

Supporting Information

Somorjai et al. 10.1073/pnas.1100045109

SI Materials and Methods

Quantification of Tail Regeneration Profiles. The distance y between the myosepta of one myomere at a distance two to three myomeres from the amputation plane was used as an internal reference to control for appropriate scale and size of imaged tails (Fig. S1H). Although the mean of y did not differ across timepoints ($P > 0.5$), this reference ensured consistency of data. Specifically, at each timepoint n , the size of the regenerating tail (x) was divided by this internal reference, such that $t_n = x/y$. Normalized growth (increase in regenerate length) was then determined by subtracting tail size at amputation (t_0) from tail size at each subsequent timepoint t_n from 1 (8 d) to 15 wk (105 d). It was not possible to directly measure blastema size, because remodeling at the interface between original musculature and blastema post-amputation confounded accurate estimates of the amputation plane by day 14. The formula $t_n - t_{n-1}$ was also used to calculate periodic growth intervals; these data are an estimate of weekly growth rates that are directly comparable from 2 to 10 wk (i.e., identical intervals of 7 d).

Scoring of Tail Regenerate Quality. Scoring of regenerates was graded on a scale of zero to five based on severity of the phenotype after 8 wk. A final quality score of zero would indicate no regeneration or an inability to reconstitute the original structure, and a score of five would correspond to perfect regeneration. No individuals were considered to have regenerated perfectly, because tails did not reach their preamputation size (0; maximal size = 1) by the time of evaluation. The ability to make all tissue types (fin, notochord, muscle, and neural tube) was also scored as one; only a single individual failed to obviously differentiate anything other than fin and was scored zero. Several additional criteria of quality were evaluated independently and were assigned 0.5 points, including quality of differentiated muscle (no obvious differentiated muscle = 0, muscle fibers present throughout the regenerate = 1), axis of the regenerate (angle deviating from the horizontal, dorsoventral defects = 0; horizontal, no dorsoventral defects = 1), and notochord defects (notches or hypomorphic notochord regenerate width = 0, no defects = 1). Points were added for each individual to calculate its quality score, and mean scores were calculated for each age/size class (Fig. S1).

Quantification of Proliferation and Pax3/7 Maximal Distribution. Quantification and confocal image processing were performed using ImageJ. Maximal extent of proliferation [phosphohistone H3 (PH3)] and Pax3/7 expression were measured from the tip of the blastema to the most anterior point of antibody staining. A total of 15 day-15 ($n = 7$ small, $n = 8$ large), 14 day-17 ($n = 8$ small, $n = 6$ large), and 22 day-20 ($n = 10$ small, $n = 12$ large) regenerate sections were quantified (two to three animals per category). Because of small sample sizes and unknown data distribution, nonparametric statistics were used to test for significance. Graphs show means \pm SE. For quantification of proliferation, the total number of PH3⁺ cells in all sections were counted; proportions of PH3⁺ cells were determined for different tissues (neural tube, notochord, muscle/blastema, and epidermis) and zones (blastema, dedifferentiation zone, and outside the regenerate) and compared using the nonparametric adjusted χ^2 test. In all tests, $P \leq 0.05$ was considered statistically significant, but only those tests passing the Bonferroni correction for multiple comparisons are shown (Fig. S7). All statistics were performed using Statistica 2.0 software (Statsoft).

In Situ Hybridizations. Whole-animal or amputated tails were fixed in 4% Paraformaldehyde (PFA) in Mops overnight at 4 °C or 2 h at room temperature and stored at -20 °C in 70% ethanol. After a graded ethanol series, tails were embedded in paraffin, sectioned at 10–15 μ m, and mounted on DAKO slides. In situ hybridizations were performed as previously described using antisense probes against *msx*, *chordin*, *pax3/7* (1), and *soxB2*. *msx* was amplified with forward primers CAATCCACGCTACCGACGAGCA and AC-CGAGAAGCTGCCCTTCAGTG and reverse primers TGAC-GAGTGGATGAACCCAGCGTA and TCTGGCAAGCTGTACGGGCAGTC. PCR products were cloned into pGEM-t Easy (Promega) vectors and subcloned into pBluescript SK+ (Invitrogen). After sequencing verification, *msx*-, *pax3/7*-, and *chordin*-containing vectors were linearized using XhoI, and the cDNA was in vitro transcribed using T3 polymerase. The *soxB2* probe was provided by Manuel Irimia (University of Barcelona, Spain).

