFKPQKQFLAYPSFPYYQKPYRAYYPVYFPAPKPRFYTNPALSFDYNFGGS
MGYGLKPPNRRYRDRPKTRDWKWAQAPQRSQSPYPQPDTVISNYILPHPRTYQALPMP
KPRSPPSGRAPSYIYPPDYVYDEPESPQESDEELENFIEKIYYNRRLF*

A



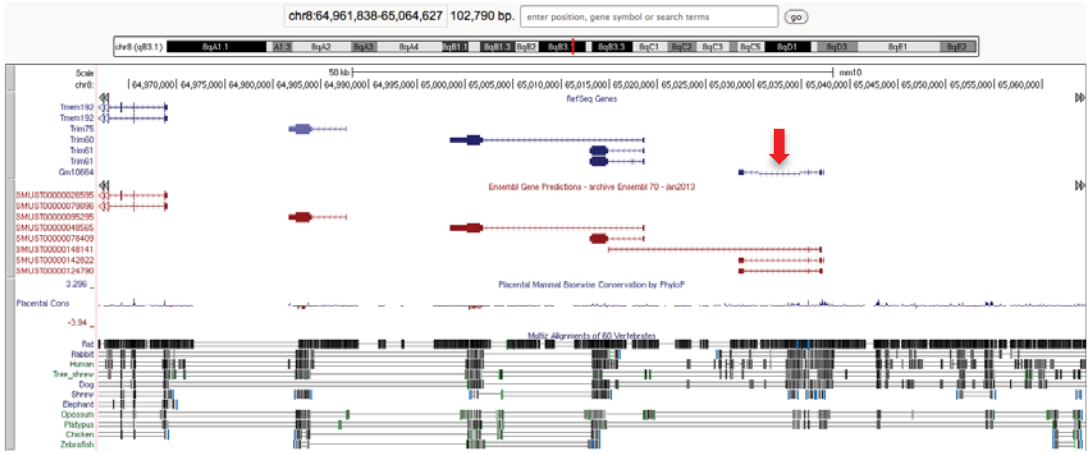
Danio rerio
ENSDARG00000094729, si:ch211-42i9.4
non-coding RNA

B



Homo sapiens
uncharacterized *LOC100506013*
non-coding RNA

C



Mus musculus
predicted gene 10664 (*Gm10664*)
non-coding RNA

fig. S3

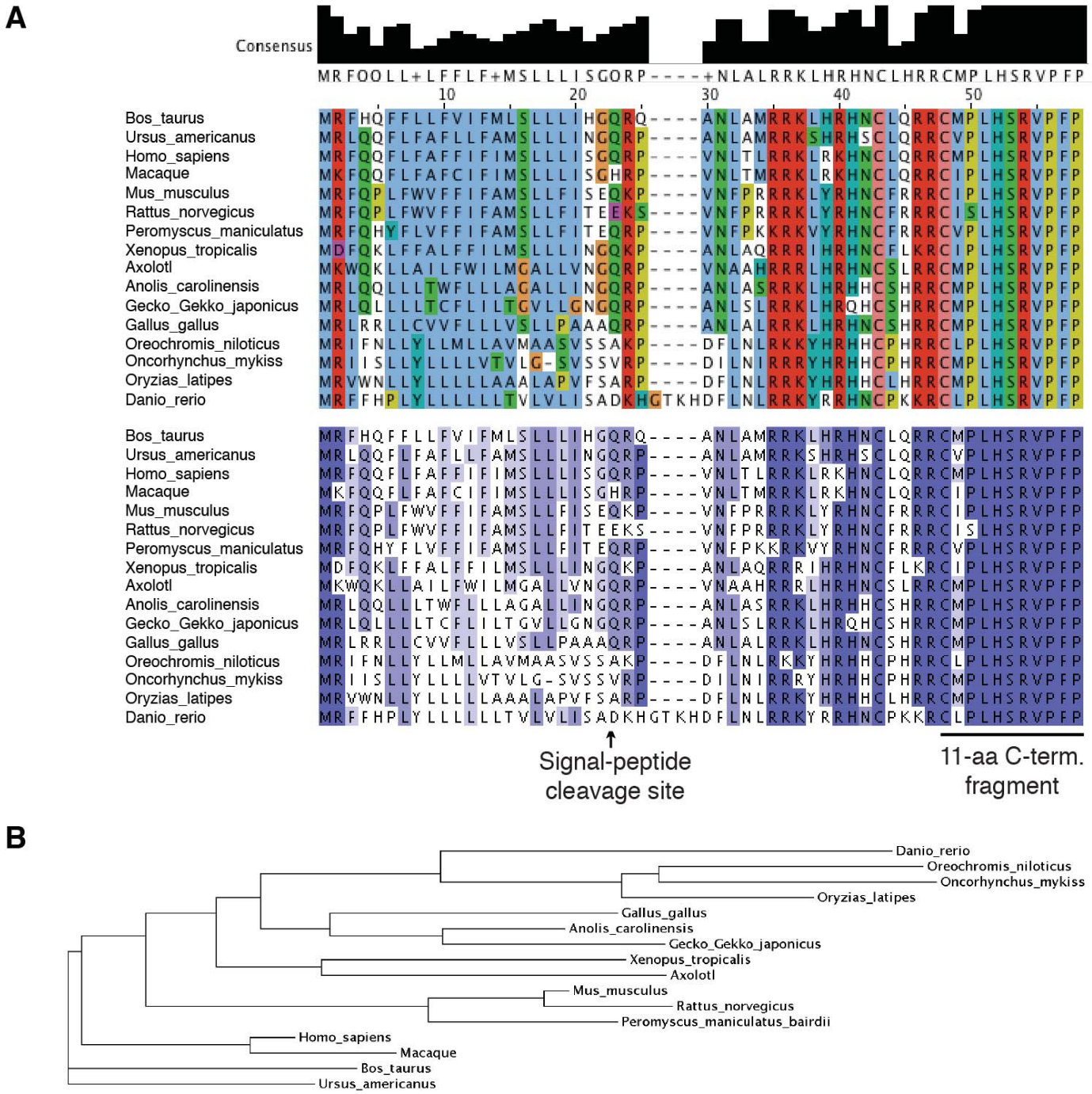
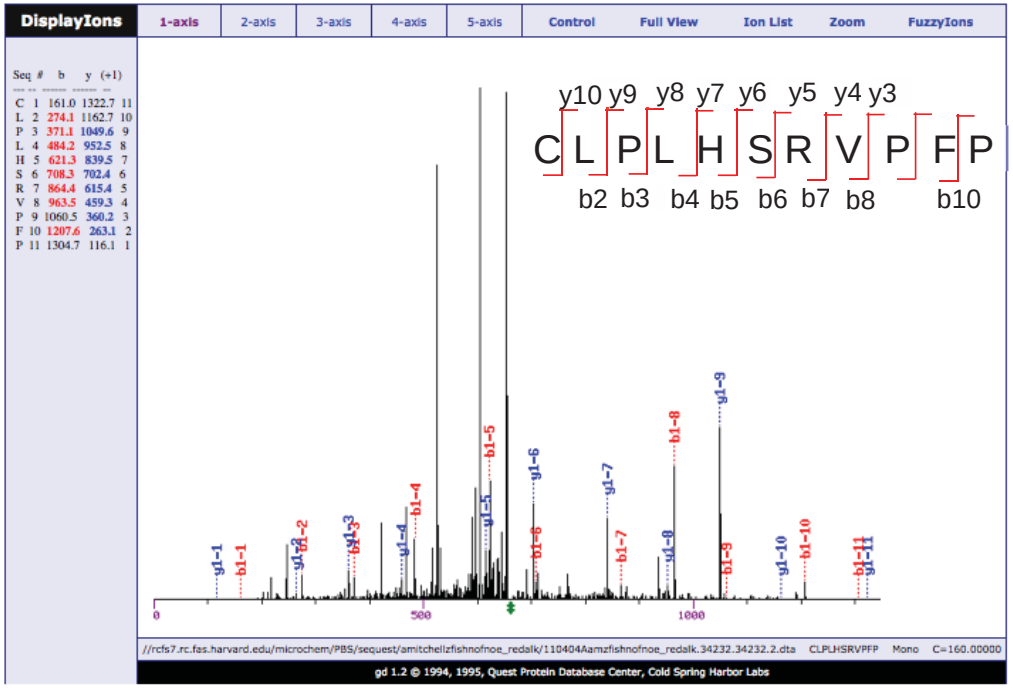
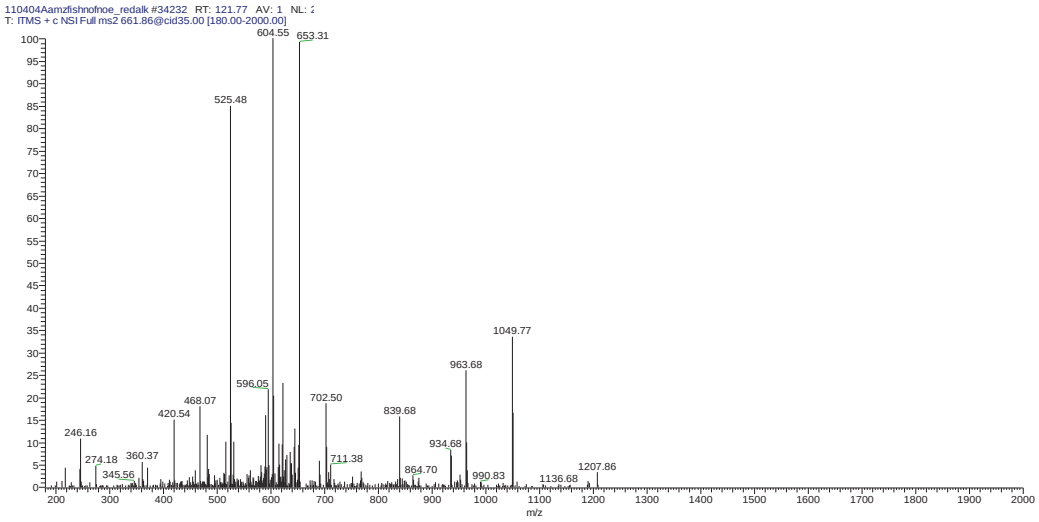

