#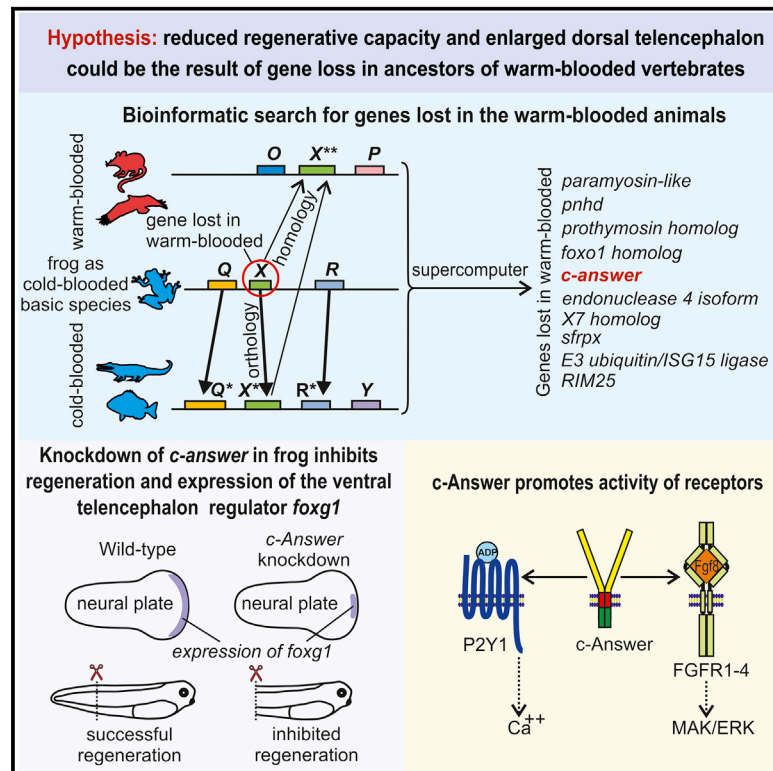 Bioinformatics Screening of Genes Specific for Well-Regenerating Vertebrates Reveals *c-answer*, a Regulator of Brain Development and Regeneration

Graphical Abstract

Authors

Daria D. Korotkova, Vassily A. Lyubetsky, Anastasia S. Ivanova, ..., Maria B. Tereshina, Leonid Peshkin, Andrey G. Zaraisky

Correspondence

azaraisky@yahoo.com

In Brief

Poor regeneration in warm-blooded animals could be caused by gene loss in evolution. Here, we describe a bioinformatic approach to identify lost genes. One of these, *c-answer*, regulates regeneration and brain development in cold-blooded animals. Thus, loss of *c-answer* could promote evolutionary changes related to regeneration and brain development in warm-blooded animals.

Highlights

- New bioinformatics approach detects 8 genes lost by warm-blooded animals in evolution

- One of the lost genes encodes membrane protein c-Answer

- c-Answer regulates regeneration and brain development in *Xenopus laevis*

- c-Answer promotes MAP/ERK pathway and Ca^{2+}-dependent purinergic pathway

CellPress

Article

Bioinformatics Screening of Genes Specific for Well-Regenerating Vertebrates Reveals *c-answer*, a Regulator of Brain Development and Regeneration

Daria D. Korotkova,[1] Vassily A. Lyubetsky,[2] Anastasia S. Ivanova,[1] Lev I. Rubanov,[2] Alexander V. Seliverstov,[2] Oleg A. Zverkov,[2] Natalia Yu. Martynova,[1] Alexey M. Nesterenko,[1,3] Maria B. Tereshina,[1] Leonid Peshkin,[4] and Andrey G. Zaraisky[1,5,*]

[1]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences (IBCH RAS), 16/10 Miklukho-Maklaya str., Moscow 117997, Russia
[2]The Institute for Information Transmission Problems, Russian Academy of Sciences (IITP RAS), 19 Bolshoy Karetny str., Moscow 127051, Russia
[3]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 1/40 Leninskie Gory, Moscow 119991, Russia
[4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[5]Lead Contact
*Correspondence: azaraisky@yahoo.com
https://doi.org/10.1016/j.celrep.2019.09.038

SUMMARY

The molecular basis of higher regenerative capacity of cold-blooded animals comparing to warm-blooded ones is poorly understood. Although this difference in regenerative capacities is commonly thought to be a result of restructuring of the same regulatory gene network, we hypothesized that it may be due to loss of some genes essential for regeneration. We describe here a bioinformatic method that allowed us to identify such genes. For investigation in depth we selected one of them encoding transmembrane protein, named "c-Answer." Using the *Xenopus laevis* frog as a model cold-blooded animal, we established that c-Answer regulates regeneration of body appendages and telencephalic development through binding to fibroblast growth factor receptors (FGFRs) and P2ry1 receptors and promoting MAPK/ERK and purinergic signaling. This suggests that elimination of *c-answer* in warm-blooded animals could lead to decreased activity of at least two signaling pathways, which in turn might contribute to changes in mechanisms regulating regeneration and telencephalic development.

INTRODUCTION

Restructuring of *cis*-regulatory elements of the gene network, which consists of the same set of genes, is thought to underlay most evolutionary events, in particular, the reduction of the appendage regenerative capacity in birds and mammals (warm-blooded animals) compared to those of well-regenerating fishes, amphibians, and reptiles (cold-blooded animals) (Rodríguez-Trelles et al., 2003; Wray, 2007). However, recently we demonstrated that genes encoding the thioredoxin-domain-containing secreted protein Ag1 and the small GTPases Ras-

dva1 and Ras-dva2, which are essential for the regeneration of body appendages in fishes and frogs, were eliminated during the evolution of birds and mammals (Ivanova et al., 2013, 2018; Tereshina et al., 2014). Based on this finding, we hypothesized that the low regenerative capacity of warm-blooded animals could be at least partially explained by extinction of *ag1* and *ras-dva1/2* during evolution. In turn, assuming these data are correct, one may suppose that other genes lost during the evolution of birds and mammals may still be involved in the regulation of regeneration in well-regenerating cold-blooded vertebrates. In the present work, we proposed a bioinformatics method (algorithm and computer program) to perform a systematic search for such genes combined with experimental testing of the involvement of the predicted genes in *Xenopus laevis* tadpole tail and hindlimb bud regeneration.

Using the developed approach, we identified several genes missing in warm-blooded animals and selected those demonstrating increased expression during regeneration of the amputated tadpole tail and hindlimb bud. Based on the protein sequence analysis, we selected one gene that encoded a previously unknown putative membrane protein and studied its functions in depth. We demonstrated that this gene was expressed predominantly in the presumptive neural plate beginning from the late gastrula stage and was sharply activated in cells of the wound epithelium at the first day after amputation of the tadpole tail and hindlimb buds. Downregulation of this gene by anti-sense morpholino and CRISPR/Cas9 diminished the overall tadpole size, specifically the eye size, and retarded tadpole tail regeneration. Conversely, overexpression of the identified gene elicited the reverse effects (i.e., an increase in the telencephalon and eyes, including ectopic eye differentiation, and restoration of tail regeneration during the "refractory" period, i.e., at stages 45–47, when the tail normally cannot regenerate). We also showed that the membrane protein encoded by this gene could bind two types of receptors involved in signaling that regulated development of the telencephalon and eyes and regeneration: the fibroblast growth factor receptors (FGFRs) FGFR1–4 and the extracellular ADP receptor

P2ry1. As we further demonstrated, if overexpressed in the *Xenopus laevis* embryos, this protein promotes MAPK/ERK and purinergic signaling activated by these receptors. Accordingly, we named this protein c-Answer after cold-blooded animal-specific wound epithelium receptor-binding protein. Together with our previous data on the *ag1* and *ras-dva* genes, the results obtained confirm that these significant evolutionary changes resulting in loss of the ability to regenerate major body appendages together with progressive evolution of the telencephalon, which are characteristics of warm-blooded animals, may have been caused by loss of a gene set in the ancestral species that regulates regeneration and brain development in cold-blooded species.

RESULTS

Bioinformatics Approach Used to Search for Missing Genes in Warm-Blooded Vertebrates

We developed an approach aimed at identifying genes present in cold-blooded animals that have no orthologs (i.e., direct homologs) in warm-blooded species (Lyubetsky et al., 2017; Zverkov et al., 2015). For the purposes of our study, we regarded orthologs as a pair of homologs in distinct species that maintained local genomic synteny (i.e., had at least one pair of independent homologous genes in their vicinity). The assumption was that if a given cold-blooded animal gene was lost during evolution in warm-blooded animals, then its intra-chromosome neighbors in cold-blooded animals should not have homologs in warm-blooded animals in the vicinity of any homolog of the lost gene. Homology was established using protein sequences.

To establish orthology, we used a two-step strategy. The genome of *Xenopus tropicalis*, which presumably contains a full collection of genes present in cold-blooded animals and is well sequenced, was used as a reference or basic species. At the first step, using a previously developed algorithm (called ClusterZSL) (Lyubetsky et al., 2013; Rubanov et al., 2016; Zverkov et al., 2012, 2015) for each gene of *X. tropicalis*, we formed a cluster of the most homologous genes from the selected representatives of cold- and warm-blooded species with well-sequenced and annotated genomes. Several orthology inference methods, including local synteny consideration, were compared previously using five mammalian genomes (Jun et al., 2009).