Immunohistochemistry and Histology. Immunohistochemistry was performed on both paraffin microtome (10–15 μ m) and vibratome (50–100 μ m) sections and whole-mount tails using standard procedures. Briefly, samples were rinsed in Tris Buffer (TBS), washed extensively with 0.1% Tween (TBST), permeabilized with the addition of 0.5% Triton-X (TBSTx), and finally blocked with 5% heat-inactivated sheep serum and 2% BSA. Immunohistochemistry on paraffin sections using DP312 required a 15-min epitope retrieval step at 95 °C in Tris, pH 8.0. Primary antibodies were added overnight at 4 °C, including α Pax3/7 clone DP312 at 1:30 (provided by N. Patel, University of California, Berkeley, CA) (2), α -acetylated tubulin at 1:500 (Sigma), and α -PH3 at 1:500 (Millipore). The DP312 antibody specifically reacts with Pax3/7 on Western blots in bacterially expressed *Drosophila* protein, also confirmed by immunohistochemistry in *Drosophila*, *Schistocerca*, the chordates *Ciona* and zebrafish, and cnidarians (2). After several 1-h-long washes in TBST and an additional blocking step, samples were incubated overnight at 4 °C using secondaries Alexa-Fluor 488 and Alexa-Fluor 568 at 1:400 (Invitrogen). For F-actin staining, samples were fixed in 4% PFA in Mops and directly placed in PBS, bypassing ethanol storage. Phalloidin-568 (Invitrogen) was added during the secondary incubation step. Nuclei were stained using DAPI (1:5,000 of 5 mg/mL stock) added during the last wash step before mounting in Prolong Gold antifade reagent (Invitrogen). Images were acquired with a Leica SPII confocal microscope. Mallory triple staining was performed on microtome sections after a standard deparaffinization step.

Classifier Development. The discrete regeneration data were fitted to a Gompertz function (see below). Linear discriminant analysis was used to develop the classifier. The classifier was cross-validated with the leave one out approach, because sample sizes were too small for creating an independent test set that could be classified reliably. The seven normalized regenerative growth timepoints were projected onto a class proximity plane. Basically, the class proximity plane is a 2D display of a particular class proximity measure that represents an animal's relative degree of belongingness to the two classes. Thus, if an animal is more similar (less dissimilar) to animals in the small-animal class (class 1) than to those animals in the large-animal class (class 2), then we assign the animal to class 1. Otherwise, it is allocated to class 2. The degree of dissimilarity is determined by an a priori selected dissimilarity/distance measure. Because of its natural occurrence in

the linear discriminant analysis classification, we use the Mahalanobis distance. Similarly, we can select a class proximity measure. Of the many possibilities, we chose the class centroids for the class proximity measure. The rationale and methodology behind the classifier development used here to classify regeneration growth curves in the two groups are as outlined below.

The biological expectation is that the tail regeneration is slow initially, then accelerates, and finally, slows down until the new tail reaches its final length. We may model this process by fitting the data to some S-shaped curve. An obvious candidate is a Gompertz growth curve (3), a sigmoid function of time.

Gompertz curves (3–6) have been used to model several different types of time series. Of the many biological applications, the increase of animal weights over time, growth of tumor sizes (7), and development of organisms are of particular relevance. Stages of our classifier development are further detailed in reference (8).

We may describe the curve by Eq. S1,

$$y(t) = C^* \exp(-b^* \exp(kt)), \quad [\text{S1}]$$

where $y(t)$ is the time-dependent measurable quantity (here, the new tail length) representing the growth behavior to be modeled. C is the (assumed known) upper limit of $y(t)$ [$y(t) \rightarrow C$ as time $t \rightarrow \infty$], and b and k are the coefficients obtained by fitting Gompertz curves to the time courses.