A
Consensus
MRFQOLL+LFFLF+MSLLLISGQRP----+NLALRRKLHRHNCLHRRCMPLHSRVPFP
10
20
30
40
50
Bos_taurus
Ursus_americanus
Homo_sapiens
Macaque
Mus_musculus
Rattus_norvegicus
Peromyscus_maniculatus
Xenopus_tropicalis
Axolotl
Anolis_carolinensis
Gecko_Gekko_japonicus
Gallus_gallus
Oreochromis_niloticus
Oncorhynchus_mykiss
Oryzias_latipes
Danio_rerio
Signal-peptide cleavage site
11-aa C-term. fragment
B
Danio_rerio
Oreochromis_niloticus
Oncorhynchus_mykiss
Oryzias_latipes
Gallus_gallus
Anolis_carolinensis
Gecko_Gekko_japonicus
Xenopus_tropicalis
Axolotl
Mus_musculus
Rattus_norvegicus
Peromyscus_maniculatus_bairdii
Homo_sapiens
Macaque
Bos_taurus
Ursus_americanus

fig. S4

CLPLHSRVPFP

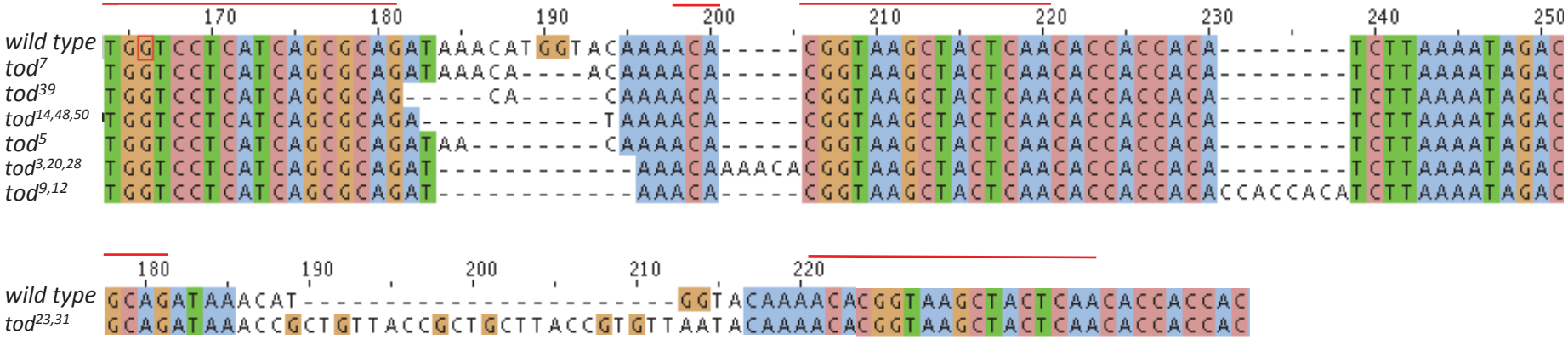
Sf=0.94 TIC=3.5e5





A gDNA

B TALEN-induced *toddler* mutations (DNA)



C Toddler protein alleles

	Signal peptide	secreted propeptide
Toddler wild type	MRFFHPLYLLLLLLTVLVLISA	DKHGTKHDFLNLRRKYRRHNCPKKRCLPLHSRVPFP*
Toddler^{7}	MRFFHPLYLLLLLLTVLVLISA	DKQQNTIFST*
Toddler^{39}	MRFFHPLYLLLLLLTVLVLISA	AQNTIFST*
Toddler^{14,48,50}	MRFFHPLYLLLLLLTVLVLISA	DKTRFSQLEAEISQTQLPEETLSTSSLQSTFPLRFYDAPGKH*
Toddler^{5}	MRFFHPLYLLLLLLTVLVLISA	DNKTRFSQLEAEISQTQLPEETLSTSSLQSTFPLRFYDAPGKH*
Toddler$^{3,\ 20,\ 28}$	MRFFHPLYLLLLLLTVLVLISA	DKQNTIFST*
Toddler$^{9,\ 12}$	MRFFHPLYLLLLLLTVLVLISA	DKHGKLLNTTTPPHLKIDKF*
Toddler$^{23,\ 31}$	MRFFHPLYLLLLLLTVLVLISA	DKPLLPLLTVLIQNTIFST*

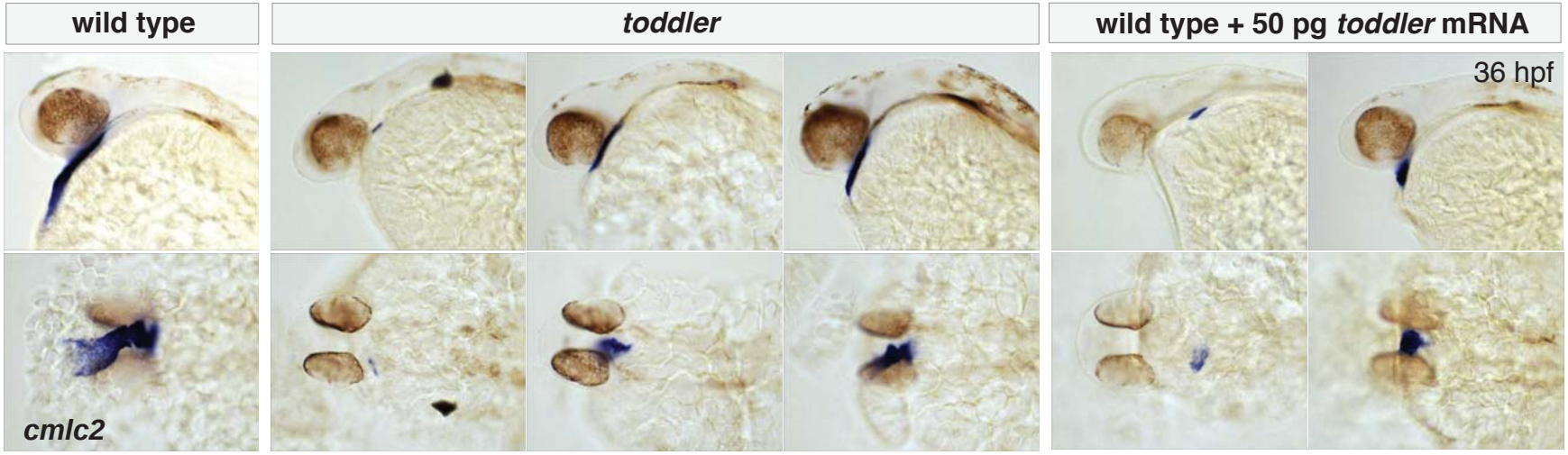
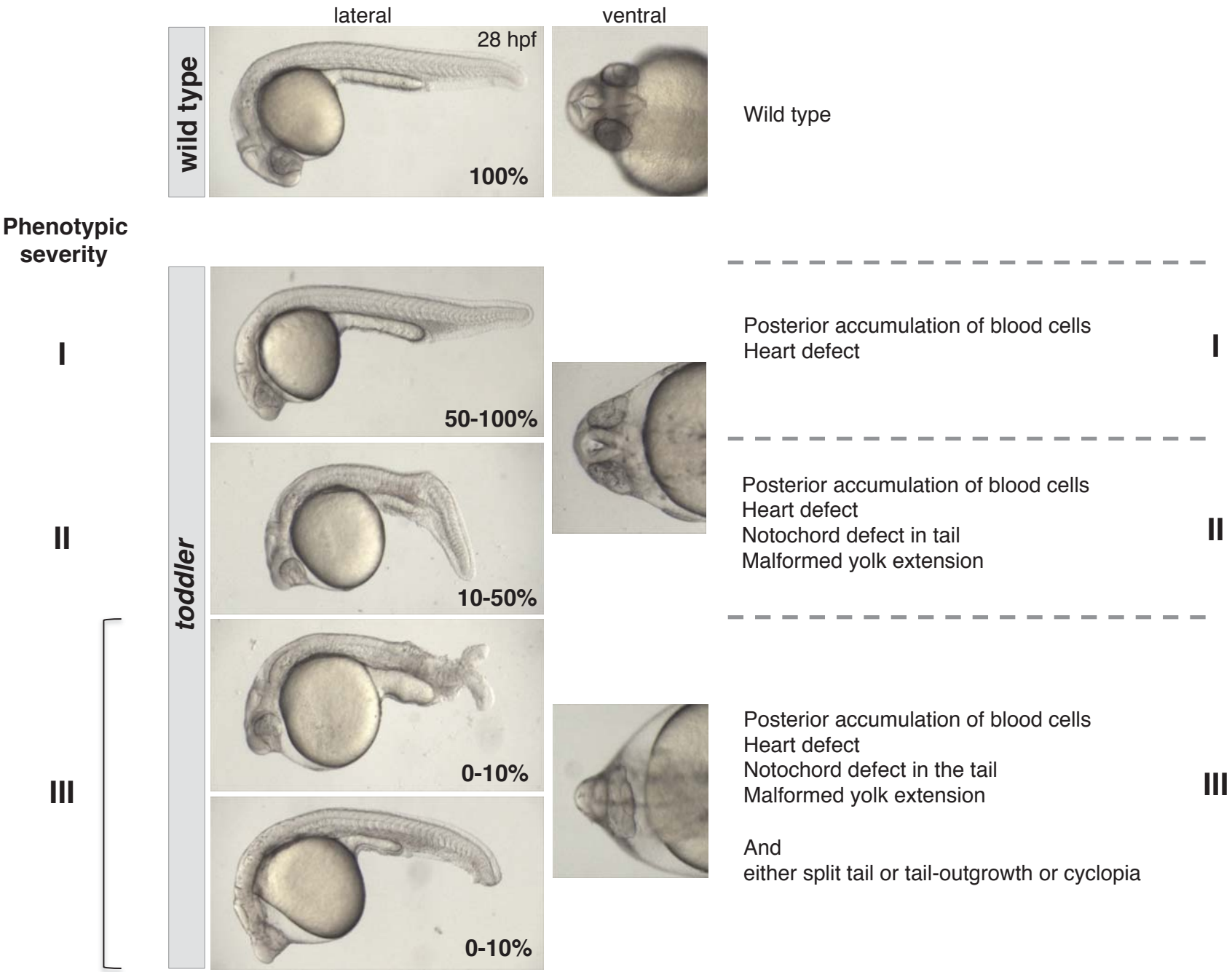
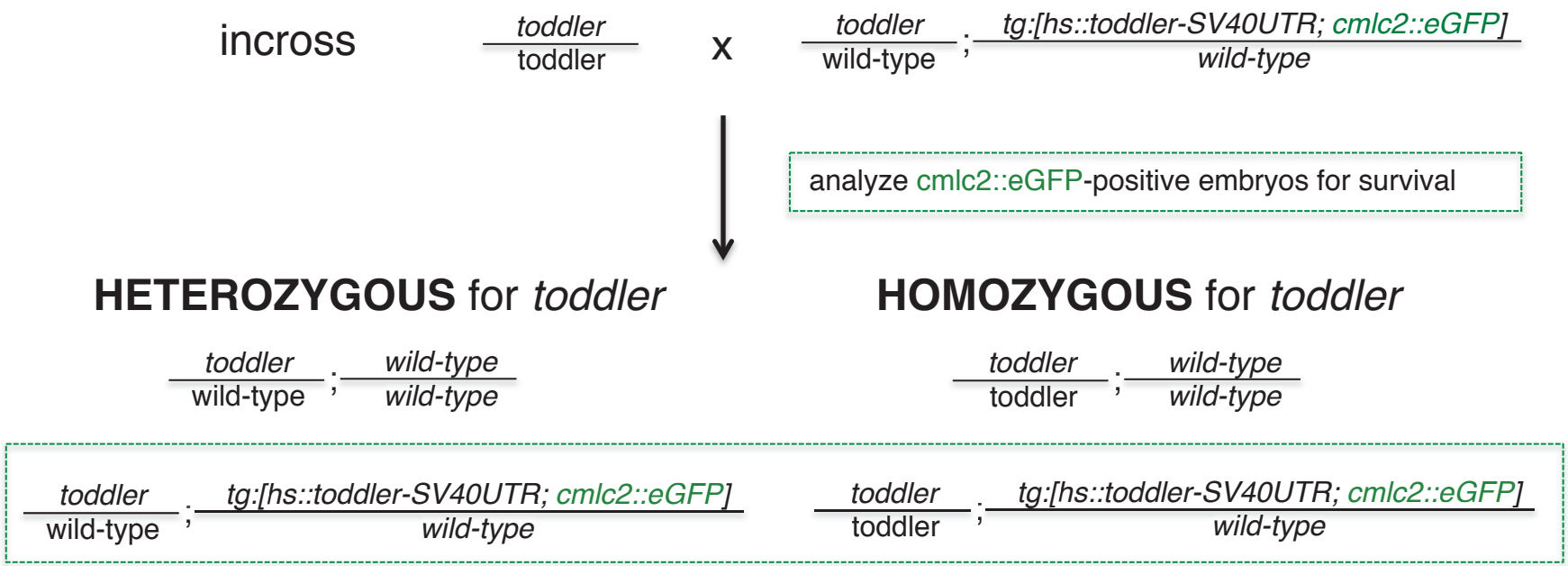

wild type
toddler
wild type + 50 pg toddler mRNA
36 hpf
cmlc2

fig. S7


lateral
ventral
wild type
28 hpf
100%
Wild type
Phenotypic severity
toddler
50-100%
Posterior accumulation of blood cells
Heart defect
I
I
10-50%
Posterior accumulation of blood cells
Heart defect
Notochord defect in tail
Malformed yolk extension
II
II
0-10%
Posterior accumulation of blood cells
Heart defect
Notochord defect in the tail
Malformed yolk extension
III
III
0-10%
And
either split tail or tail-outgrowth or cyclopia