At the second step, we chose homologs within each cluster that would maintain local genomic synteny in cold-blooded but not in warm-blooded species. These homologs were considered as genes lost during evolution by warm-blooded animals.

The developed algorithm can operate on any two juxtaposed groups of animals (e.g., anamniotes versus amniotes, apodes versus tetrapodes, short- versus long-living, etc.) that henceforth are referred to as the "lower" and "upper" sets. These groups can be generated based on any trait that is present in lower species but not in upper species. In our study, the lower set includes well-regenerating cold-blooded animals (fish, amphibians, and reptiles), whereas the upper set includes poorly regenerating warm-blooded animals (birds and mammals). In turn, the sets are subdivided into parts. Using the latter term, we designated a group of selected species belonging to some biological class of vertebrates.

The numeric parameters p_i and q_j were assigned to each upper and lower part, respectively. Then, in the basic species (*X. tropicalis*), we tried to identify genes with disrupted synteny in at least p_i upper species of the i-th part and undisrupted synteny in at least q_j lower species of the j-th part; given this approach, some upper or lower species would lack or have the gene, respectively. In an extreme case, p_i equals the number of upper species, and $q_j = 1$ for all i and j. A decreasing number of gene paralogs were used as an additional condition with a numeric parameter r_j. Specifically, if the number of paralogs of a gene in the basic species exceeds the number of its paralogs in an upper species in accordance with the specified parameter $r_j \geq 1$, then the gene is considered lost.

All of the protein-coding genes from the basic species were tested using the 2- and 3-species modes. In the 2-species mode, homolog X^* of gene X in each lower species was checked for synteny. In this mode, the following two conditions were verified to establish that a given gene had no orthologs in the upper species. (1) A pair of genes Y and Z are defined in the basic species as different from X and each other and co-localize within a window of the size $2l$; their homologs X^*, Y^*, and Z^* must co-localize within windows of the size $2l_1$ in the lower species, as shown by the bold arrows (Figure 1A). (2) There is no homolog X^{**} in the upper species or its synteny is disrupted.

The latter means that for each gene S within the window in the basic species, there is no homolog within windows of the size $2l_2$ among the upper species (Figure 1A). The genes Y, Z, and S will be referred to as witnesses. The algorithm parameters specify the desired number of witnesses. Here, 2 Mbp was chosen as the value of $l = l_1 = l_2$ based on our empirical observation that at least one witness could be always found at a distance of 2 Mbp or less for all well-established orthologs. In the 3-species mode, the following two conditions are regarded to state that gene X in the basic species has no ortholog in the upper species: (1) X has no witnesses near any of its homologs in the upper species, and (2) any ortholog of X in the lower species also has no witnesses near any of its own homologous upper species (Figure 1B).

The computer implementation of the described method for the prediction of gene losses and gains between several groups of species is freely available at http://lab6.iitp.ru/en/lossgainrsl/. The program is deeply parallelized and can operate on a super-computer, which is essential if a large number of complete genomes are considered jointly or the synteny blocks consist of many neighboring genes.

Let us denote the set of genes of the basic species obtained in the 2-species mode by the 2-species list and that obtained in the 3-species mode by the 3-species list. The intersection of the 2- and 3-species lists is referred to as the gene list for given definitions of homology (see Method Details). Thus, the intersection of the gene lists for specific definitions is referred to as the list of lost genes and is the output of our method and program. Figures 1C and 1D illustrate the cases in which gene X in the basic species has orthologs in the upper species and thus should be excluded from the final list of lost genes.

For the genomic data specified in the STAR Methods section, our program predicted the following genes as lost in warm-blooded animals: *ENSXETG00000033176* (previously unknown
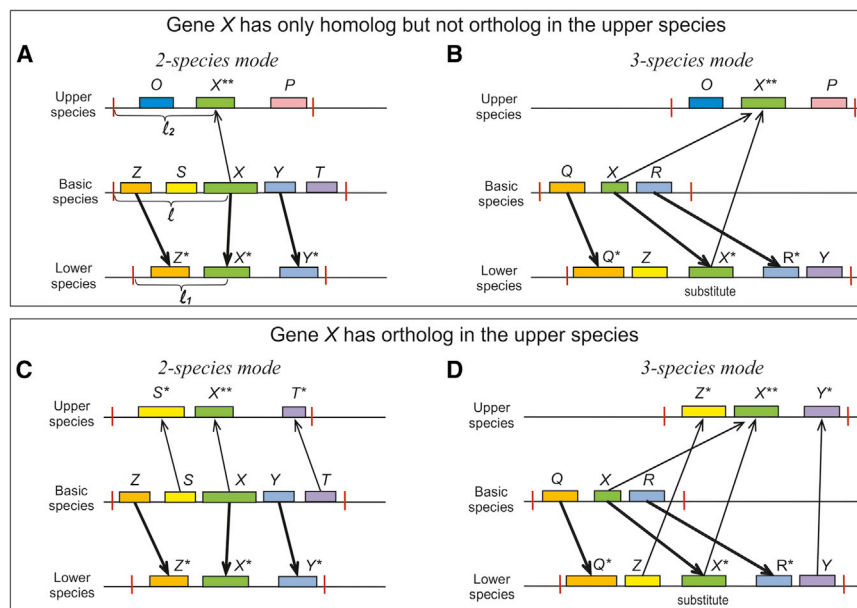
Figure 1. The 2- and 3-Species Selection Modes for Genes with No Orthologs in the Upper Species

(A and B) The schemes demonstrate the case when gene *X* is considered to be the lost gene for the 2- and 3-species modes. Gene *X* only has homolog *X*** but no ortholog in the upper species because the neighbors of *X* in the basic (for 2-species mode, A) or lower (for 3-species mode, B) species (*Z*, *S*, *Y* or *Q**, *R**, *Y*, respectively) have no homologs in the upper species, where the neighbors of gene *X*** are *O* and *P*. The red ticks indicate the borders of the "window" in which local synteny has been checked. The size of the window, *l*, was chosen in this particular work to be 2 Mbp for all genes: $l = l_1 = l_2$. Thin and bold arrows indicate homologous and orthologous genes in the upper and lower species, respectively.

(C and D) The schemes demonstrate the case in which gene *X* is considered to be preserved for the 2- and 3-species modes, respectively. Gene *X* has an ortholog *X*** in the upper species because at least one of the neighbors of *X* in the basic (for 2-species mode, C) or lower (for 3-species mode, D) species (*Z*, *S*, *Y* or *Q**, *R**, *Y*, respectively) has a homolog in the upper species, where the neighbors of gene *X*** are *Y** and *S**.

gene). Due to its specific expression in the wound epithelium of the regenerating tail and hindlimb bud and the ability of the protein to bind some membrane receptors (see below), we named this gene *c-answer* after cold-blooded animal-specific wound epithelial receptor-binding protein), *ENSXETG00000016048* (*foxo1 homolog*, *forkhead box O1 homolog*), *ENSXETG00000006008* (*prothymosin homolog*), *ENSXETG00000023966* (*sfrpx*, *secreted frizzled-related*), *ENSXETG00000025525* (*pnhd*, *pinhead, secreted inhibitor of Wnt/β-Catenin signaling*), *ENSXETG00000030282* (*E3 ubiquitin/ISG15 ligase TRIM25*), ENSXETG00000031627 (uncharacterized *F-box protein*), ENSXETG00000033120 (uncharacterized *endonuclease 4 isoform X7homolog*), and *ENSXETG00000033543* (uncharacterized *paramyosin-like*).

To confidently confirm that all of the detected genes have no orthologs in warm-blooded animals, they have all been checked for the absence of possible local synteny with genes in warm-blooded animals in more stringent conditions, using enlarged to 5 Mbp window, $l = l_1 = l_2$.

Notably, three genes missing in all placental mammals and in the majority of birds and reptiles (*ag1*, *ras-dva1*, and *ras-dva2*), which we empirically found previously (Ivanova et al., 2013, 2015, 2018; Tereshina et al., 2014), were also detected by our program as lost genes if some species of warm-blooded animals in which these genes were present were excluded before processing. Thus, these testing results confirm the validity of our approach.

Some of the Identified Genes Are Involved in the Processes of Regeneration of the *X. laevis* Tadpole Hindlimb Bud and Tail

The temporal expression patterns of orthologs of the *X. tropicalis* genes identified during the bioinformatics screening were analyzed during regeneration of the *X. laevis* tadpole tail by qRT-PCR. In the experiment, we used tissue samples cut from the stumps of tails on days 0, 1, 3, and 6 post-amputation (dpa), where the 0 dpa sample was the piece of stump harvested immediately after amputation. These 0 dpa samples were used as controls. As a result, we detected a significant increase in the expression levels of four genes relative to the expression of two housekeeping genes (*ef1alfa* and *odc*) on 1 dpa in the tail regenerates compared to those at 0 dpa: *paramyosin-like*, *c-answer*, *sfrpx*, and *foxo1 homolog* (Figure 2A). By 6 dpa, the expression levels of all these genes returned to their respective basal levels. The revealed increase in expression of these four genes suggests their potential roles in regeneration. No activation during regeneration was revealed for other five genes identified in our bioinformatic screening (see as an example the results of qRT-PCR for *pnhd* on Figure 2A).