The inflection point T_{INFL} , at which the growth reaches its maximum, is at time $T_{\text{INFL}} = -\ln(b)/k$ when $y(T_{\text{INFL}}) = C/e$. Here, e is Euler's number ($e = 2.71828\dots$), the base of the natural logarithm. Taking natural logarithms two times, we get a linear equation in time (Eq. S2):

$$\ln(\ln(C/y(t))) = \ln b + kt. \quad [\text{S2}]$$

The experimental time courses to be fitted are displayed in Fig. S8. In the plots, the value $C = 3.225$ is the common median length of the 64 animal tails before cutting (i.e., at $t = 0$). This finding is a compromise, an extrapolated estimate of what, on average, the tail lengths would be for all animals if/when they are fully regrown. We could have used the different C value limits of the different classes (viz. $C_{\text{large}} = 2.695$ and $C_{\text{small}} = 3.455$). However, the empirical relationship $T_{\text{INFL}}(C_{\text{small}})/T_{\text{INFL}}(C_{\text{large}}) \sim C_{\text{small}}/C_{\text{large}}$ holds quite well; thus, all T_{INFL} values calculated with $C = 3.225$ fall between $T_{\text{INFL}}(L_{\text{small}})$ and $T_{\text{INFL}}(L_{\text{large}})$.

In Fig. S9, we display the best linear regression fits to Eq. S2 (Fig. S9A) and its residuals (Fig. S9B) and the best quadratic regression fits to Eq. S4 (Fig. S9C) and its residuals (Fig. S9D) for median small animals (MedS). The goodness of either type of

fit is also comparable for SQ1 and SQ3. We obtained statistically equivalent results for the median large animals (MedL), LQ1, and LQ3 curves.

Based on the essentially parabolic (not random) time dependence of the residuals of the linear regression fit (Fig. S9 shows MedS), we require some generalization of the Gompertz curve (Eq. S1). The linear term, kt , in Eq. S1 may be replaced by a general function of time t , $F(t)$. A quadratic version of $F(t)$, given by Eq. S3, is the simplest possible such generalization (6):

$$y(t) = C^* \exp(-b^* \exp(k_1 t + k_2 t^2)). \quad [\text{S3}]$$

The unknown parameters b , k_1 , and k_2 in Eq. S3 may still be determined by a linear least squares regression fit of Eq. S4:

$$\ln(\ln(C/y(t))) = \ln b + k_1 t + k_2 t^2. \quad [\text{S4}]$$

Again, the inflection point T_{INFL} occurs when $y(t) = C/e$ and may be obtained by solving the quadratic equation for t (Eq. S5):

$$\ln b + k_1 t + k_2 t^2 = 0. \quad [\text{S5}]$$

In Table S1, we display the accuracy for both the linear and quadratic regression fits and the computed values for the inflection points $T_{\text{INFL}}^{\text{L}}$ and $T_{\text{INFL}}^{\text{Q}}$.

R^2 is the coefficient of determination (i.e., the ratio of explained and total variation), with a range $0 \leq R^2 \leq 1$; $R^2 = 1$ corresponds to a perfect fit. R_{STDEV} is the SD of the residuals: the closer to zero, the better the fit.

In terms of both the R^2 and R_{STDEV} values, the quadratic regression fits numerically outperform the linear ones. However, the results for the linear inflection points $T_{\text{INFL}}^{\text{L}}$, determined from Eq. S2, are statistically indistinguishable ($P < 0.0001$) from their quadratic equivalents, $T_{\text{INFL}}^{\text{Q}}$, both for the small and large animals.

In Table S2, we give the P values for both the intercept and the slope for the linear regression fits. A small value for P (e.g., $P < 0.05$ or < 0.01) means that, with 95% or 99% confidence, neither intercept nor slope is zero. (For the quadratic regression fits, equivalent P values are not readily calculable.)

In Table S3, we collect the F ratios and corresponding P values for the linear and quadratic fits. The F ratio is a measure of how much of the variance of the data is explained by fitting to the given equation vs. how much of the variance is explained by chance. The larger the F ratio, the better the fit, which is also expressed by a very small P value. Although the results in Table S3 clearly show the superiority of the quadratic fits, the P values for the linear case are statistically more than satisfactory.

- Somorjai I, Bertrand S, Camasses A, Haguenaer A, Escriva H (2008) Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev Genes Evol* 218:703–713.
- Lagha M, et al. (2008) Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program. *Genes Dev* 22:1828–1837.
- Gompertz B (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos Trans R Soc Lond* 123:513–583.