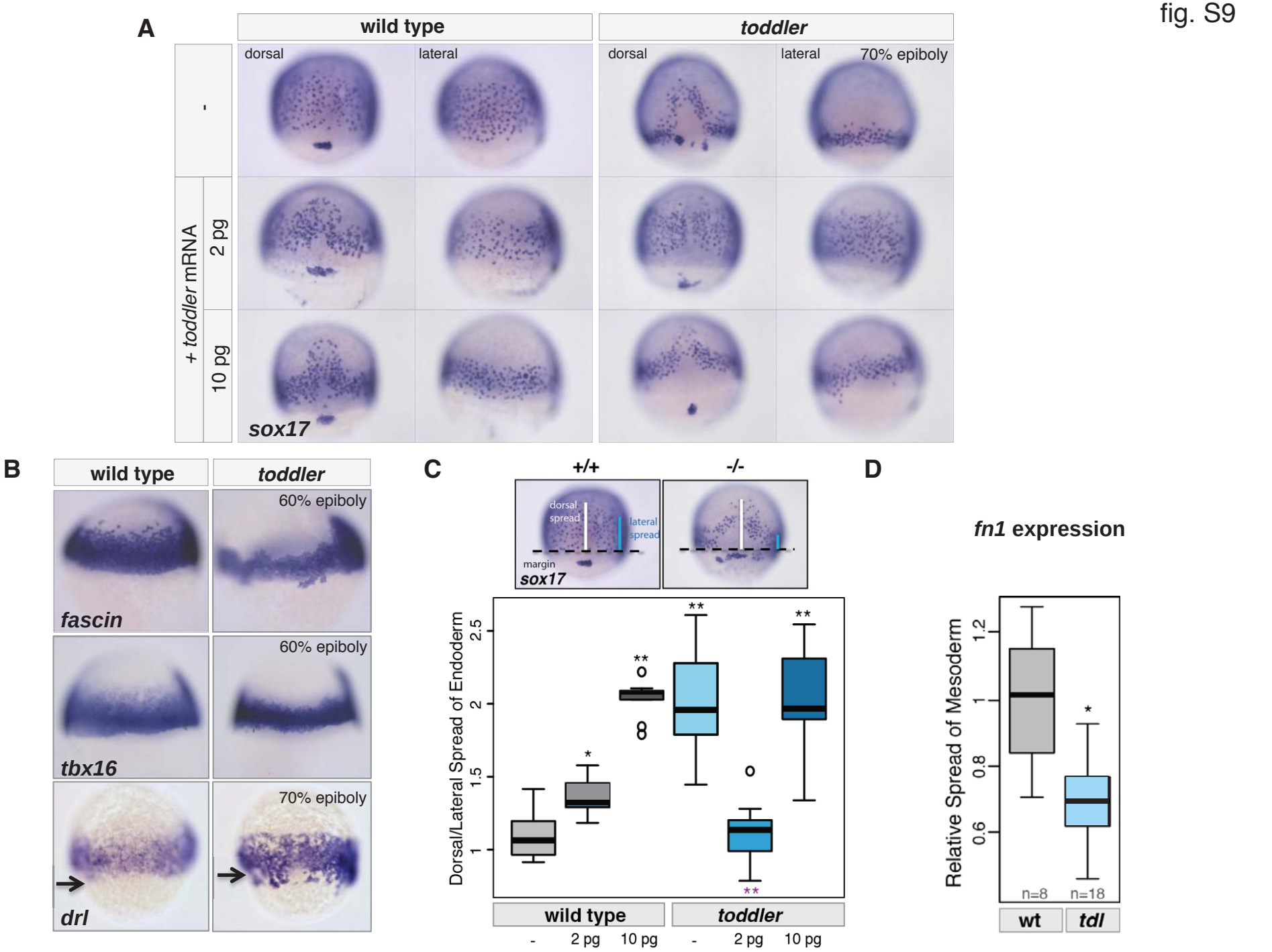
A

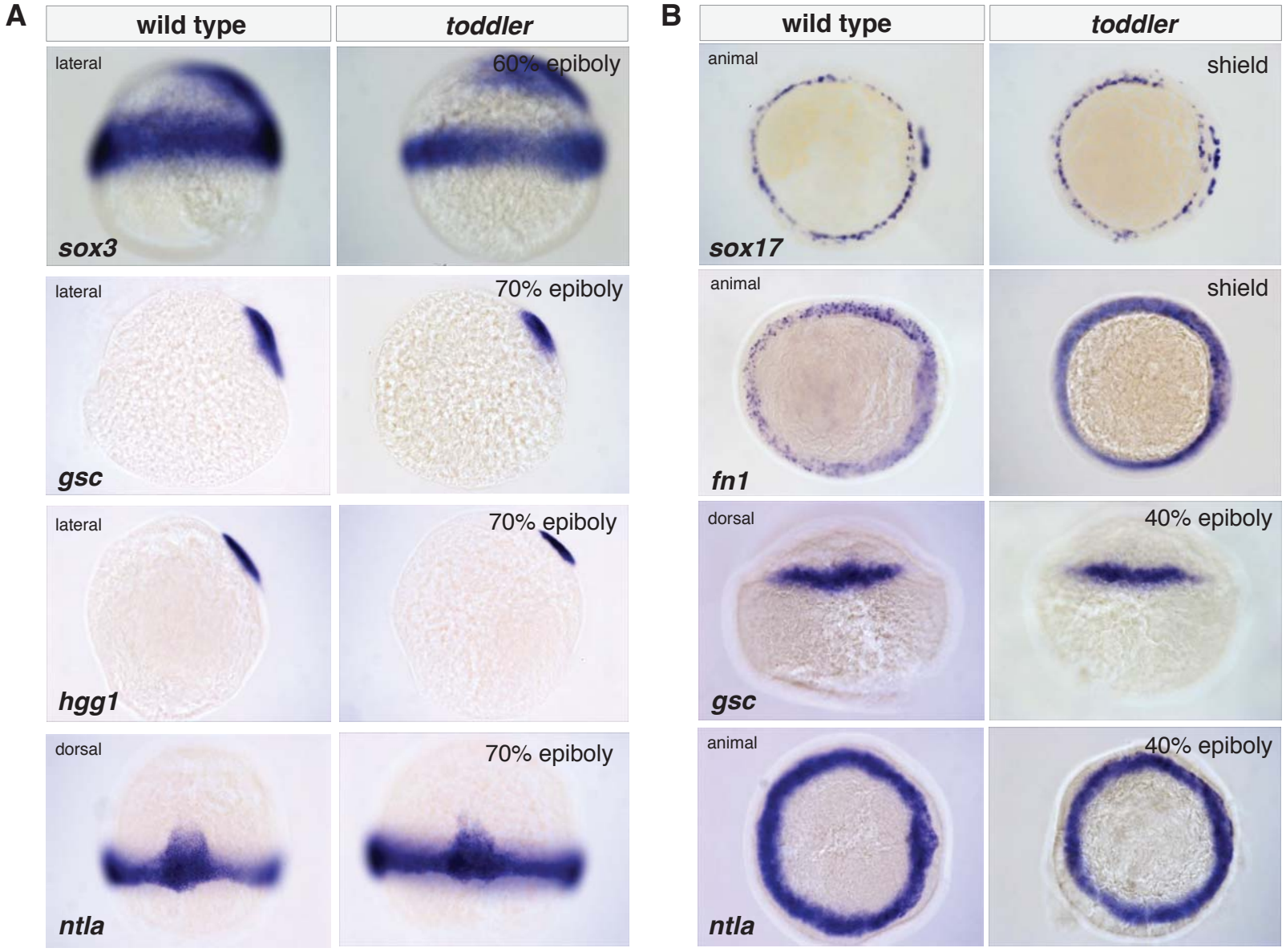


B

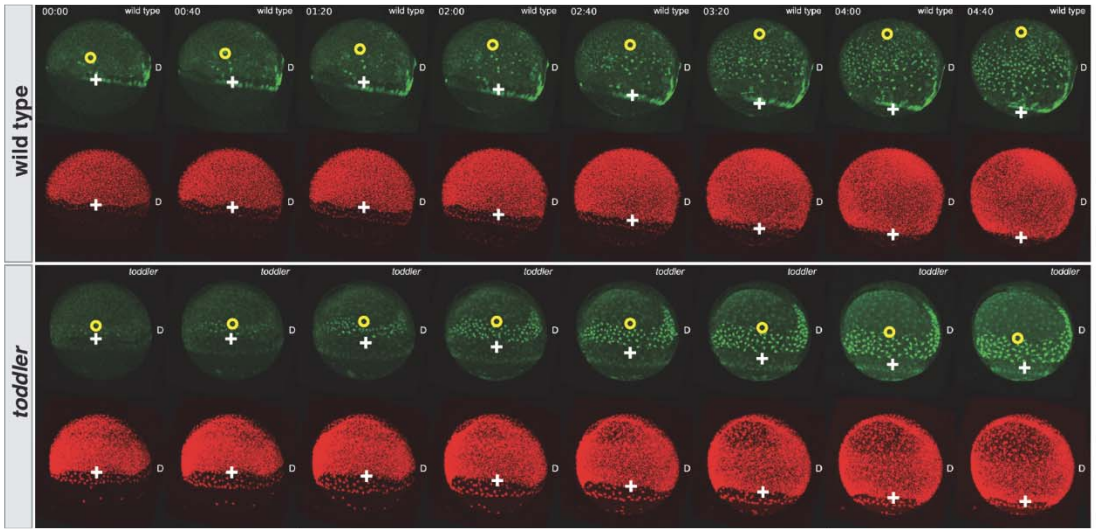
	At 60 hpf (hours post fertilization)	
time of heat-shock (hs for 45 min)	**number of cmlc2::GFP⁺ embryos**	**number of cmlc2::GFP⁺ survivors (normal circulation + heart beat)**
-	25	13
40% epiboly	29	25
50% epiboly	30	30
75% epiboly	28	16