Then, to verify whether the identified genes were critical for regeneration, we tested the effects of their downregulation on tadpole tail regeneration. To this end, we injected early *X. laevis* embryos with anti-sense morpholino oligonucleotides (MOs) to the mRNAs of these genes (see testing results for the MO efficiency and specificity in Figures S1A–S1D). The tadpole tails injected with MOs were amputated at stage 41, and their regeneration capacity was compared. However, we were able to perform these experiments only for *paramyosin-like*, *c-answer*, and *sfrpx*. Injection of MOs for *foxo1 homolog* had dramatic effects that did not allow the embryos to develop after gastrulation. Therefore, we concluded that this gene probably performed some housekeeping function that was critical for embryonic development.

Among the three genes whose knockdown had no lethal effects, we observed clear inhibition of tail regeneration only for knockdown of *c-answer* (Figures 2B and 2B'). Based on these
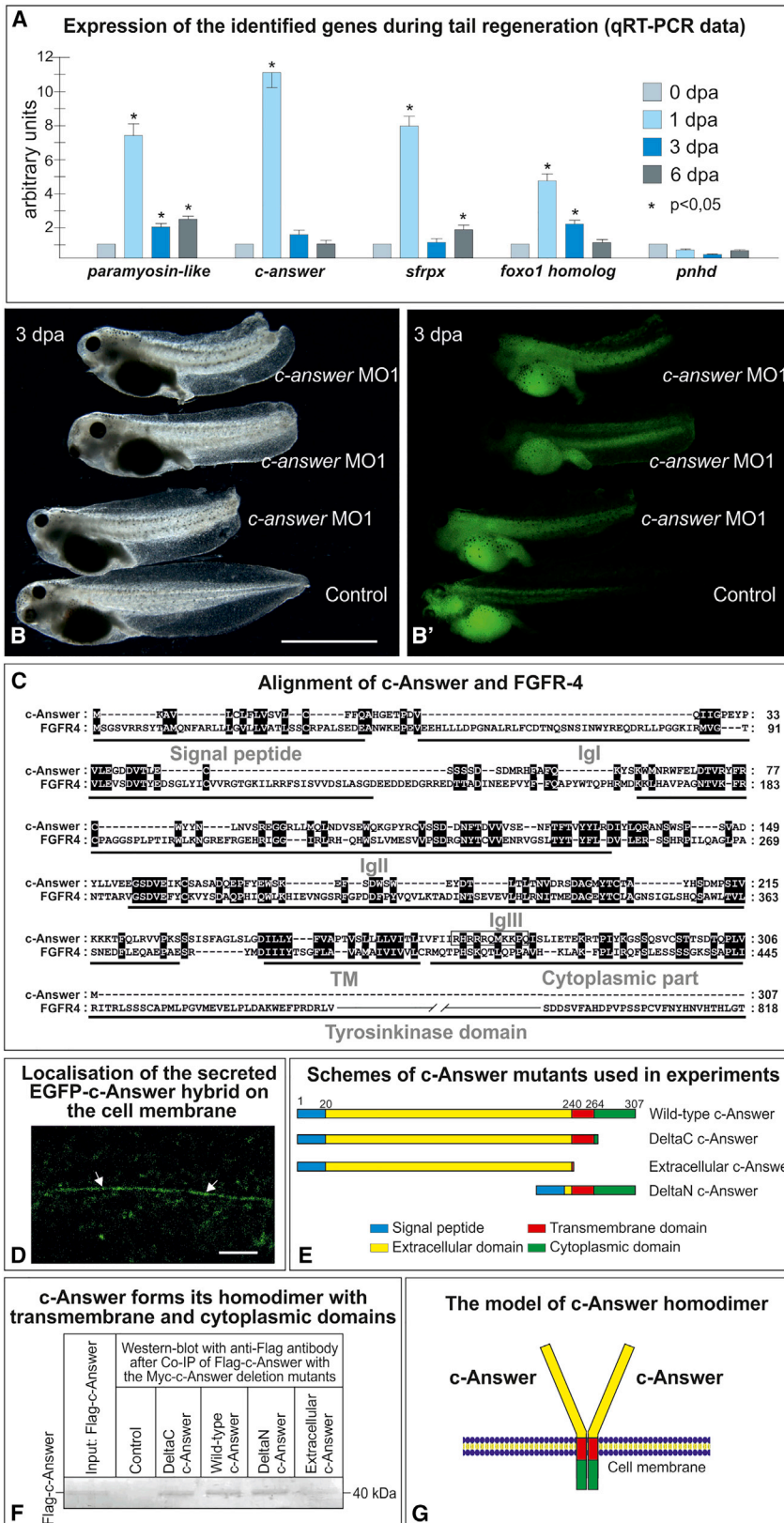
Figure 2. Investigation of the Expression and Function of Genes Identified As a Result of Bioinformatic Screening during *X. laevis* Tadpole Tail Regeneration and Analysis of the Structure of c-Answer Protein

(A) qRT-PCR analysis of *c-answer* expression in the blastema during tail regeneration at the indicated day post-amputation. All of the results were normalized to the geometric mean expression of two housekeeping genes (*ef1alfa* and *odc*) as previously described (Ivanova et al., 2013).

(B) Injection of anti-sense morpholino oligonucleotides to *c-answer* mRNA inhibits tadpole tail regeneration.

(B') Fluorescent image of the same embryos shown in (B) demonstrating the distribution of the co-injected tracer fluorescein lysine dextran (FLD).

(C) Alignment of c-Answer with FGFR4.

(D) Localization of the secreted hybrid of EGFP and c-Answer on the membrane of the animal cap cell as revealed by confocal microscopy.

(E) Schematics of the c-Answer deletion mutants used in the co-immunoprecipitation and functional analysis experiments.

(F) Western blotting with a FLAG-antibody after coIP of FLAG-c-Answer with different Myc-tagged c-Answer deletion mutants, as shown in (C).

(G) The c-Answer homodimer model according to the coIP results shown in (D).

Scale bars: 1 mm in (B) and 200 nm in (D).

See also Figures S1 and S2.

results, we chose this gene for further investigation. To be sure in the specificity of *c-answer* MO effects, we also used another MO to mRNA of this gene: *c-answer* MO2. Testing of the efficiency of this MO2 is shown in Figure S1E.

Finally, to understand if the observed effect was the result of *c-answer* downregulation during the process of regeneration or if it was a remote consequence of the inhibition of the function of this gene in early embryogenesis, we downregulated *c-answer* by injecting *c-answer vivoMO* into distal part of the tails' stumps immediately after amputation at the tadpole's stages 40–42 according to the protocol developed earlier (Ivanova et al., 2018). As a result, we observed similar tail regeneration inhibition, as was seen in case of the conventional *c-answer* MO injections in early embryos (Figures S1F–S1I). Thus, one may conclude that the function of *c-answer* is essential directly during the tail regeneration.

c-Answer Is a Homodimer-Forming Transmembrane Protein Homologous to FGFRs

By analyzing the protein primary structure, we have established that c-Answer is a transmembrane protein, which has an overall structure resembling that of the single-path receptors. In all tested species, including *X. tropicalis* (GenBank: MG865735), *X. laevis* (GenBank: MG865736), *Danio rerio* (GenBank: MG865737), *Ambystoma mexicanum* (GenBank: MG865738), *Gekko japonicus* (GenBank: MK318921), and *Python bivittatus* (GenBank: MK318922), c-Answer contains the amino terminal signal peptide, two immunoglobulin (Ig)-like domains, a single transmembrane helix, and a short cytoplasmic region (Figures 2C and S2A). Although the putative extracellular region of c-Answer has no high homology with any known proteins, it appears to be most homologous to the Ig-like D2 and D3 domains of the extracellular regions of the FGFRs 1–4 and demonstrates the highest homology with the FGFR4 receptor (Figure 2C). Interestingly, these FGFR domains are critical for their binding to fibroblast growth factors (FGFs) (Johnson et al., 1990; Lemmon and Schlessinger, 2010)

Meanwhile, the cytoplasmic part of c-Answer has none of the tyrosine kinase domains characteristic of FGFRs. Thus, c-Answer can be putatively considered a receptor lacking intrinsic kinase activity. Interestingly, a highly positively charged juxtamembrane domain was revealed in the cytoplasmic part of c-Answer (Figures 2C and S2A).

To confirm experimentally that c-Answer is a transmembrane protein, we investigated subcellular localization of its hybrid with EGFP. To this end, EGFP-Answer was translated from the synthetic mRNA injected into the *X. laevis* embryos, and its localization in the animal hemisphere at the middle gastrula stage was observed via confocal microscopy. Indeed, a portion of EGFP-c-Answer was localized to cell membranes (Figures 2D and S2B for EGFP-c-Answer co-localization with the cytoplasmic membrane marker, mKate2-mem).

Transmembrane proteins of multicellular organisms often form homodimers (Pogozheva and Lomize, 2018). Therefore, we tested the ability of c-Answer to form a homodimer by co-immunoprecipitation (coIP) in embryos co-injected with mRNA encoding the wild-type c-Answer or its deletion mutants tagged using the Myc and FLAG epitopes, respectively. The following

FLAG-tagged c-Answer deletion mutants were used: extracellular c-Answer (c-Answer lacking the transmembrane and cytoplasmic domains), deltaN-c-Answer (c-Answer lacking the extracellular domain), and deltaC-c-Answer (c-Answer lacking the cytoplasmic domain) (Figure 2E). As a result, we established that the transmembrane and cytoplasmic domains of c-Answer were primarily involved in homodimer formation (Figures 2F and 2G).