- Tsoularis A (2001) Analysis of logistic growth models. *Res. Lett. Inf. Math. Sci.* 2:23–46.
- Jukić D, Kralik G, Scitovski R (2004) Least-squares fitting Gompertz curve. *J Comput Appl Math* 169:359–375.
- Winsor CP (1932) The Gompertz curve as a growth curve. *Proc Natl Acad Sci USA* 18: 1–8.
- Wheldon TE (1988) *Mathematical Models in Cancer Research*, Adam Hilger, Bristol.
- Somorjai RL, Dolenko B, Nikulin A, Roberson W, Thiessen N (2011) Class proximity measures—dissimilarity-based classification and display of high-dimensional data. *J Biomed Inform* 44:775–788.

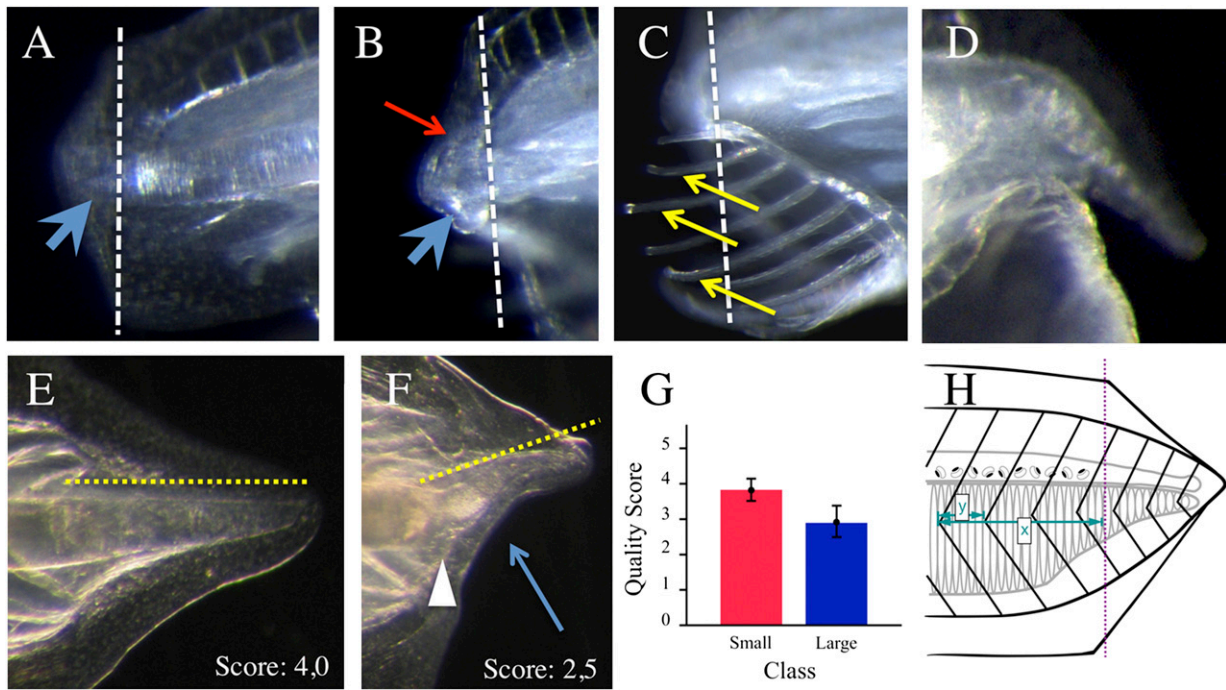


Fig. S1. (A–C) Details of anterior regeneration, 4 wk postamputation, with amputation plane (white dashed line) through the notochord (A) and anterior to the velum (B and C). Blue arrowheads show notochord regenerates, red arrow shows the neural tube, and yellow arrows show buccal cilia. (D) A 10-wk regenerate after amputation through the hepatic diverticulum. Note the small tail with notochord. (E and F) Details of scoring of posterior regenerates with amputation through the postanal tail. (E) Near perfect regenerate (score = 4.0). (F) Defects in regeneration (score = 2.5) include axis and structural defects. The yellow dotted line indicates the orientation of the axis. The white arrowhead shows the joint between the original notochord and the regenerated notochord, which is smaller in diameter. The blue arrow points to a dent, where ventrally, regeneration did not proceed correctly. (G) Quantification of regeneration scores in large and small animals indicates that tails regenerate less well in larger animals (small score mean = 3.83 ± 0.31 , large score mean = 2.95 ± 0.47 , $P \leq 0.001$). (H) Schematic diagram of quantification of tail regeneration scheme, where the purple dotted line marks the amputation plane. x , Size of regenerate at t_i ; y , internal reference measure. Here, t_0 is shown, because x is measured at day 0 when amputation is performed.