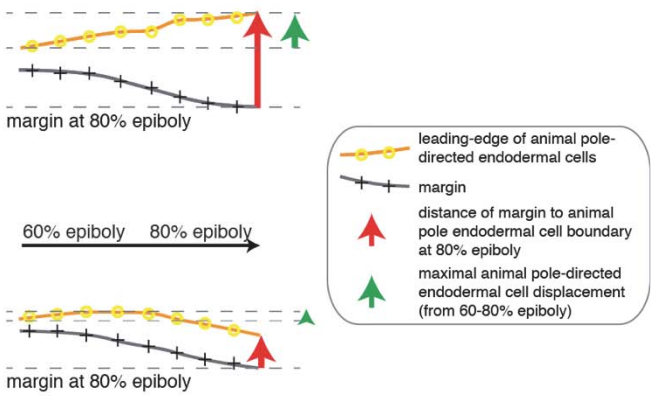
fig. S9


A
wild type
toddler
dorsal
lateral
dorsal
lateral
70% epiboly
-
+ toddler mRNA
2 pg
10 pg
sox17
B
wild type
toddler
60% epiboly
fascin
60% epiboly
tbx16
70% epiboly
drl
C
+/+
-/-
dorsal spread
lateral spread
margin
sox17
Dorsal/Lateral Spread of Endoderm
wild type
toddler
-
2 pg
10 pg
-
2 pg
10 pg
D
fn1 expression
Relative Spread of Mesoderm
n=8
n=18
wt
tdl


A
wild type
toddler
lateral
60% epiboly
sox3
lateral
70% epiboly
gsc
lateral
70% epiboly
hgg1
dorsal
70% epiboly
ntla
B
wild type
toddler
animal
shield
sox17
animal
shield
fn1
dorsal
40% epiboly
gsc
animal
40% epiboly
ntla

A


wild type
toddler
00:00
00:40
01:20
02:00
02:40
03:20
04:00
04:40
D


margin at 80% epiboly
60% epiboly
80% epiboly
margin at 80% epiboly
leading-edge of animal pole-directed endodermal cells
margin
distance of margin to animal pole endodermal cell boundary at 80% epiboly
maximal animal pole-directed endodermal cell displacement (from 60-80% epiboly)

B

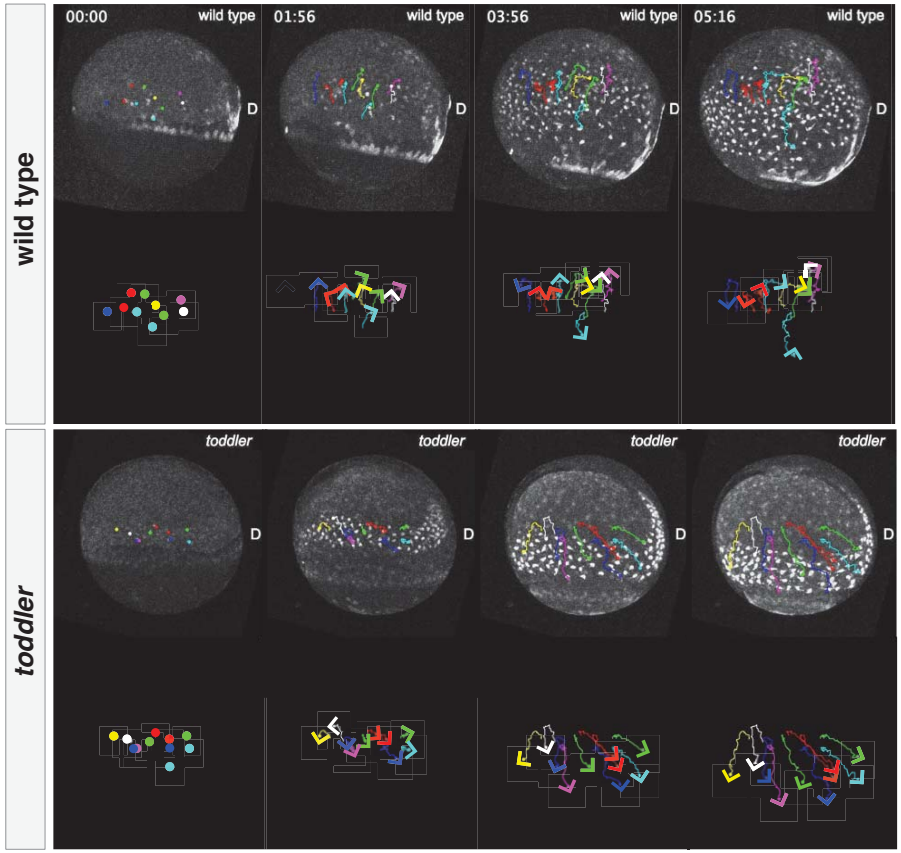
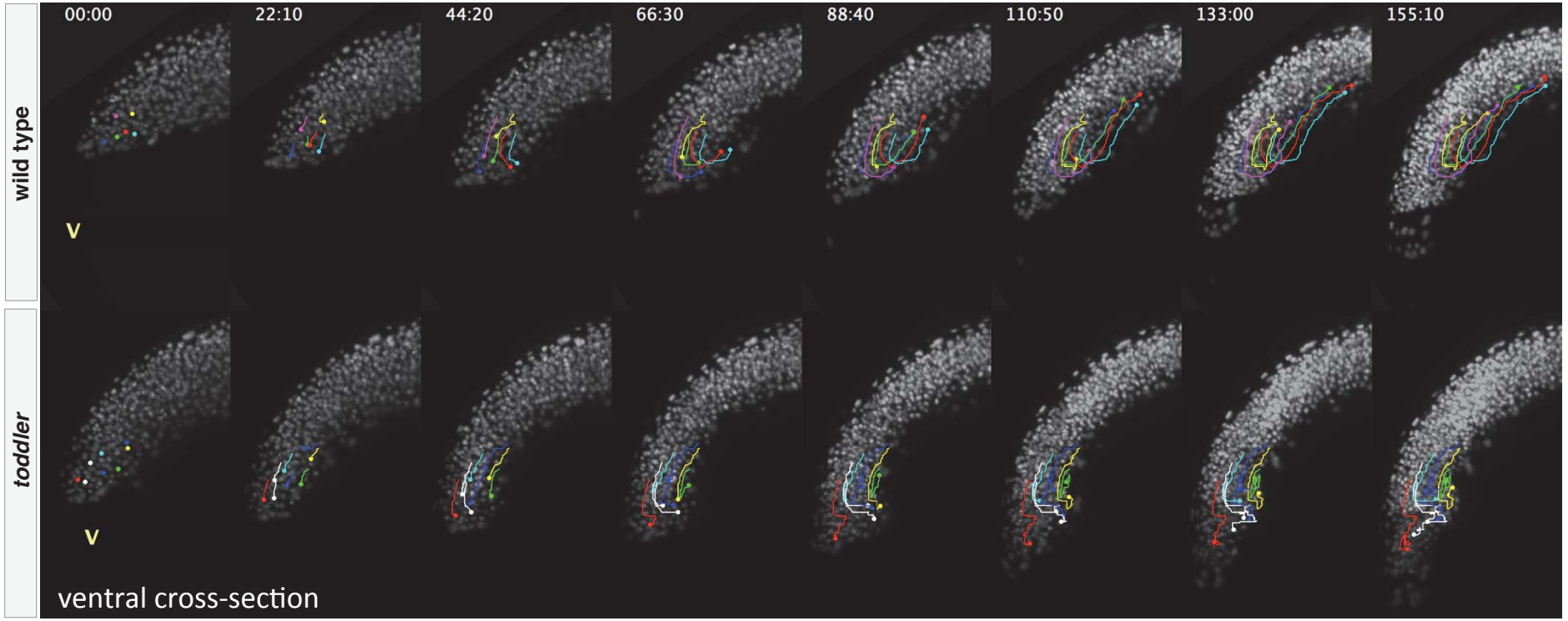
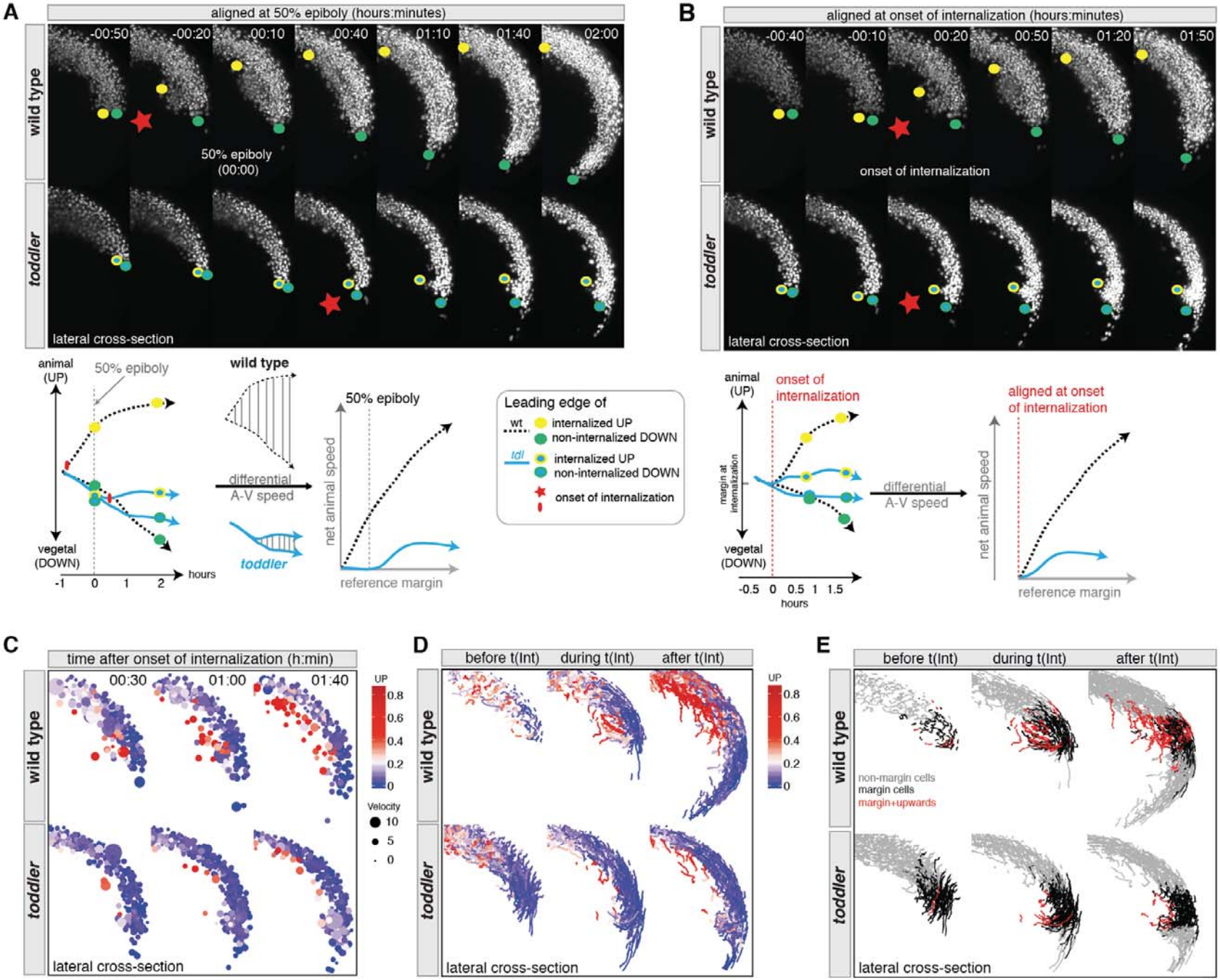
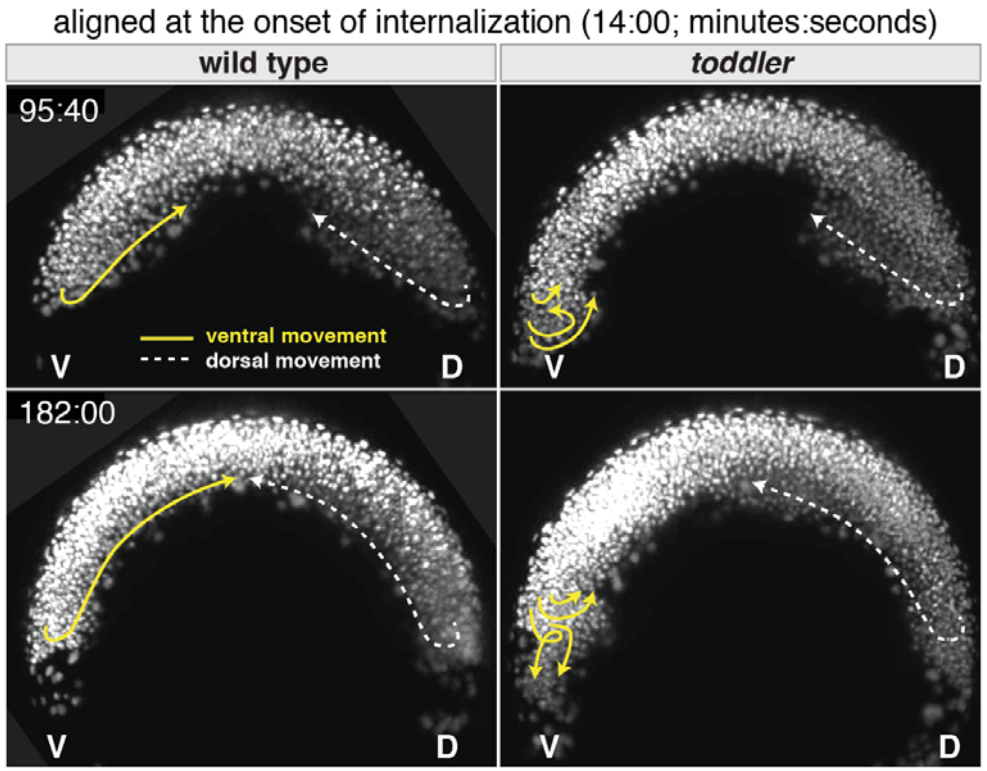