Analysis of c-Answer Expression in Embryonic Development and during Regeneration

According to a genome-scale database of mRNA dynamics during *X. laevis* embryogenesis (Yanai et al., 2011), *c-answer* transcripts are present at a low level in the embryo until the late gastrula stage, when their concentration begins to increase gradually (Figure 3A). Consistent with this finding, *c-answer* transcripts were revealed by whole-mount *in situ* hybridizations beginning from the late gastrula stage throughout the dorsal ectoderm, with maximum expression located within the neurectoderm (Figure 3B). As shown by sectioning neurula stage embryos, *c-answer* expression was localized primarily in cells of the inner layer of the anterior neurectoderm and the trunk axial mesoderm (Figure 3C).

The single-cell RNA sequencing results (Briggs et al., 2018) also support our experimental data on *c-answer* expression. Although according to the single-cell data *c-answer* expression at a low level is distributed rather uniformly and is present in all germ layers (ectoderm, mesoderm, and endoderm), some tissue subtype clusters, especially those of the presumptive forebrain with eye primordia and the cement gland, are enriched in *c-answer* transcripts, which was confirmed by their co-localization in the same clusters with transcripts of the corresponding marker genes *foxg1* (anterior neural fold, telencephalon, eyes, and brachial arches), *rax* (anterior neural fold and eye primordia), and *ag1* (cement and hatching glands) (Figure S3).

Consistent with the qRT-PCR analysis results (Figure 2A), an increase in *c-answer* expression was seen during regeneration of the stumps of both tails and the hindlimb buds at 1 dpa. Importantly, in both cases, *c-answer* was expressed in the wound epithelium, which indicated that the involvement of *c-answer* in the regulation of this tissue was critical for regeneration (Figures 3F–3I).

In sum, the *in situ* hybridization and qRT-PCR experiment results and the single-cell sequencing data indicate that *c-answer* may be involved in the early development of the forebrain, eyes, and cement gland as well as in body appendage regeneration.

c-Answer Downregulation and Overexpression Have Opposite Effects on Brain Development and Regeneration

To investigate the physiological function of *c-answer* during CNS development and tail regeneration, we analyzed the effects of its downregulation using two different methods: injections of antisense MOs to *c-answer* mRNA and the CRISPR/Cas9 knockout of the *c-answer* gene. Two different MOs (see the MO1 and MO2 structures in Figure S1) and two different synthetic (or single) guide RNA (sgRNA) target sites were used in these experiments (see positions of the target sites in Figure S4).
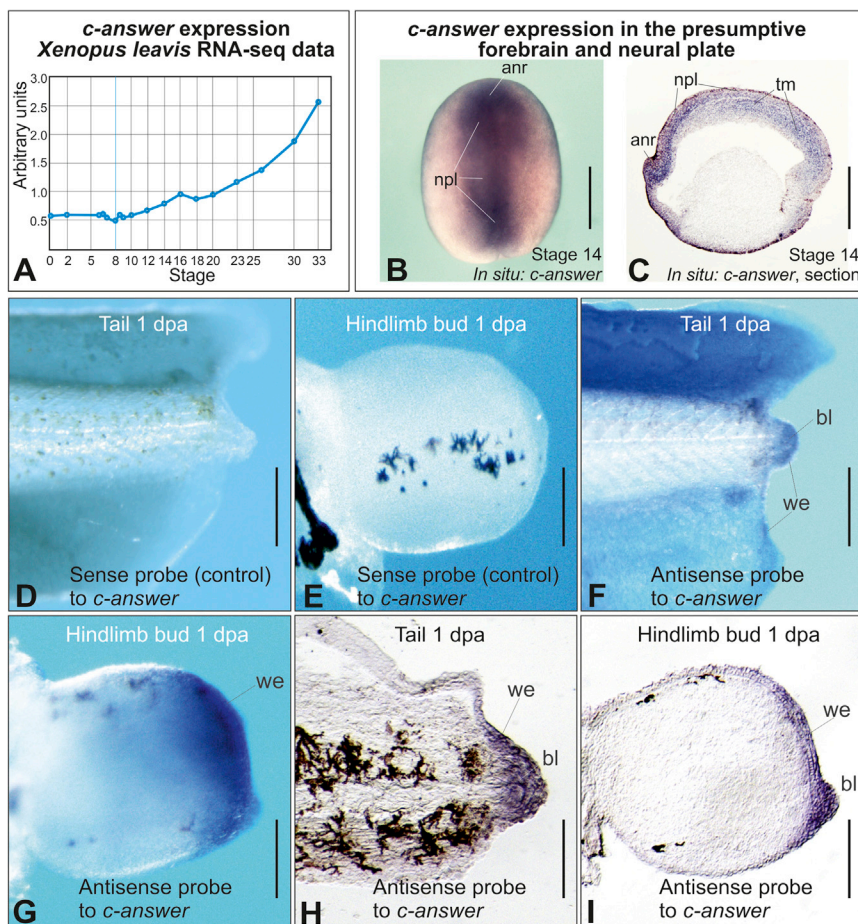
Figure 3. Analysis of *c-answer* Expression during the Early Embryonic Development and in the Processes of Tail and Hindlimb Bud Regeneration of *X. laevis*

(A) Temporal expression of c-Answer as revealed by RNA sequencing (RNA-seq) analysis (Yanai et al., 2011).

(B and C) *In situ* hybridization with a c-Answer probe in the whole-mounted middle neurula (B) and the frozen sagittal section of the embryo at the middle neurula stage (C). Anr, anterior neural ridge; npl, neural plate; tm, trunk mesoderm.

(D and E) Whole-mount *in situ* hybridization with a sense probe to *c-answer* (control) of the amputated tail (D) and hindlimb bud (E) at day 1 post-amputation.

(F and G) Whole-mount *in situ* hybridization with an anti-sense probe to *c-answer* of the amputated tail (F) and hindlimb bud (G) on day 1 post-amputation. Increased c-Answer expression is observed in the wound epithelium (we) and blastema (bl).

(H and I) Frozen histological sections of the regenerating tail (H) and hindlimb bud (I) hybridized whole-mount with an anti-sense probe to *c-answer*.

Scale bars: 500 μm in (B) and (C) and 50 μm in (D)–(I).

See also Figure S3.

As a result, we observed tadpoles with a specific phenotype (morphant/CRISPRants thereafter) that included reduction of the telencephalon, eyes, and cement gland together with a diminishing overall body size in each case with these gene downregulation methods (Figures 4A, 4B, S6A, and S6B). However, the frequencies of the mutant phenotype differed sharply in experiments with MO and CRISPR/Cas9. Thus, 75% and 80% of embryos injected with MO1 and MO2, respectively, demonstrated this phenotype, which was especially evident when the MO was injected in only one of the two blastomeres at the 2-cell stage (Figure 4A). At the same time, a similar phenotype was detected in only 21% or 23% of embryos out of n = 700 or n = 600 accordingly in 6 experiments injected with the mixture of sgRNA1 or sgRNA2 and the Cas9 protein, respectively. Notably, this comparatively low percentage of the mutant phenotype in cases of CRISPR/Cas9 knockout corresponded well with the low number of embryos with the percentage of mutations in the *c-answer* gene exceeding 70% (see results of embryos genotyping by next generation sequencing after the CRISPR/Cas9 procedure in Figure S5). This finding indicates that the phenotypic effect of *c-answer* knockout may be primarily observed in embryos with *c-answer* mutated in the majority of their cells. Importantly, tadpoles demonstrated diminished size survive at least up to meta-

morphosis, which indicates that the knockout of *c-answer* was not lethal, which in turn confirms that animals with mutations in this gene could have indeed survived in evolution.

When we tested the regeneration capacity of tadpoles with downregulated *c-answer*, a retardation of tail regeneration was observed in 77% and 80% of the tadpoles (stage 41) injected with MO1 and MO2, respectively, and in 92% and 94% of the tadpoles with *c-answer* knocked out by sgRNA1 and sgRNA2, respectively, with the CRISPR/Cas9 method (Figures 4C–4E, S6B, S6D, and S6F). This much higher percentage of abnormal tail regeneration in the case of CRISPR/Cas9 treatment compared to the low overall percentage of embryos with the mutant phenotype obviously occurred because the regeneration experiments were performed using preliminarily selected tadpoles, which had a reduced body size and thus presumably had *c-answer* knocked out in most of their cells.