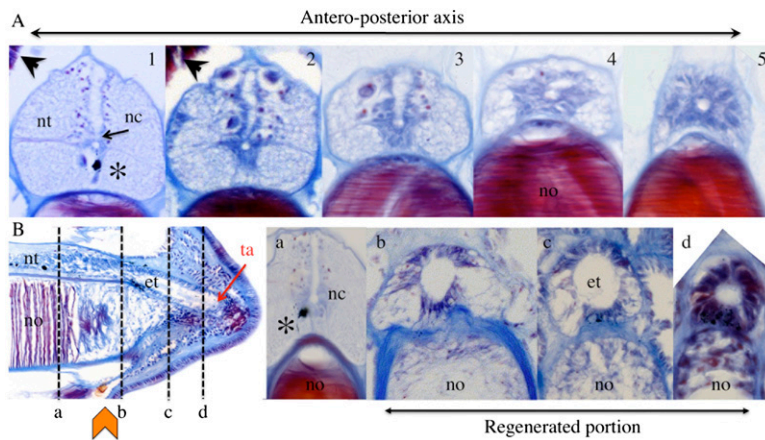


Fig. S2. (A) Representative serial transverse sections through the neural tube of an uninjured adult amphioxus tail, illustrating the changes in morphology proceeding from anterior (proximal) to posterior (distal). Mallory triple staining shows connective and neural tissue in blue and muscle in red or purple. Arrowheads in 1 and 2 highlight the last leftside myotome tail, which is absent in the distal sections 3–5. Proximal sections contain rhabdomeric photoreceptors (Hesse ocelli) associated with a black pigment spot (*). The neural tube is always closely associated with the notochord below and dorsally by a medial septum (1 and 2) that is absent in more posterior sections. The distal section (5) shows how the neural tube narrows to become a simple tube reminiscent of an ependymal tube, extending directly from the neural canal (black arrow in 1, nc). (B, Left) A representative longitudinal section through the midline of a 17-d (stage 3) regenerate, indicating the location of transverse sections shown in B, Right (a–d). By this stage, the morphology of the ependymal tube is very similar to the morphology of the original neural tube, whereas the notochord remains undifferentiated. et, Ependymal tube; nc, neural canal; no, notochord; nt, neural tube; ta (red arrow), terminal ampulla. The yellow block arrowhead marks the amputation plane.