wild type
toddler
00:00
01:56
03:56
05:16
D

fig. S12

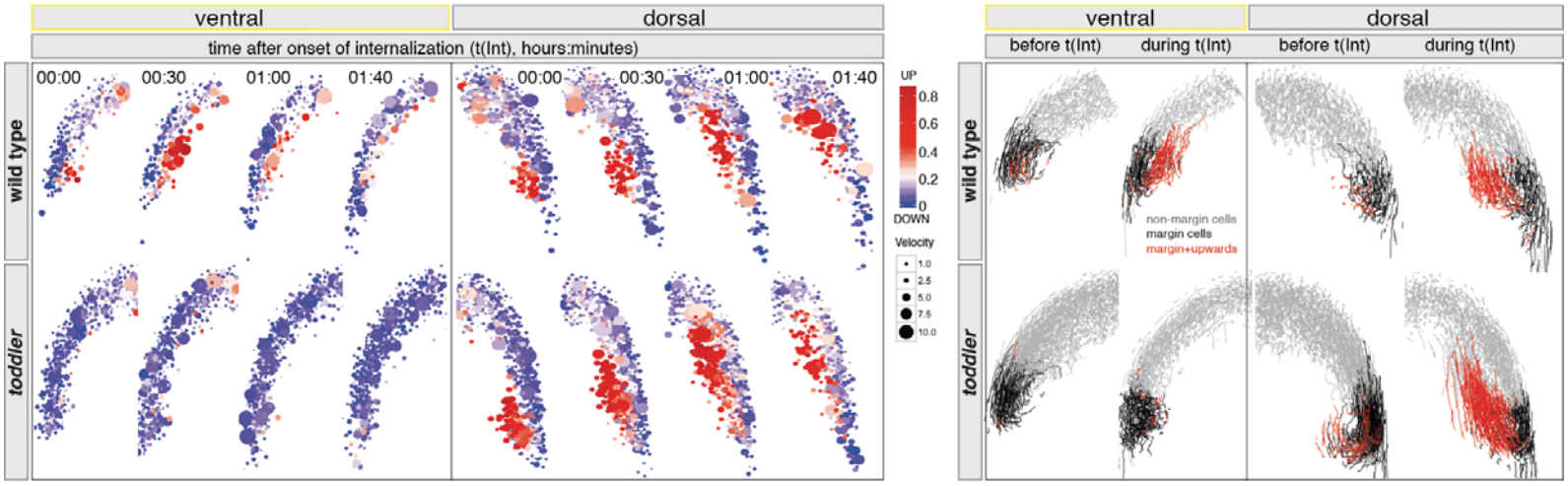

wild type
toddler
00:00
22:10
44:20
66:30
88:40
110:50
133:00
155:10
V
V
ventral cross-section

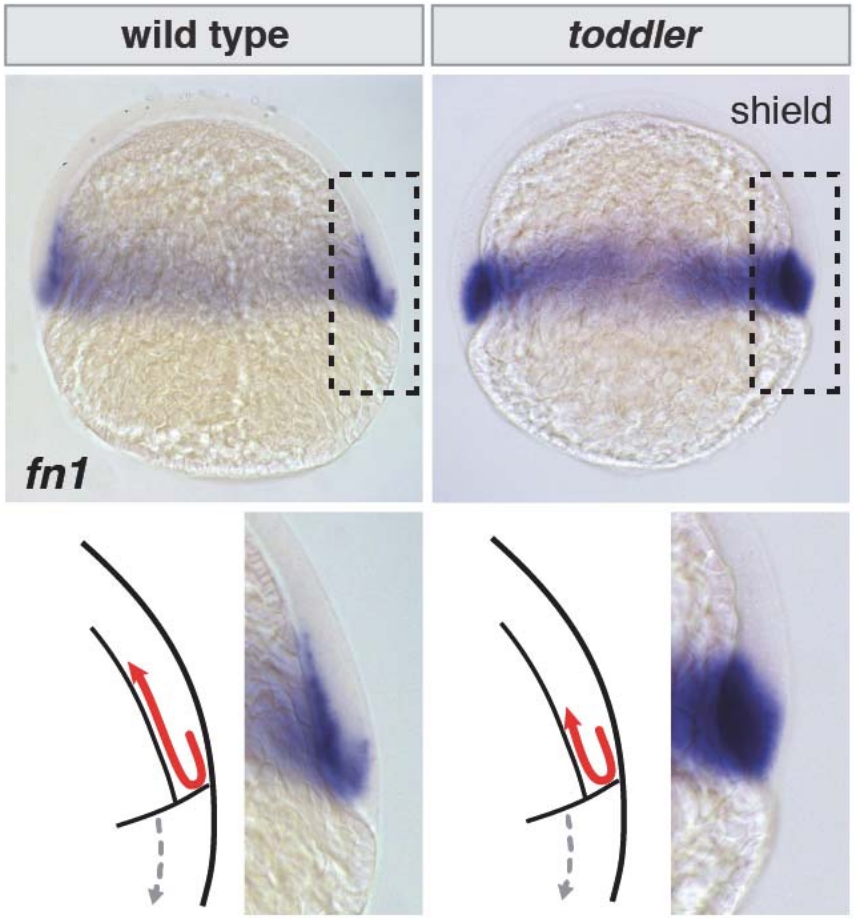

A
aligned at 50% epiboly (hours:minutes)
wild type
toddler
-00:50
-00:20
00:10
00:40
01:10
01:40
02:00
50% epiboly
(00:00)
lateral cross-section
B
aligned at onset of internalization (hours:minutes)
wild type
toddler
-00:40
-00:10
00:20
00:50
01:20
01:50
onset of internalization
lateral cross-section
animal
(UP)
50% epiboly
vegetal
(DOWN)
hours
-1
0
1
2
wild type
differential
A-V speed
toddler
50% epiboly
net animal speed
reference margin
Leading edge of
wt
internalized UP
non-internalized DOWN
tdl
internalized UP
non-internalized DOWN
onset of internalization
animal
(UP)
onset of
internalization
margin at
internalization
vegetal
(DOWN)
-0.5
0
0.5
1
1.5
hours
differential
A-V speed
aligned at onset
of internalization
net animal speed
reference margin
C
time after onset of internalization (h:min)
wild type
toddler
00:30
01:00
01:40
lateral cross-section
UP
0.8
0.6
0.4
0.2
0
Velocity
10
5
0
D
before t(Int)
during t(Int)
after t(Int)
wild type
toddler
lateral cross-section
UP
0.8
0.6
0.4
0.2
0
E
before t(Int)
during t(Int)
after t(Int)
wild type
toddler
non-margin cells
margin cells
margin+upwards
lateral cross-section