The involvement of *c-answer* in the regulation of regeneration was also confirmed at the molecular level, through demonstration of the inhibitory effect of *c-answer* downregulation on the activation of several important regulators of the tail regeneration (Figure 4F). Interestingly, as one may see, downregulation of *c-answer* resulted in a strong inhibition of two other genes missing in the warm-blooded vertebrates: *ag1* and *ras-dva1*. This indicates possible tight coupling of these three genes in the same regulatory gene net and allows to suppose that the loss of one of them during evolution might stimulate the following loss of two others.
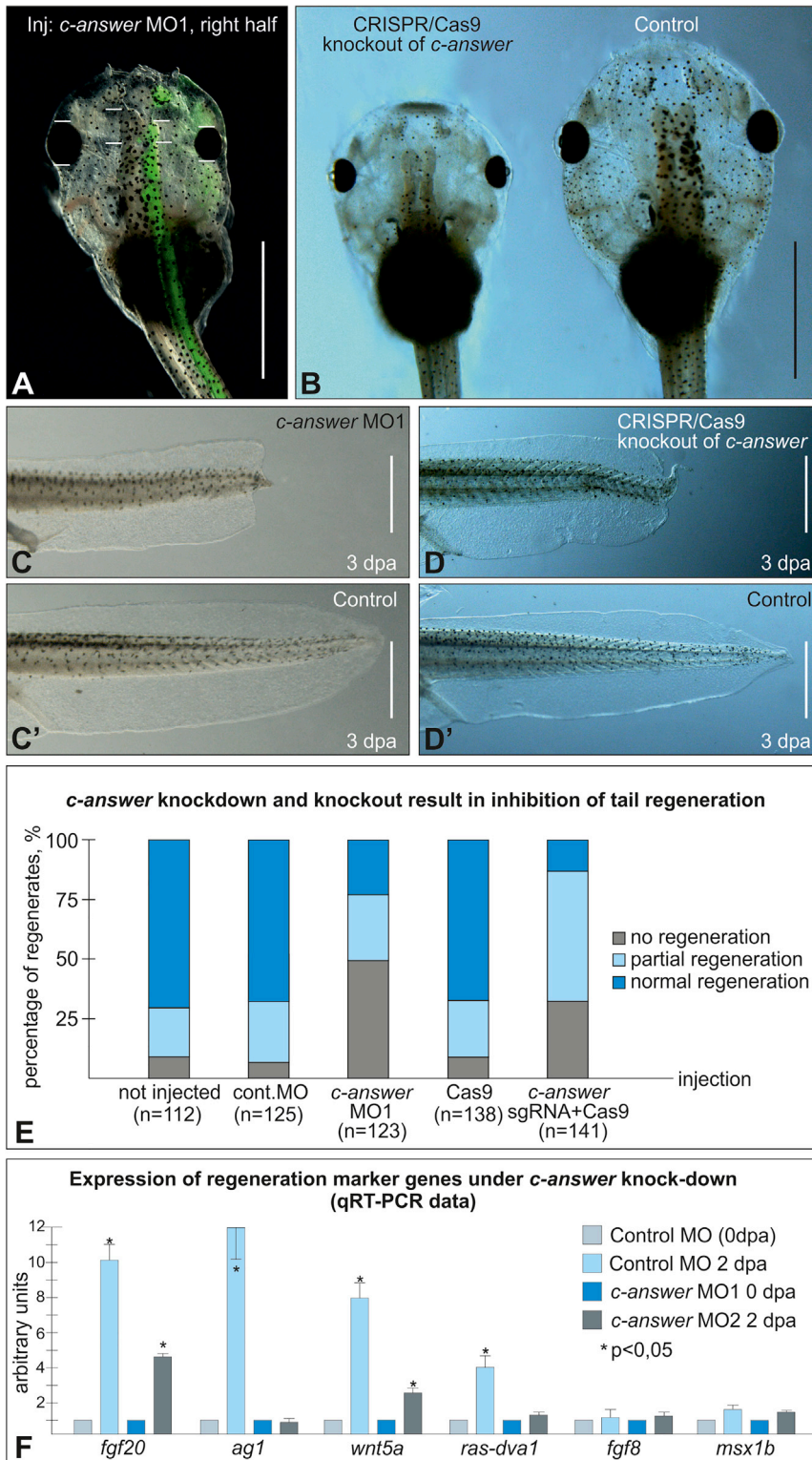
Figure 4. Effects of *c-answer* Downregulation by Knockdown with Anti-Sense Morpholino Oligonucleotides and the CRISPR/Cas9 Knockout on Tadpole Brain Development and Tail Regeneration

(A) *c-answer* knockdown with *c-answer* MO1 injections into the dorsal right blastomere at the 4-cell stage results in diminishing of the overall tadpole size, especially of the forebrain and eye (n = 180; 85% morphants), compared to that of the left side (control). Overlay with the fluorescent image demonstrates the distribution of the co-injected tracer FLD.

(B) Tadpole in which c-Answer was knocked out in 2nd exon with CRISPR/Cas9 technology has a smaller size then the wild-type tadpole at the same stage.

(C, C', D, and D') Tail regeneration in tadpoles with *c-answer* knockdown (C) or knockout (D) is inhibited compared to that of the wild-type control (see E for quantification).

(E) Diagram showing the distribution of tail regeneration phenotypes in tadpoles injected with the indicated MO or components of the CRISPR/Cas9 system.

(F) Analysis of the indicated marker genes expression by qRT-PCR in tips of the amputated tails of *Xenopus laevis* tadpoles at stage 41 (see for the scheme of the experiments and for specific primers structures; Ivanova et. al. 2018). The results were normalized by geometric median of *odc* and *ef1alfa* housekeeping genes expression described in Ivanova et al. (2013). The expression level of each gene at 0 dpa in the control piece of tissue was taken as one arbitrary unit.

Scale bars: 1 mm in (A) and (B) and 250 μm in (C)–(D').

Figures S1, S4, S5, and S6.

Given that *c-answer* is intensively expressed in the presumptive forebrain region beginning from the neurula stage, we verified whether the observed forebrain malformations were caused by the lack of *c-answer* activity during neurulation. To this end, we studied the effects of the *c-answer* MO1 injections on expression of forebrain-specific genes at the midneurula stage. Consistent with the effects observed in tadpoles, a moderate reduction of the telencephalic regulator *foxg1* (80%, n = 68), eye regulators *rax* (70%, n = 58), and *pax6* (75%, n = 64) expression was detected (Figures 5A and 5B; Figure S7A). Interestingly, in all cases, the areas in which *c-answer* downregulation inhibited regulator-gene expression were located in the lateral regions of their normal expression domains, corresponding to the presumptive dorsal part of the telencephalon and eyes. No inhibitory effects were seen in the medial regions, which give rise
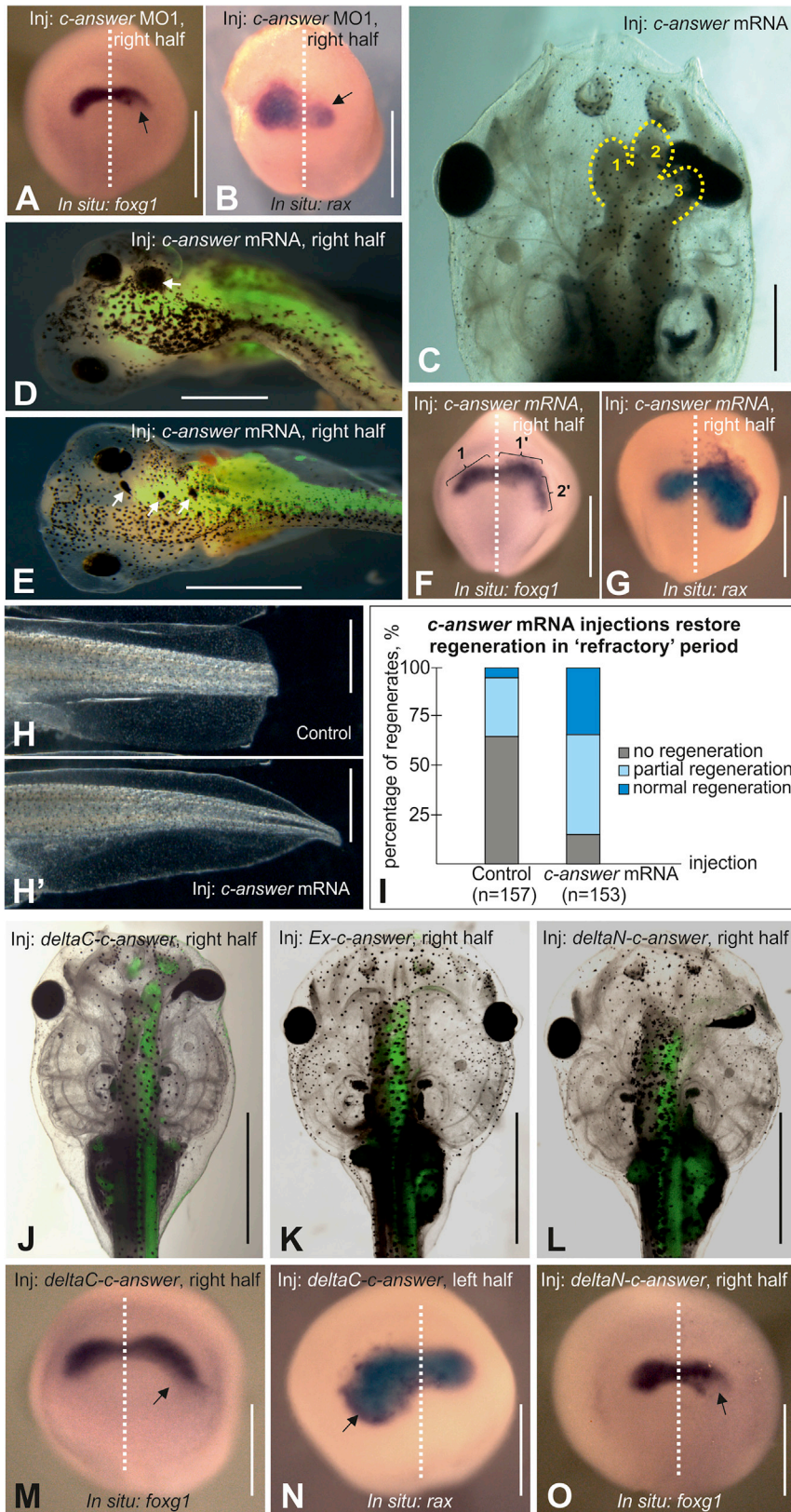
In sum, we concluded that downregulation of *c-answer* led to both a reduction of the forebrain and retardation of tail regeneration. No such effects were observed in the embryos injected with the control *mis-c-answer* MO (not shown).