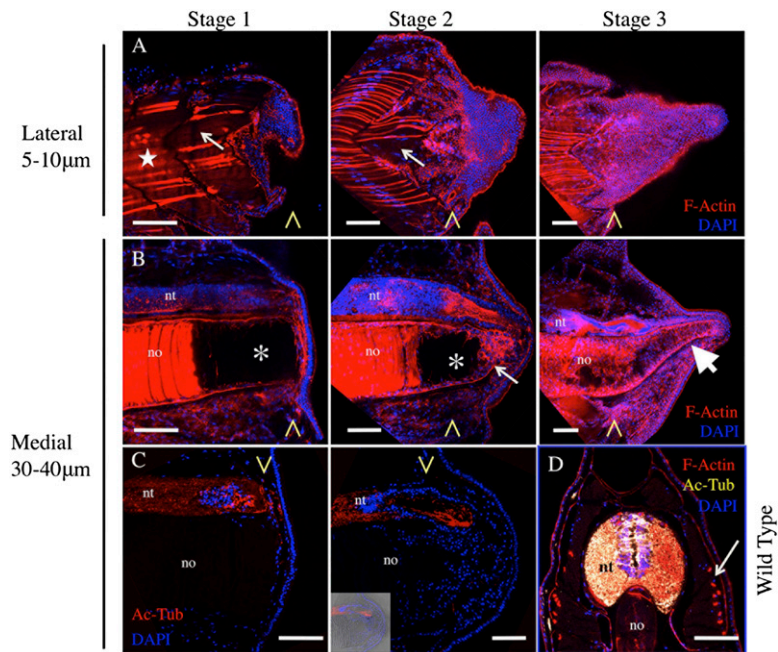


Fig. S3. Dynamic changes in F-actin and acetylated tubulin accompany muscle dedifferentiation and neural tube elongation during tail regeneration. (A) Lateral sections through the degenerating muscle fibers (white arrow) of stages 1–3 regenerates, which was evidenced by Phalloidin labeling of F-actin. The process of muscle degradation is complex and not always uniform in time or space. The white star indicates a portion of the image that does not directly focus on the striated lateral muscle fibers but rather, is immediately below/medial to them. (B) Sagittal sections through the midline of the same individuals shown in A. By stage 1, the F-actin–positive lamellar structure of the notochord is lost at a considerable distance from the amputation plane (white asterisks). At stage 2, F-actin begins to accumulate around the nuclei at the tip of the regenerating notochord (white arrow). This reorganization of the distal notochord continues until, by stage 3, virtually the entire notochord structure has been reformed (thick white arrow). (C) Acetylated tubulin labels the regenerating axons of the neural tube, first extending beneath the wound epithelium (stage 1) and then clearly labeling the ependymal bulb at stage 2. C, *Inset* shows the stage 2 regenerate confocal image overlain by the Nomarski image. (D) F-actin and acetylated tubulin colocalize in the unamputated neural tube and neurons (white), but only F-actin is present in the lateral fibers of the axial muscle (white arrow) and the notochord. A–C represent sagittal confocal sections of the regenerate; the exception is D, which is an anterior transverse plane of an uninjured amphioxus. Note the relatively larger size of the neural tube relative to the notochord. Yellow arrowheads show the plane of amputation. no, notochord; nt, neural tube. (Scale bar: 50 μ m.)

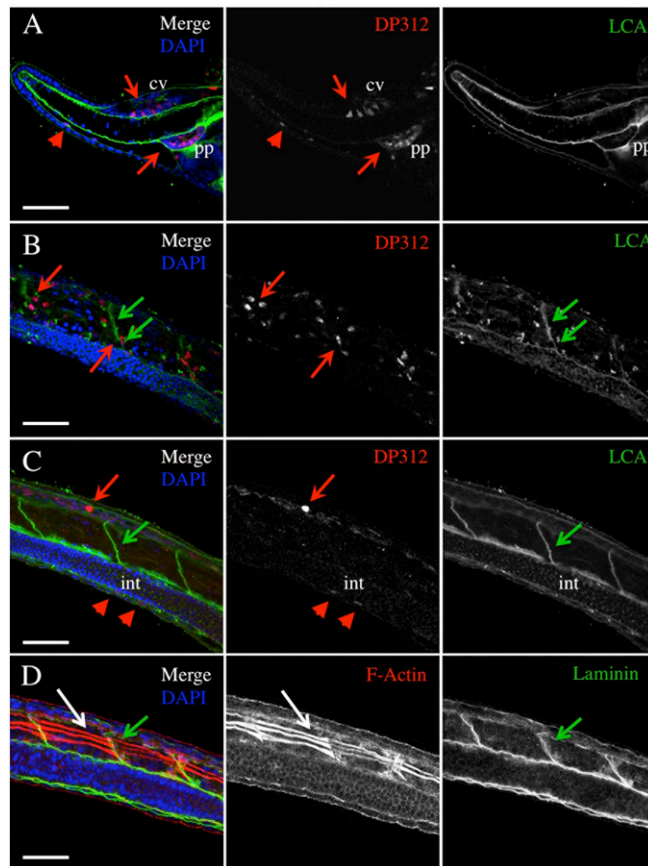


Fig. S4. DP312 expression in larvae. (A) Confocal image of an amphioxus larva (76 h) head, showing Pax3/7 protein in the cerebral vesicle (cv), preoral pit (pp, red arrows), and isolated cells of the anterior coelom (red arrowhead). (B and C) DP312 labels superficial lateral cells immediately below the epidermis and overlying the axial musculature (B, red arrows) as well as large neurons of the neural tube (C, red arrows) and isolated cells of the nascent subintestinal vein (red arrowheads) sandwiched between the intestine (int) and the epidermis. Fluorescent lectin (LCA) labels the polysaccharides of the basal lamina (A–C, green arrowheads) among other structures. Note the location of these DP312+ cells associated with the basal lamina of the myosepta. (D) The myosepta linking the lateral musculature (white arrows, F-actin) are not only labeled with conjugated LCA but also laminin (green arrow), confirming that they contain basal lamina (compare B with D, which were taken at the same confocal depth). Anterior is always to the left and dorsal is up. (Scale bar: 5 μ m.)