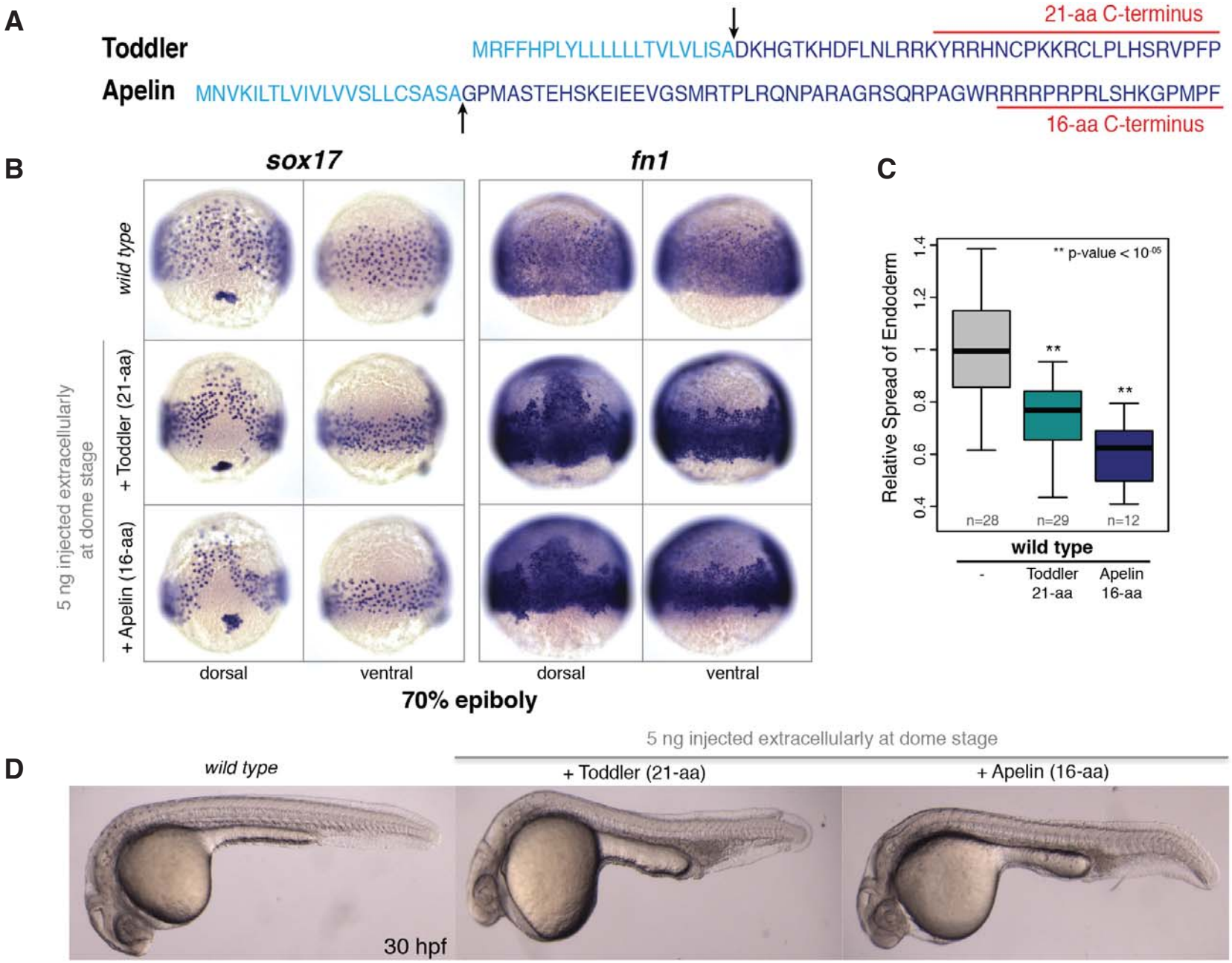
A

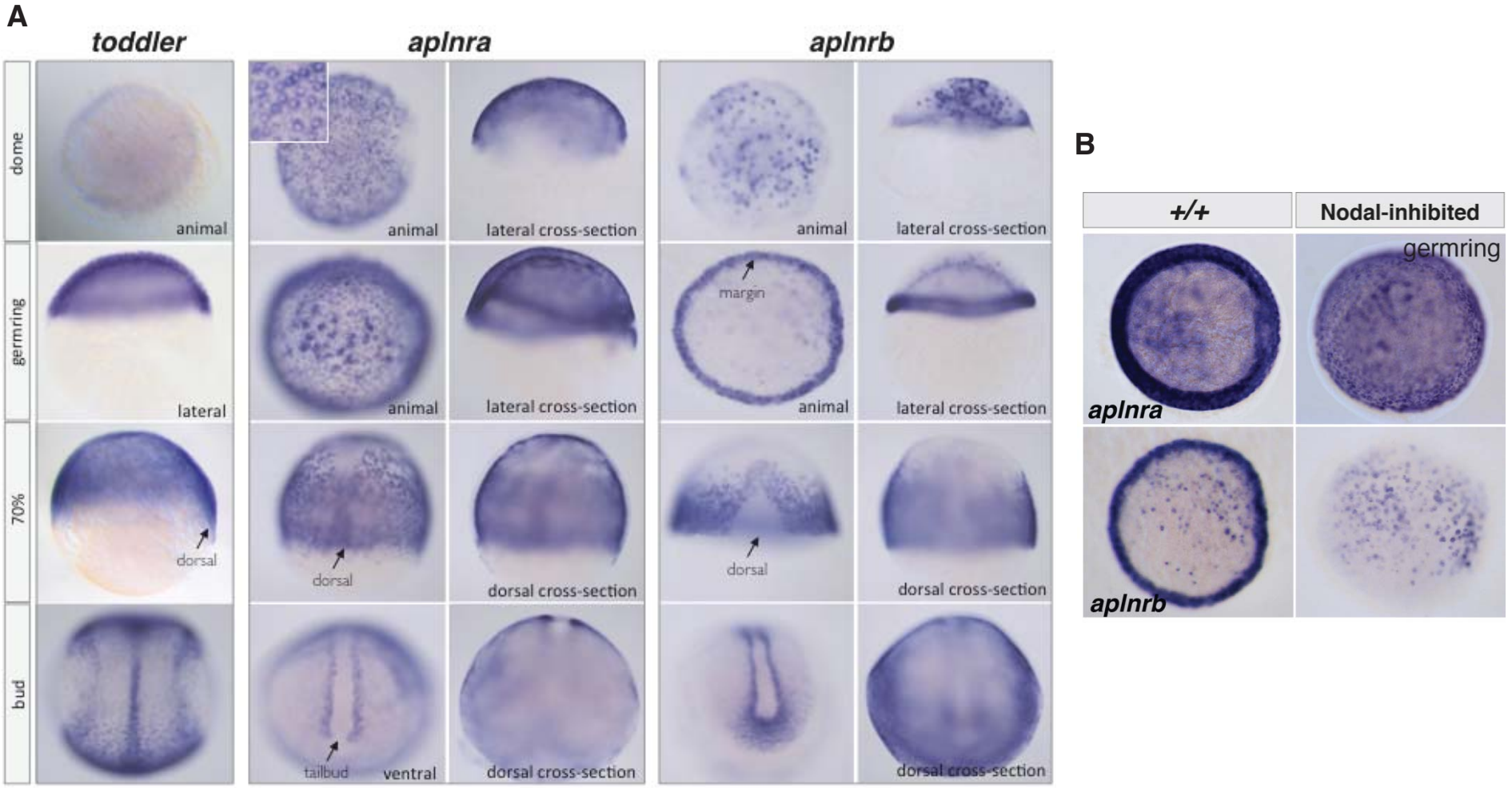

aligned at the onset of internalization (14:00; minutes:seconds)
wild type
toddler
95:40
182:00
ventral movement
dorsal movement
V
D

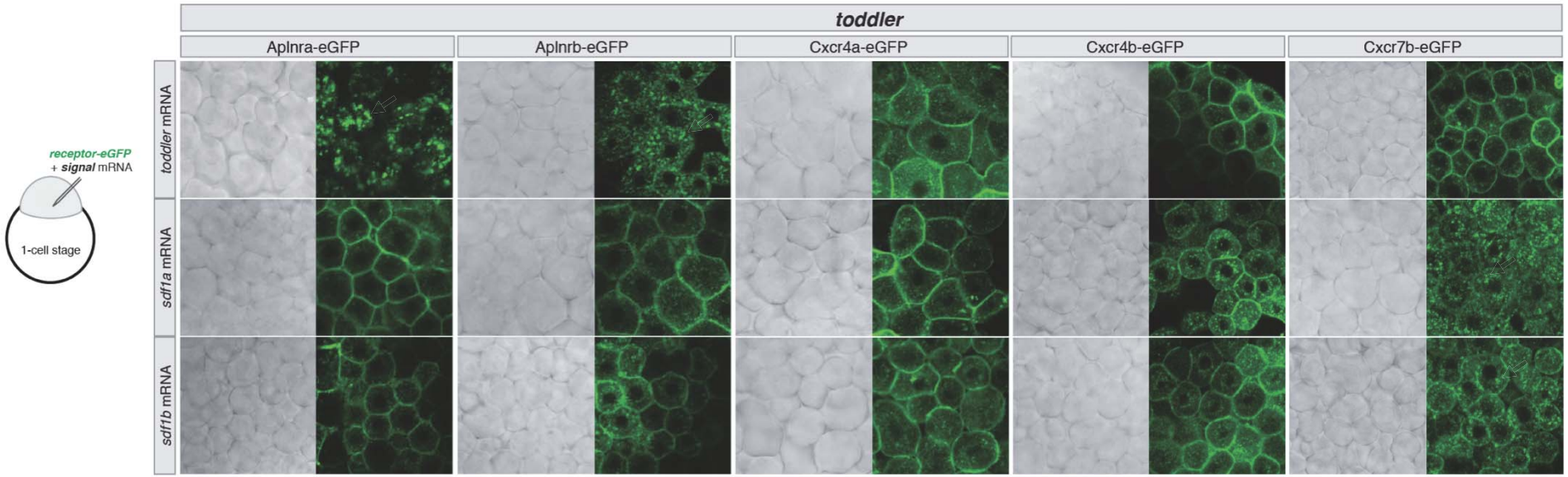
B

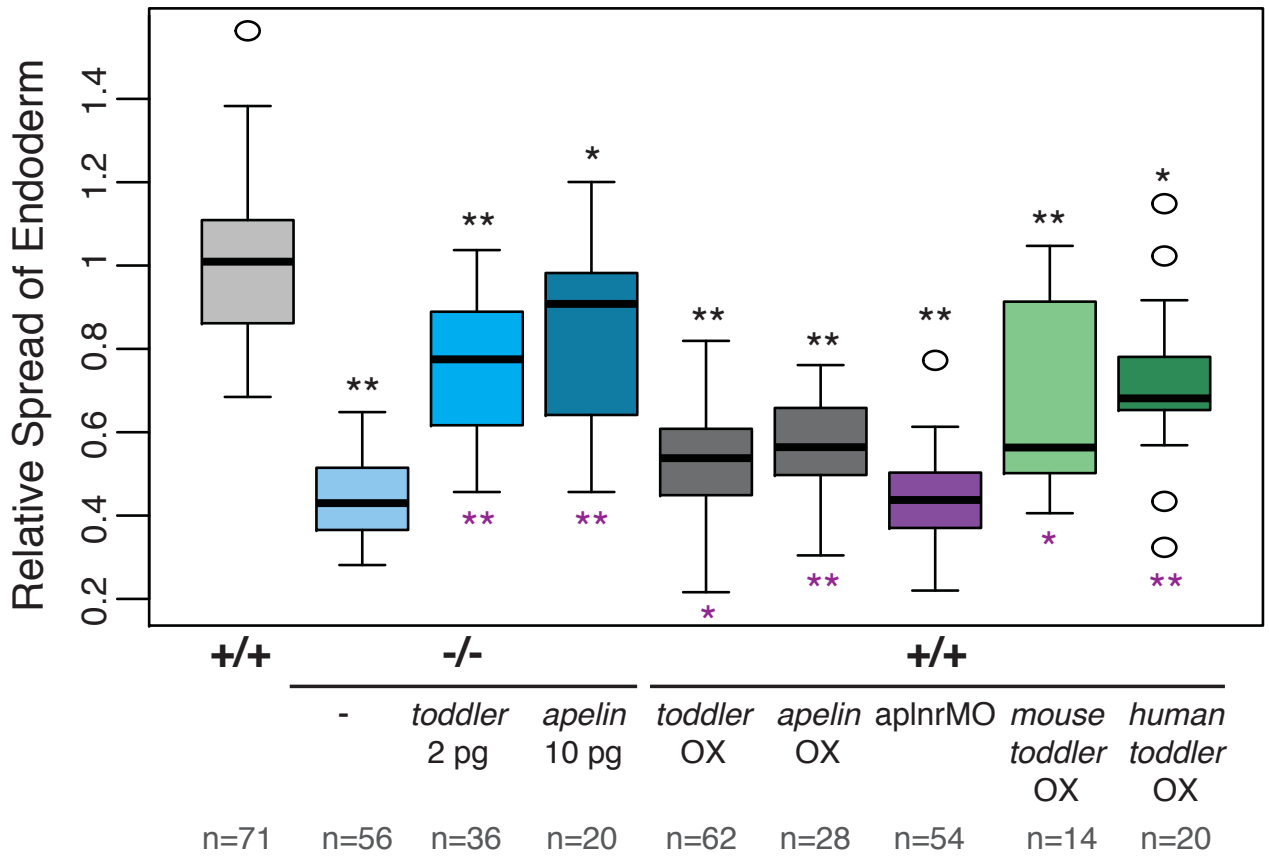

ventral
dorsal
time after onset of internalization (t(Int), hours:minutes)
00:00
00:30
01:00
01:40
wild type
toddler
UP
0.8
0.6
0.4
0.2
0
DOWN
Velocity
1.0
2.5
5.0
7.5
10.0
before t(Int)
during t(Int)
non-margin cells
margin cells
margin+upwards