Figure 5. Effects of the Overexpression of c-Answer and Its Deletion Mutants on Tadpole Brain Development and Tail Regeneration

(A and B) Inhibition of *foxg1* (A) and *rax* (B) expression on the right side of middle neurula embryos by injection into the right dorsal blastomere at the 4-cell stage.

(C) Overexpression of wild-type *c-answer* results in the development of an ectopic telencephalic hemisphere (3) in addition to normal hemispheres (1 and 2).

(D and E) Ectopic eye differentiation in tadpoles overexpressing wild-type *c-answer* in the form of one normal-sized eye (D) or three small eyes (E).

(F and G) Ectopic expression of *foxg1* (F) and *rax* (G) on the right side of middle neurula embryos following the injection of *c-answer* mRNA into the right dorsal blastomere at the 4-cell stage.

(H and H') Overexpression of wild-type *c-answer* rescues tail regeneration during the "refractory" period (see I for quantification).

(I) Diagram showing the distribution of the regenerating tail phenotypes in the control tadpoles and those overexpressing wild-type *c-answer*. Tails were amputated during the refractory period.

(J) Overexpression of the deltaC-c-Answer mutant leads to a telencephalon size increase and ectopic RPE differentiation.

(K) Overexpression of the extracellular domain of c-Answer results in a slight increase in the telencephalic and eye size on the injected (right) side.

(L) Overexpression of the deltaN-c-Answer mutant inhibits the development of the telencephalon and eye on the injected side.

(M and N) Ectopic *foxg1* (M) and *rax* (N) expression on the left side of middle neurula embryos injected with deltaC-c-Answer mRNA into the left dorsal blastomere at the 4-cell stage.

(O) Inhibition of *foxg1* expression in the lateral part of the endogenous right expression domain of *foxg1*.

Scale bars: 500 μm in (A), (B), (D)–(G), and (M)–(O); 1 mm in (C) and (L); and 250 μm in (H) and (H').

See also Figure S7.

predominantly to the ventral parts of the telencephalon and eyes. No effects of *c-answer* knockdown were revealed the case of *six3* (n = 60), which is expressed throughout the anterior neural ectoderm, *xanf1/hesx1* (n = 75), marking by its expression the presumptive forebrain territory, and *en1* (n = 52), whose expression marks the presumptive mid-hindbrain border (Figures S7E, S7G, and S7J). These results confirm the specificity of *c-answer* function and indicate that it seemingly operates downstream independently of *six3* and *xanf1/hesx1*, whose mutations are known to cause in humans holoprosencephaly and septo-optic dysplasia, respectively (Wallis et al., 1999; Dattani et al., 1998) (Figure 3A). Notably, we showed earlier that *xanf1/hesx1* in its turn also operated downstream or independently of *six3*, as downregulation of *xanf1/hesx1* did not affect *six3* expression (Ermakova et al., 2007).

Then, we investigated *c-answer* gain-of-function effects by unilateral *c-answer* mRNA overexpression. As a result, two types of effects were revealed in the head region. First, an increase in telencephalic differentiation ranging from a slight increase of the telencephalon on the injected side to an additional part of the telencephalon was detected in embryos (n = 74, 70%) (Figure 5C). In addition, ectopic cement glands were frequently observed in these embryos (n = 74, 52%). Second, a range of eye phenotypes was observed on the injected side in almost all of these embryos (Figures 5D and 5E). These phenotypes included disrupted eye development, ectopic retinal pigment epithelium (RPE), RPE extensions, expansion of the RPE, positioning of the eye cup adjacent to forebrain, and even development of a secondary eye in rare cases (Figure 5D). Importantly, these malformations were accompanied by expansion of the expression zones of the telencephalic regulator *foxg1* (n = 63, 79%) and eye regulators *rax* (n = 54, 89%) and *pax6* (n = 60, 80%) during neurulation (Figures 5F, 5G, and S7B). However, as in the case of *c-answer* knockdown, no effects were detected for *six3* (n = 57) and *en1* (n = 48) (Figures S7H and S7K).

To verify that *c-answer* overexpression could influence tail regeneration, we studied the ability of exogenous *c-answer* to rescue regeneration during the so-called refractory period (stages 45–47), when tail regeneration appears to be blocked for natural reasons, using the previously developed approach (Ivanova et al., 2018). As a result, we observed a significant increase in the number of tadpoles with a restored tail regeneration capacity during the refractory period (Figures 5H and 5I).

To test if the injected c-answer mRNA can be really preserved in the tail tissues until stages when tail amputations were done, we tested the presence of this mRNA by qRT-PCR in the tail tips of the control and injected tadpoles. The qRT-PCR signal of *c-answer* mRNA is much higher in the tail tips of stage 46 tadpoles developed from the embryos injected with c-Answer mRNA than in the control ones. This confirms that the injected c-Answer mRNA is indeed preserved at least until stage 46 in concentrations several times higher than the endogenous c-answer mRNA.

Obviously, taken together, all this indicates that *c-answer* can stimulate regeneration during the refractory period.

c-Answer Deletion Mutants Require a Transmembrane Domain to Influence CNS Development and Regeneration

Because c-Answer has extracellular and cytoplasmic regions, we decided to test the effects of deletion mutants lacking various domains to shed light on the function of c-Answer at the molecular level. Therefore, we investigated the effects on brain development and tail regeneration of the extracellular c-Answer (c-Answer lacking the transmembrane and cytoplasmic domains), deltaN-c-Answer (c-Answer lacking the extracellular domain), and deltaC-c-Answer (c-Answer lacking the cytoplasmic domain) mutants (Figure 2E).

When *deltaC-c-answer* mRNA was injected, we observed effects resembling those of wild-type c-Answer overexpression, i.e., an increase in the telencephalon (n = 75, 75%) and ectopic differentiation of the eye pigment epithelium (n = 83, 65%) (Figure 5J). However, in contrast to the wild-type *c-answer* mRNA injections, differentiation of the ectopic pigment epithelium was always observed in the vicinity of the normal eye. Ectopic retinal pigment epithelial differentiation was never detected far from the normal eye, like it was observed in case of the wild-type c-Answer mRNA injections. Moreover, well-structured secondary eyes were never detected in these embryos. At the same time, the eye on the side injected with *deltaC-c-answer* mRNA was frequently enlarged, similar to that in the embryos injected with wild-type *c-answer* mRNA (Figure 5J). Consistent with the increased telencephalon and eyes, an expansion of *foxg1* (n = 60, 78%), *pax6* (n = 55, 75%), and *rax* (n = 58, 80%) expression was detected on the injected side at the neurula stage (Figures 5M and 5N; Figure S7D). Thus, the deltaC mutant injections resembled the wild-type *c-answer* mRNA injections but differed from the latter by exhibiting a more severe and frequent increase in the telencephalon and less ectopic eye differentiation.

Interestingly, the effects of deltaC-c-Answer upon tail regeneration resembled those of *c-answer* MOs. Namely, retardation of regeneration was seen in tadpoles developing with deltaC-*c-answer* mRNA injected into their blastomeres, which gave rise to the tail tissues (28% no regeneration, 54% partial regeneration, and 18% normal regeneration; 36/70/24).

Given that the deltaC mutant of c-Answer, which contains its extracellular part and the transmembrane domain, caused effects resembling those of wild-type c-Answer, we tested whether the same effects could be elicited by the extracellular part of c-Answer (extracellular mutant). However, none of the brain outgrowth and ectopic eye differentiation seen in the case of wild-type c-Answer or its deltaC mutant were observed. Additionally, we did not observe any retardation in tail regeneration. At the same time, a slight increase in the telencephalon and eye size was seen in 30% of the injected tadpoles (n = 83) (Figure 5K). Therefore, we conclude that although the extracellular part of c-Answer is necessary for its function, it cannot effectively operate without the transmembrane domain.