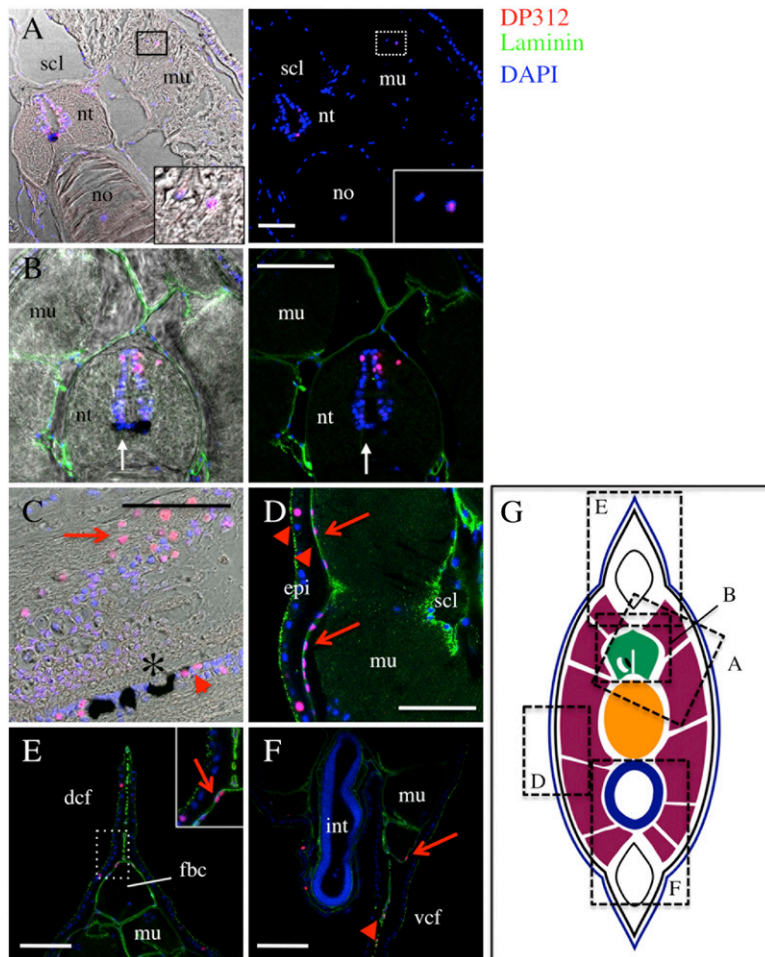


Fig. S5. DP312 expression in adult amphioxus. (A) DP312 is expressed in very few isolated cells of the musculature (mu; *Inset*) and the central canal of the neural tube. Note the large cavities, termed sclerocoels. (B) Detail of the neural tube. Some but not all nuclei are labeled. (C) Longitudinal section through the neural tube (position marked by the white arrow in *B*). Large neurons of the central canal express DP312 (red arrows) as do neurons of the floor plate (red arrowheads), where one can clearly identify the pigment cups of the rhabdomic photoreceptors or Hesse ocelli (*). (D) DP312 is also expressed in a few isolated cells throughout the epidermis (red arrowhead) as well as infrequently in a layer of flattened cells within the basal lamina overlying the musculature (red arrows). (E and F) Some cells within the dorsal (E, *Inset*) and ventral fin chambers (F) also express DP312 (red arrows) as do cells of the fin mesenchyme (D, arrowheads). (G) Position of transverse sections A and B, and D–F are outlined by black dotted boxes. C is not visible, because it is a longitudinal section through the neural tube. Muscle is represented in purple, epidermis and epithelium of the intestine are represented in blue, neural tube is represented in green, and the notochord is represented in yellow. dcf, dorsal caudal fin; epi, epidermis; int, intestine; no, notochord; nt, neural tube; scl, sclerocoel; vcf, ventral caudal fin. (Scale bar: 50 μm .)

Table S3. *F* ratios and corresponding *P* values for the linear and quadratic fits

	Linear		Quadratic	
	<i>F</i> ratio	<i>P</i>	<i>F</i> ratio	<i>P</i>
SQ3	41.0	0.001	836.0	<0.001
MedS	58.1	0.001	613.0	<0.001
SQ1	48.5	0.001	571.0	<0.001
LQ3	125.0	<0.001	383.0	<0.001
MedL	104.0	<0.001	468.0	<0.001
LQ1	150.0	<0.001	824.0	<0.001

See text following Eq. S5.