wild type
toddler
shield
fn1


A
Toddler
MRFFHPLYLLLLLLTVLVLISADKHGTKHDFLNLRRKYRRHNCPKKRCLPLHSRVPFP
21-aa C-terminus
Apelin
MNVKILTLVIVLVVSLLCSASAGPMASTEHSKEIEEVGSMRTPLRQNPARAGRSQRPAGWRRRRPRPRLSHKGPMPF
16-aa C-terminus
B
sox17
fn1
wild type
+ Toddler (21-aa)
+ Apelin (16-aa)
5 ng injected extracellularly at dome stage
dorsal
ventral
dorsal
ventral
70% epiboly
C
Relative Spread of Endoderm
** p-value < 10-05
n=28
n=29
n=12
wild type
-
Toddler 21-aa
Apelin 16-aa
D
5 ng injected extracellularly at dome stage
wild type
+ Toddler (21-aa)
+ Apelin (16-aa)
30 hpf


A
toddler
aplnra
aplnrb
dome
germring
70%
bud
animal
lateral
dorsal
tailbud
ventral
lateral cross-section
dorsal cross-section
margin
B
+/+
Nodal-inhibited
germring
aplnra
aplnrb


toddler
Aplnra-eGFP
Aplnrb-eGFP
Cxcr4a-eGFP
Cxcr4b-eGFP
Cxcr7b-eGFP
toddler mRNA
sdf1a mRNA
sdf1b mRNA
receptor-eGFP
+ signal mRNA
1-cell stage


Relative Spread of Endoderm
0.2
0.4
0.6
0.8
1
1.2
1.4
+/+
-/-
+/+
-
toddler 2 pg
apelin 10 pg
toddler OX
apelin OX
aplnrMO
mouse toddler OX
human toddler OX
n=71
n=56
n=36
n=20
n=62
n=28
n=54
n=14
n=20

References and Notes

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, ed. 5, 2007).

2. C.-P. Heisenberg, Y. Bellaïche, Forces in tissue morphogenesis and patterning. *Cell* **153**, 948–962 (2013). doi:10.1016/j.cell.2013.05.008 Medline

3. L. Solnica-Krezel, D. S. Sepich, Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* **28**, 687–717 (2012). doi:10.1146/annurev-cellbio-092910-154043 Medline

4. J. B. Wallingford, Planar cell polarity and the developmental control of cell behavior in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **28**, 627–653 (2012). doi:10.1146/annurev-cellbio-092910-154208 Medline

5. S. Nowotschin, A.-K. Hadjantonakis, Cellular dynamics in the early mouse embryo: from axis formation to gastrulation. *Curr. Opin. Genet. Dev.* **20**, 420–427 (2010). doi:10.1016/j.gde.2010.05.008 Medline

6. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011). doi:10.1016/j.cell.2011.10.002 Medline

7. G.-L. Chew, A. Pauli, J. L. Rinn, A. Regev, A. F. Schier, E. Valen, Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development* **140**, 2828–2834 (2013). doi:10.1242/dev.098343 Medline

8. A. Pauli, E. Valen, M. F. Lin, M. Garber, N. L. Vastenhouw, J. Z. Levin, L. Fan, A. Sandelin, J. L. Rinn, A. Regev, A. F. Schier, Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591 (2012). doi:10.1101/gr.133009.111 Medline

9. I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, D. P. Bartel, Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011). doi:10.1016/j.cell.2011.11.055 Medline

10. A. S. Hassan, J. Hou, W. Wei, P. A. Hoodless, Expression of two novel transcripts in the mouse definitive endoderm. *Gene Expr. Patterns* **10**, 127–134 (2010). doi:10.1016/j.gep.2010.02.001 Medline

11. M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root, E. S. Lander, lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011). doi:10.1038/nature10398 Medline

12. D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, J. K. Joung, FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460–465 (2012). doi:10.1038/nbt.2170 Medline

13. J. D. Sander, L. Cade, C. Khayter, D. Reyon, R. T. Peterson, J. K. Joung, J. R. Yeh, Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698 (2011). doi:10.1038/nbt.1934 Medline

14. T. Mizoguchi, H. Verkade, J. K. Heath, A. Kuroiwa, Y. Kikuchi, Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521–2529 (2008). doi:10.1242/dev.020107 Medline

15. G. Pézeron, P. Mourrain, S. Courty, J. Ghislain, T. S. Becker, F. M. Rosa, N. B. David, Live analysis of endodermal layer formation identifies random walk as a novel gastrulation movement. *Curr. Biol.* **18**, 276–281 (2008). doi:10.1016/j.cub.2008.01.028 Medline

16. M. Doitsidou, M. Reichman-Fried, J. Stebler, M. Köprunner, J. Dörries, D. Meyer, C. V. Esguerra, T. Leung, E. Raz, Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell* **111**, 647–659 (2002). doi:10.1016/S0092-8674(02)01135-2 Medline

17. M. Haskel-Ittah, D. Ben-Zvi, M. Branski-Arieli, E. D. Schejter, B. Z. Shilo, N. Barkai, Self-organized shuttling: generating sharp dorsoventral polarity in the early Drosophila embryo. *Cell* **150**, 1016–1028 (2012). doi:10.1016/j.cell.2012.06.044 Medline

18. G. Venkiteswaran, S. W. Lewellis, J. Wang, E. Reynolds, C. Nicholson, H. Knaut, Generation and dynamics of an endogenous, self-generated signaling gradient across a migrating tissue. *Cell* **155**, 674–687 (2013). doi:10.1016/j.cell.2013.09.046 Medline

19. E. Donà, J. D. Barry, G. Valentin, C. Quirin, A. Khmelinskii, A. Kunze, S. Durdu, L. R. Newton, A. Fernandez-Minan, W. Huber, M. Knop, D. Gilmour, Directional tissue migration through a self-generated chemokine gradient. *Nature* **503**, 285–289 (2013). Medline

20. E. M. Powell, W. M. Mars, P. Levitt, Hepatocyte growth factor/scatter factor is a motogen for interneurons migrating from the ventral to dorsal telencephalon. *Neuron* **30**, 79–89 (2001). doi:10.1016/S0896-6273(01)00264-1 Medline

21. P. Giacobini, A. Messina, S. Wray, C. Giampietro, T. Crepaldi, P. Carmeliet, A. Fasolo, Hepatocyte growth factor acts as a motogen and guidance signal for gonadotropin hormone-releasing hormone-1 neuronal migration. *J. Neurosci.* **27**, 431–445 (2007). doi:10.1523/JNEUROSCI.4979-06.2007 Medline

22. G. Kirfel, B. Borm, A. Rigort, V. Herzog, The secretory β-amyloid precursor protein is a motogen for human epidermal keratinocytes. *Eur. J. Cell Biol.* **81**, 664–676 (2002). doi:10.1078/0171-9335-00284 Medline

23. S. Nair, T. F. Schilling, Chemokine signaling controls endodermal migration during zebrafish gastrulation. *Science* **322**, 89–92 (2008). doi:10.1126/science.1160038

24. X.-X. I. Zeng, T. P. Wilm, D. S. Sepich, L. Solnica-Krezel, Apelin and its receptor control heart field formation during zebrafish gastrulation. *Dev. Cell* **12**, 391–402 (2007). doi:10.1016/j.devcel.2007.01.011 Medline

25. I. C. Scott, B. Masri, L. A. D'Amico, S. W. Jin, B. Jungblut, A. M. Wehman, H. Baier, Y. Audigier, D. Y. Stainier, The G protein-coupled receptor Agtrl1b regulates early development of myocardial progenitors. *Dev. Cell* **12**, 403–413 (2007). doi:10.1016/j.devcel.2007.01.012 Medline

26. S. Paskaradevan, I. C. Scott, The Aplnr GPCR regulates myocardial progenitor development via a novel cell-non-autonomous, $G\alpha_{i/o}$ protein-independent pathway. *Biol. Open* **1**, 275–285 (2012). doi:10.1242/bio.2012380 Medline

27. X. Li, I. Roszko, D. S. Sepich, M. Ni, H. E. Hamm, F. L. Marlow, L. Solnica-Krezel, Gpr125 modulates Dishevelled distribution and planar cell polarity signaling. *Development* **140**, 3028–3039 (2013). doi:10.1242/dev.094839 Medline

28. F. Lin, S. Chen, D. S. Sepich, J. R. Panizzi, S. G. Clendenon, J. A. Marrs, H. E. Hamm, L. Solnica-Krezel, $G\alpha_{12/13}$ regulate epiboly by inhibiting E-cadherin activity and modulating the actin cytoskeleton. *J. Cell Biol.* **184**, 909–921 (2009). doi:10.1083/jcb.200805148 Medline

29. M. Costa, E. T. Wilson, E. Wieschaus, A putative cell signal encoded by the folded gastrulation gene coordinates cell shape changes during Drosophila gastrulation. *Cell* **76**, 1075–1089 (1994). doi:10.1016/0092-8674(94)90384-0 Medline

30. S. Parks, E. Wieschaus, The Drosophila gastrulation gene concertina encodes a G α-like protein. *Cell* **64**, 447–458 (1991). doi:10.1016/0092-8674(91)90652-F Medline

31. C. D'Aniello, E. Lonardo, S. Iaconis, O. Guardiola, A. M. Liguoro, G. L. Liguori, M. Autiero, P. Carmeliet, G. Minchiotti, G protein-coupled receptor APJ and its ligand apelin act downstream of Cripto to specify embryonic stem cells toward the cardiac lineage through extracellular signal-regulated kinase/p70S6 kinase signaling pathway. *Circ. Res.* **105**, 231–238 (2009). doi:10.1161/CIRCRESAHA.109.201186 Medline