Then, we investigated the effects of the c-Answer mutant lacking the extracellular part (deltaN mutant). As a result, a set of abnormalities resembling those elicited by the *c-answer* MO, including the diminishing telencephalon and eye size (n = 80, 75%) (Figures 5L and 5O) and retardation of regeneration (24% no regeneration, 49% partial regeneration, and 27% normal
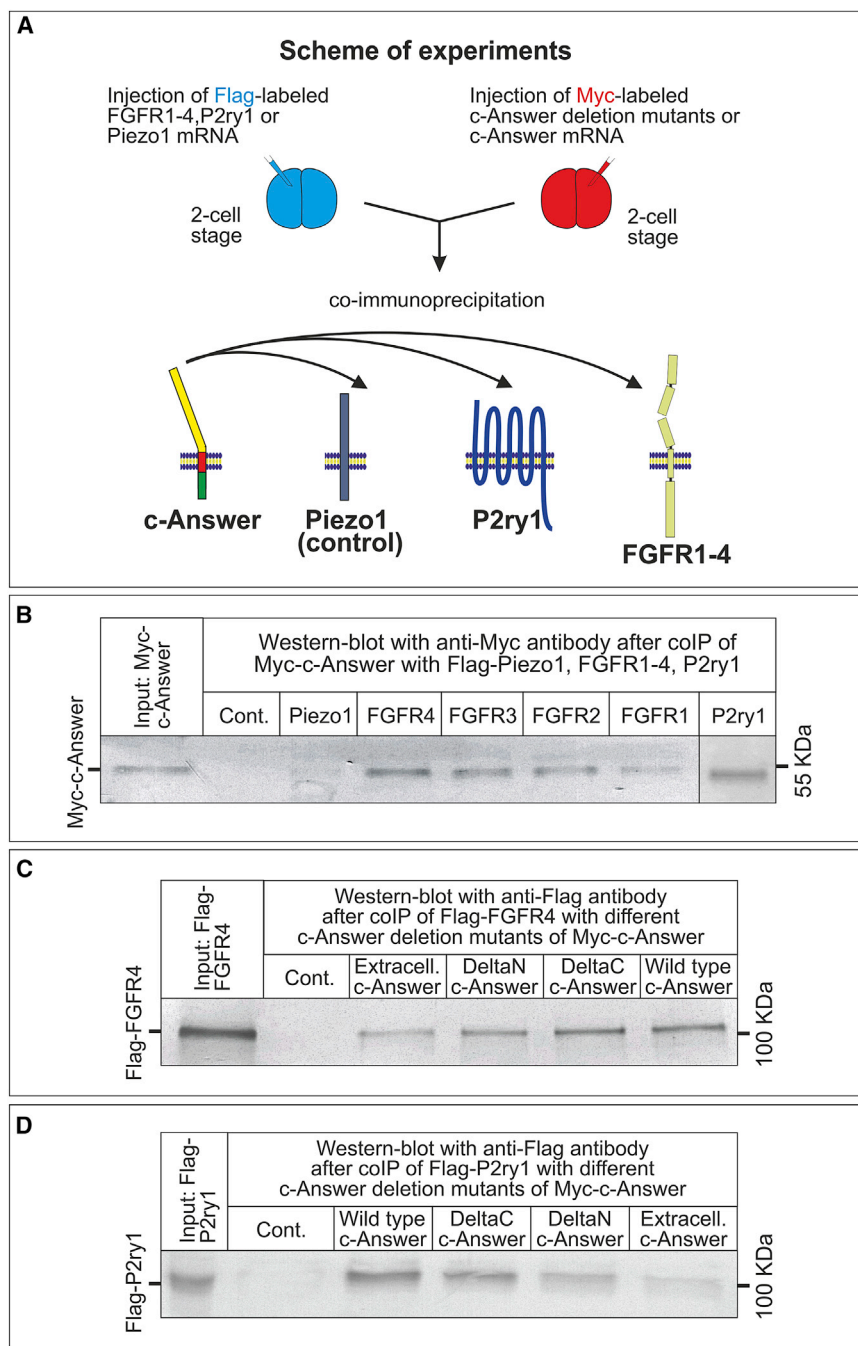
Figure 6. Analysis of the Binding Capacity of c-Answer and Its Deletion Mutants to the FGFR1–4 and P2ry1 Receptors

(A) Schematic of the experiments. FLAG- and Myc-tagged proteins were separately expressed from synthetic mRNAs in the embryos, and the embryonic extracts were mixed for coIP and analyzed by western blotting.

(B–D) Western blotting analysis with anti-Myc or anti-FLAG antibodies after coIP of Myc-c-Answer with FLAG-Piezo1, FGFR1–4, P2Y1 (B), of FLAG-FGFR4 with different c-Answer deletion mutants of Myc-c-Answer (C), or FLAG-P2ry1 with different deletion mutants of Myc-c-Answer (D).

Cont. means the control western blotting with indicated antibody (Myc or FLAG) after application on anti-Myc or anti-FLAG resin of only Myc-c-Answer, FLAG-FGFR4, or FLAG-P2Y1, without preliminary coIP.

c-Answer Interacts with FGFR1–4 and P2ry1 but Not with Fgf8

Based on the experimental results regarding c-Answer misexpression as well as its potential function as a transmembrane protein, we may speculate that it is involved in signaling that regulates the early stages of telencephalic and eye differentiation. We focused on two types of signaling because they both operate in the anterior neural plate and their disruption has effects similar to those observed in the c-Answer misexpression experiments. The first type is Fgf8 signaling, which activates expression of the telencephalic master regulator *foxg1* in cells of the anterior neural border (ANB) and is essential for the ventral and dorsal telencephalon development (Aboitiz and Montiel, 2007; Danesin and Houart, 2012; Houart et al., 1998; Lupo et al., 2002; Shimamura and Rubenstein, 1997). The second is ADP purinergic signaling through the transmembrane receptor P2ry1, which was shown to play important role during eye and telencepalic differentiation (Harata et al., 2016; Massé et al., 2007). Similar to overexpression of wild-type *c-answer*, overexpression of *fgf8* enhance the *foxg1* expression (Shimamura and Rubenstein, 1997), whereas overexpression of *p2ry1* induces ectopic eye differentiation, as was observed in the case of the wild-type *c-answer* mRNA injections (Massé et al., 2007). Recently, knockdown of *p2ry1* was shown to result in a reduction of the eye and telencephalon, similar to the case of *c-answer* downregulation (Harata et al., 2016).

Given this correlation of misexpression effects, we tested whether c-Answer could directly interact with the extracellular

regeneration; 29/59/32), was observed. Obviously, this finding indicates that the c-Answer deltaN mutant is likely to operate as its dominant-negative mutant.

In summary, the data obtained indicate that the transmembrane part of c-Answer is essential for its activity. The isolated extracellular domain had no significant effect on brain development and regeneration. In turn, c-Answer deprived of the extracellular (deltaN) or cytoplasmic part (deltaC) operated in the case of regeneration as the dominant-negative mutants that could seemingly compete with wild-type c-Answer.
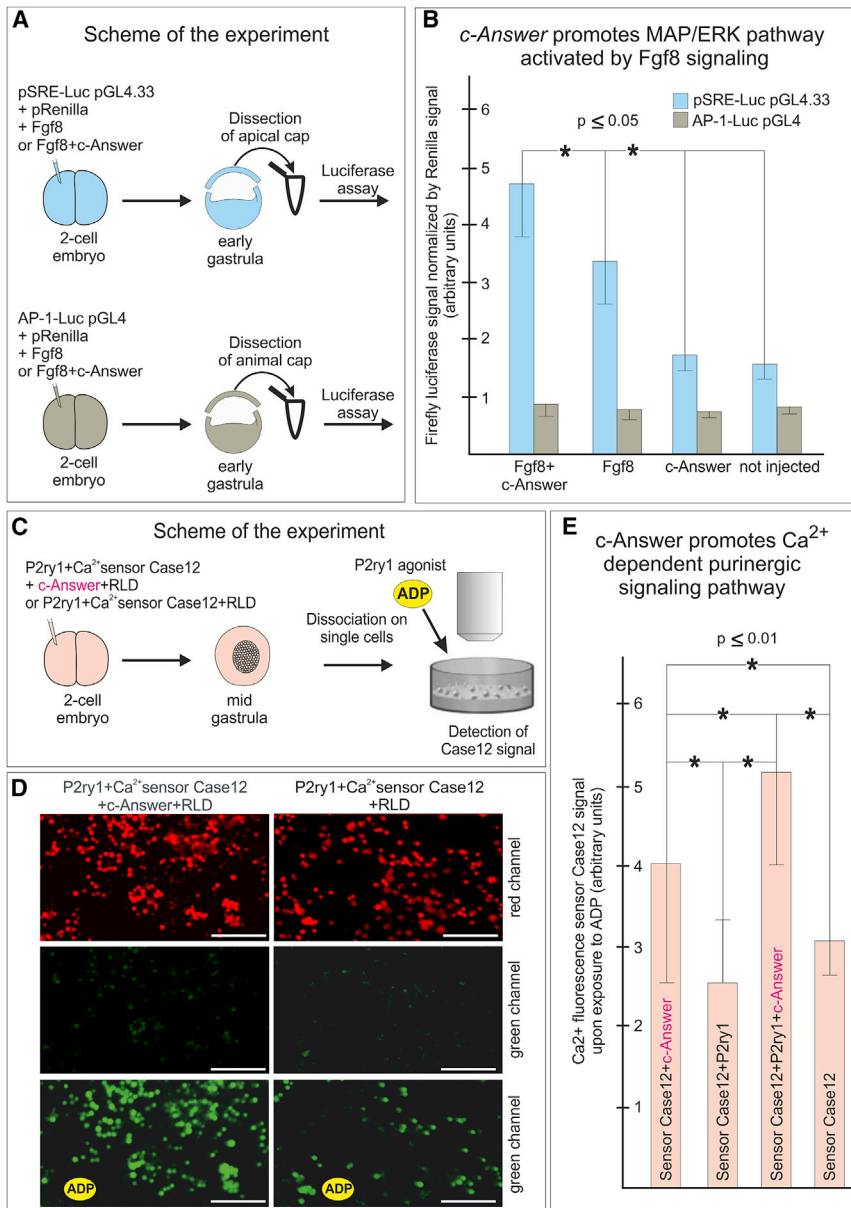
Figure 7. c-Answer Promotes Signaling through the FGFR4 and P2ry1 Receptors, Related to Video S1 and Video S2

(A) Schematic of the experiment performed to analyze the effects of c-Answer on the expression of the MAP/ERK pathway luciferase reporter pSPE-Luc pGL4.33. The stress-activated reporter AP-1-Luc pGL4 was used as a control.