32. I.-N. E. Wang, X. Wang, X. Ge, J. Anderson, M. Ho, E. Ashley, J. Liu, M. J. Butte, M. Yazawa, R. E. Dolmetsch, T. Quertermous, P. C. Yang, Apelin enhances directed cardiac differentiation of mouse and human embryonic stem cells. *PLOS ONE* **7**, e38328 (2012). doi:10.1371/journal.pone.0038328 Medline

33. D. Tempel, M. de Boer, E. D. van Deel, R. A. Haasdijk, D. J. Duncker, C. Cheng, S. Schulte-Merker, H. J. Duckers, Apelin enhances cardiac neovascularization after myocardial infarction by recruiting $Aplnr^+$ circulating cells. *Circ. Res.* **111**, 585–598 (2012). doi:10.1161/CIRCRESAHA.111.262097 Medline

34. Y. Kang, J. Kim, J. P. Anderson, J. Wu, S. R. Gleim, R. K. Kundu, D. L. McLean, J. D. Kim, H. Park, S. W. Jin, J. Hwa, T. Quertermous, H. J. Chun, Apelin-APJ signaling is a critical regulator of endothelial MEF2 activation in cardiovascular development. *Circ. Res.* **113**, 22–31 (2013). doi:10.1161/CIRCRESAHA.113.301324 Medline

35. M. Inui, A. Fukui, Y. Ito, M. Asashima, Xapelin and Xmsr are required for cardiovascular development in Xenopus laevis. *Dev. Biol.* **298**, 188–200 (2006). doi:10.1016/j.ydbio.2006.06.028 Medline

36. D. N. Charo, M. Ho, G. Fajardo, M. Kawana, R. K. Kundu, A. Y. Sheikh, T. P. Finsterbach, N. J. Leeper, K. V. Ernst, M. M. Chen, Y. D. Ho, H. J. Chun, D. Bernstein, E. A. Ashley, T. Quertermous, Endogenous regulation of cardiovascular function by apelin-APJ. *Am. J. Physiol. Heart Circ. Physiol.* **297**, H1904–H1913 (2009). doi:10.1152/ajpheart.00686.2009 Medline

37. K. Tatemoto, M. Hosoya, Y. Habata, R. Fujii, T. Kakegawa, M. X. Zou, Y. Kawamata, S. Fukusumi, S. Hinuma, C. Kitada, T. Kurokawa, H. Onda, M. Fujino, Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem. Biophys. Res. Commun.* **251**, 471–476 (1998). doi:10.1006/bbrc.1998.9489 Medline

38. D. K. Lee, R. Cheng, T. Nguyen, T. Fan, A. P. Kariyawasam, Y. Liu, D. H. Osmond, S. R. George, B. F. O'Dowd, Characterization of apelin, the ligand for the APJ receptor. *J. Neurochem.* **74**, 34–41 (2000). doi:10.1046/j.1471-4159.2000.0740034.x Medline

39. B. Tucker, C. Hepperle, D. Kortschak, B. Rainbird, S. Wells, A. C. Oates, M. Lardelli, Zebrafish Angiotensin II Receptor-like 1a (agtrl1a) is expressed in migrating hypoblast, vasculature, and in multiple embryonic epithelia. *Gene Expr. Patterns* **7**, 258–265 (2007). doi:10.1016/j.modgep.2006.09.006 Medline

40. S. Nornes, B. Tucker, M. Lardelli, Zebrafish aplnra functions in epiboly. *BMC Res. Notes* **2**, 231 (2009). doi:10.1186/1756-0500-2-231 Medline

41. M. C. Scimia, C. Hurtado, S. Ray, S. Metzler, K. Wei, J. Wang, C. E. Woods, N. H. Purcell, D. Catalucci, T. Akasaka, O. F. Bueno, G. P. Vlasuk, P. Kaliman, R. Bodmer, L. H. Smith, E. Ashley, M. Mercola, J. H. Brown, P. Ruiz-Lozano, APJ acts as a dual receptor in cardiac hypertrophy. *Nature* **488**, 394–398 (2012). doi:10.1038/nature11263 Medline

42. J. Ishida, T. Hashimoto, Y. Hashimoto, S. Nishiwaki, T. Iguchi, S. Harada, T. Sugaya, H. Matsuzaki, R. Yamamoto, N. Shiota, H. Okunishi, M. Kihara, S. Umemura, F. Sugiyama, K. Yagami, Y. Kasuya, N. Mochizuki, A. Fukamizu, Regulatory roles for APJ, a seven-transmembrane receptor related to angiotensin-type 1 receptor in blood pressure in vivo. *J. Biol. Chem.* **279**, 26274–26279 (2004). doi:10.1074/jbc.M404149200 Medline

43. E. M. Roberts, M. J. Newson, G. R. Pope, R. Landgraf, S. J. Lolait, A. M. O'Carroll, Abnormal fluid homeostasis in apelin receptor knockout mice. *J. Endocrinol.* **202**, 453–462 (2009). doi:10.1677/JOE-09-0134 Medline

44. H. Kidoya, M. Ueno, Y. Yamada, N. Mochizuki, M. Nakata, T. Yano, R. Fujii, N. Takakura, Spatial and temporal role of the apelin/APJ system in the caliber size regulation of blood vessels during angiogenesis. *EMBO J.* **27**, 522–534 (2008). doi:10.1038/sj.emboj.7601982 Medline

45. K. Kuba, L. Zhang, Y. Imai, S. Arab, M. Chen, Y. Maekawa, M. Leschnik, A. Leibbrandt, M. Markovic, J. Schwaighofer, N. Beetz, R. Musialek, G. G. Neely, V. Komnenovic, U. Kolm, B. Metzler, R. Ricci, H. Hara, A. Meixner, M. Nghiem, X. Chen, F. Dawood, K. M. Wong, R. Sarao, E. Cukerman, A. Kimura, L. Hein, J. Thalhammer, P. P. Liu, J. M. Penninger, Impaired heart contractility in Apelin gene-deficient mice associated with aging and pressure overload. *Circ. Res.* **101**, e32–e42 (2007). doi:10.1161/CIRCRESAHA.107.158659 Medline

46. A. Y. Sheikh, H. J. Chun, A. J. Glassford, R. K. Kundu, I. Kutschka, D. Ardigo, S. L. Hendry, R. A. Wagner, M. M. Chen, Z. A. Ali, P. Yue, D. T. Huynh, A. J. Connolly, M. P. Pelletier, P. S. Tsao, R. C. Robbins, T. Quertermous, In vivo genetic profiling and cellular localization of apelin reveals a hypoxia-sensitive, endothelial-centered pathway activated in ischemic heart failure. *Am. J. Physiol. Heart Circ. Physiol.* **294**, H88–H98 (2008). doi:10.1152/ajpheart.00935.2007 Medline

47. N. A. Evans, D. A. Groarke, J. Warrack, C. J. Greenwood, K. Dodgson, G. Milligan, S. Wilson, Visualizing differences in ligand-induced β-arrestin-GFP interactions and trafficking between three recently characterized G protein-coupled receptors. *J. Neurochem.* **77**, 476–485 (2001). doi:10.1046/j.1471-4159.2001.00269.x Medline

48. D. K. Lee, S. S. G. Ferguson, S. R. George, B. F. O'Dowd, The fate of the internalized apelin receptor is determined by different isoforms of apelin mediating differential interaction with β-arrestin. *Biochem. Biophys. Res. Commun.* **395**, 185–189 (2010). doi:10.1016/j.bbrc.2010.03.151 Medline

49. A. Reaux, N. De Mota, I. Skultetyova, Z. Lenkei, S. El Messari, K. Gallatz, P. Corvol, M. Palkovits, C. Llorens-Cortès, Physiological role of a novel neuropeptide, apelin, and its receptor in the rat brain. *J. Neurochem.* **77**, 1085–1096 (2001). doi:10.1046/j.1471-4159.2001.00320.x Medline

50. N. Zhou, X. Fan, M. Mukhtar, J. Fang, C. A. Patel, G. C. DuBois, R. J. Pomerantz, Cell-cell fusion and internalization of the CNS-based, HIV-1 co-receptor, APJ. *Virology* **307**, 22–36 (2003). doi:10.1016/S0042-6822(02)00021-1 Medline

51. S. C. Chng, L. Ho, J. Tian, B. Reversade, ELABELA: A hormone essential for heart development signals via the apelin receptor. *Dev. Cell* **27**, 672–680 (2013). doi:10.1016/j.devcel.2013.11.002 Medline

52. A. F. Schier, W. S. Talbot, Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39**, 561–613 (2005). doi:10.1146/annurev.genet.37.110801.143752 Medline

53. G. Barnes, A. G. Japp, D. E. Newby, Translational promise of the apelin—APJ system. *Heart* **96**, 1011–1016 (2010). doi:10.1136/hrt.2009.191122 Medline

54. M. J. Kleinz, A. P. Davenport, Emerging roles of apelin in biology and medicine. *Pharmacol. Ther.* **107**, 198–211 (2005). doi:10.1016/j.pharmthera.2005.04.001 Medline

55. A. Kasai, N. Shintani, H. Kato, S. Matsuda, F. Gomi, R. Haba, H. Hashimoto, M. Kakuda, Y. Tano, A. Baba, Retardation of retinal vascular development in apelin-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1717–1722 (2008). doi:10.1161/ATVBAHA.108.163402 Medline

56. H. Kidoya, H. Naito, N. Takakura, Apelin induces enlarged and nonleaky blood vessels for functional recovery from ischemia. *Blood* **115**, 3166–3174 (2010). doi:10.1182/blood-2009-07-232306 Medline

57. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011). doi:10.1038/nmeth.1701 Medline

58. L. Käll, A. Krogh, E. L. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004). doi:10.1016/j.jmb.2004.03.016 Medline

59. L. Käll, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. *Nucleic Acids Res.* **35**, (Web Server), W429–W432 (2007). doi:10.1093/nar/gkm256 Medline

60. M. F. Lin, I. Jungreis, M. Kellis, PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011). doi:10.1093/bioinformatics/btr209 Medline

61. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995). doi:10.1002/aja.1002030302 Medline

62. B. Boldajipour, H. Mahabaleshwar, E. Kardash, M. Reichman-Fried, H. Blaser, S. Minina, D. Wilson, Q. Xu, E. Raz, Control of chemokine-guided cell migration by ligand sequestration. *Cell* **132**, 463–473 (2008). doi:10.1016/j.cell.2007.12.034 Medline

63. M. D. Abramoff, P. J. Magalhães, S. J. Ram, Image processing with ImageJ. *Biophoton. Int.* **11**, 36–42 (2004).

64. C. Thisse, B. Thisse, High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008). doi:10.1038/nprot.2007.514 Medline

65. A. Kaufmann, M. Mickoleit, M. Weber, J. Huisken, Multilayer mounting enables long-term imaging of zebrafish development in a light sheet microscope. *Development* **139**, 3242–3247 (2012). doi:10.1242/dev.082586 Medline

66. P. J. Keller, A. D. Schmidt, J. Wittbrodt, E. H. K. Stelzer, Digital scanned laser light-sheet fluorescence microscopy (DSLM) of zebrafish and Drosophila embryonic development. *Cold Spring Harb. Protoc.* **2011**, 1235–1243 (2011). doi:10.1101/pdb.prot065839 Medline

67. K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, G. Danuser, Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008). doi:10.1038/nmeth.1237 Medline