(B) Diagram showing the Luc signal analysis results for the two reporters in the animal caps of embryos expressing the indicated proteins.

(C) Schematic of the experiment performed to analyze the effects of c-Answer on the Ca^{2+} flux in response to the addition of the P2ry1 agonist ADP to animal cap cells expressing the Ca^{2+} sensor Case12 and purinergic receptor P2ry1.

(D) Fluorescent images of cells expressing the indicated proteins before and after ADP addition.

(E) Diagram showing the results of Case12 signal analysis in animal cap cells expressing the indicated proteins.

Scale bars: 200 μm in (D).

that all of the tested deletion mutants could interact with FGFR4 and P2ry1, albeit with a weaker affinity than to wild-type c-Answer (Figures 6C and 6D). At the same time, among these mutants a somewhat stronger interaction was observed with the deltaC and deltaN mutants, which contained the transmembrane domain.

c-Answer Promotes FGF and P2ry1 Signaling

To understand how c-Answer could influence FGF signaling, we tested the effects of c-Answer on the downstream signaling pathways activated by Fgf8. To this end, we used the pGL4.33 (serum response elemtn [SRE]) reporter vector (Promega) in which the firefly luciferase was driven by response elements sensitive to the MAP/ERK pathway (i.e., the key pathway activated through tyrosine kinase receptors, including FGFRs). Another reporter [pGL4.44 (AP1), Promega)] sensitive to the stress-activated MAPK/JNK pathway that could not be activated by growth factors was used as a control.

In these experiments, we injected the embryos with the pGL4.33 (SRE) reporter in a mixture with either the *fgf8* or *fgf8* mixed with the *c-answer* mRNAs and then cut the animal caps off the injected embryos at the early gastrula stage. The luciferase signal was analyzed at the late gastrula stage equivalent (Figure 7A). As a result, we observed an increase in MAP/ERK in cells co-injected with the *fgf8* and *c-answer* mRNAs (Figure 7B). By contrast, no activation of the MAPK/JNK pathway was seen in the similarly arranged control experiments in which the reporter vector *fgf8* or *c-answer* mRNAs were injected alone (Figure 7B).

protein components of the Fgf8 and purinergic signaling pathways. To this end, we performed coIP experiments with a Myc epitope-tagged c-Answer with Fgf8a and Fgf8b, the receptors of Fgf8 (FGFR1–4), and the receptor of ADP (P2ry1) tagged with the FLAG epitope. All of these proteins were translated from the corresponding synthetic mRNAs injected in pairs into the early *X. laevis* embryos (see Method Details). As a result, we established that in these conditions, c-Answer bound to all FGFRs and P2ry1 but not to Fgf8 (Figures 6A and 6B).

Then, to understand which domains of c-Answer were responsible for its interactions with FGFR4 and P2ry1, we investigated the ability of Myc-tagged deletion mutants of c-Answer to bind FLAG-tagged receptors in a coIP test. As a result, we revealed

Thus, we concluded that c-Answer promoted MAP/ERK pathway activation by Fgf8 signaling.

Given that a key cytoplasmic mediator of purinergic signaling is Ca^{2+}, we arranged experiments with a Ca^{2+} fluorescent reporter to test the effects of c-Answer on this signaling pathway through the P2ry1 receptor. To this end, we injected the embryos with a mRNA encoding a Ca^{2+}-sensitive variant of the green fluorescent protein (Case9 reporter protein, Evrogen) mixed with either the *p2ry1* or *p2ry1* and the *c-answer* mRNAs. The animal caps of these embryos were dissociated into single cells in Ca^{2+}-free medium, and then the Ca^{2+} flux into the cytoplasm of these cells after application of ADP was monitored via measuring the fluorescence of the Case9 reporter (Figure 7C). As a result, we revealed a higher fluorescent signal in cells expressing exogenous c-Answer (Figure 7C; Video S1). Similar results were obtained in the whole embryos (Video S2).

DISCUSSION

Wide-Range Bioinformatics Screening for Genes Lost in Higher Vertebrates Introduces a Novel Method for Orthologous Gene Identification

In the present work, we have developed an algorithm and a computer program for a systematic search for genes lost at a certain step of evolution. The rationale for the systematic search for such genes is confirmed by our recent empirical finding that genes lost in poorly regenerating higher vertebrates (*ag1*, *ras-dva1*, and *ras-dva2*) still play important roles during body appendage restoration in well-regenerating fishes and amphibians (Ivanova et al., 2013, 2015, 2018; Tereshina et al., 2014).

As a result of the wide-range computer screening with the developed method for genes lost in warm-blooded vertebrates, we were able to identify only nine genes. Even assuming the technical limitations of the method discussed below, the number of lost genes seems to be surprisingly low. Obviously, this result confirms that complete gene deletion is a quite rare phenomenon and that indeed the phenotypic and physiological differences between different classes of vertebrates are mostly based on rearrangement of the genomic regulatory networks instead of changes in the gene repertoire.

Among the identified genes, only four demonstrated evident activation of expression at the very beginning of tadpole body appendage regeneration. However, only downregulation of *c-answer* specifically affected regeneration, whereas downregulation of the three other genes either demonstrated no pronounced effect upon regeneration (*paramyosin-like* and *sfrpx*) or resulted in severe overall damage of the embryos (*foxo1 homolog*). These results indicate that *paramyosin-like* and *sfrpx* probably operate downstream of the mechanisms responsible for regeneration. On the other hand, the loss of *foxo1 homolog* during evolution could not be an initial cause of the decrease in the regeneration ability. We speculated that *foxo1 homolog* executed some housekeeping function and was lost in a more recent step of evolution, when the genetic mechanisms were already significantly modified as a result of accumulation of some non-lethal mutations or loss of other genes.

Given that *c-answer* was lost in the ancestors of warm-blooded animals, the question arose as to what benefit warm-blooded vertebrates might acquire as a result of this loss that could compensate for a decrease in the regeneration capacity. One such benefit could be connected with a decline in the *foxg1* expression level, which we observed in embryos with downregulated *c-answer*. As shown in the mouse, a reduced *foxg1* expression level in the dorsal telencephalon is critical for development of this region, which significantly increased in size during the evolution of higher vertebrates (Danesin and Houart, 2012). As shown here, although downregulation of *c-answer* did not entirely eliminate *foxg1* expression, it elicited a significant decline in its expression level, especially in the presumptive dorsal regions of the *foxg1* expression territory (Figure 5A). Thus, we may speculate that the loss of *c-answer* in the ancestors of warm-blooded animals may have resulted in a decline in *foxg1* expression in the dorsal telencephalon, which in turn may have provided conditions for the progressive development of this brain region.

c-answer Is a Novel Regulator of Forebrain Development and Body Appendage Regeneration in *X. laevis*

c-answer encodes a transmembrane protein that has a primary structure that is most homologous to FGFR4. However, in contrast to the FGFRs, c-Answer has no D1 Ig-domain in the extracellular part or tyrosine kinase domain in the cytoplasmic part. As we have shown, overexpression of *c-answer* results in the promotion of telencephalic and eye differentiation. These effects are accompanied by an increase in the expression zones of the telencephalic regulator *foxg1* and eye regulators *rax* and *pax6*. By contrast, downregulation of *c-answer* elicits a reduction of the telencephalon and eye size. Consistent with the results that indicated the involvement of *c-answer* in the regulation of telencephalon and eye development, we demonstrated the ability of c-Answer to interact with the FGFR1–4 and P2ry1 receptors, which transmit FGF and ADP signaling.

Importantly, the stimulating effects of c-Answer on signaling pathways activated by Fgf8 and ADP were confirmed at the physiological and molecular levels. At the same time, we were unable to reveal the interaction of c-Answer with Fgf8, which is the key inducer of *foxg1* expression and telencephalic differentiation (Shimamura and Rubenstein, 1997). This result indicates that although c-Answer can stimulate Fgf8 signaling, it cannot directly interact with Fgf8 or is at least unable to interact with the latter in the absence of FGFR1–4.

Using the tadpole tail regeneration model, we determined that *c-answer* was necessary for appendage regeneration in addition to telencephalon development. Keeping in mind that FGF signaling, including Fgf8 and FGFR4, is a key regulatory component of the regeneration molecular machinery (Gorsic et al., 2008; Lin and Slack, 2008), we may suppose that in addition to telencephalon development, c-Answer plays a role as a positive modulator of FGF signaling during regeneration. c-Answer could play the role of an enhancer of Fgf8 signaling by compensating for the low levels of Fgf8 and its receptor during the earliest regeneration period. Furthermore, in conjunction with this finding, we may speculate that the loss of *c-answer* in warm-blooded animals may be one factor that reduces their body appendage regenerative capacity.

In contrast to FGF signaling, there is no direct evidence in the literature that purinergic signaling through P2ry1 is involved in regeneration. At the same time, a crucial role of P2Y receptors was shown during wound healing (Greig et al., 2003; Iwanaga et al., 2013). Therefore, we could not exclude the possibility that the positive modulation of P2ry1 by c-Answer revealed in the present work was important for regeneration.

Importantly, among a number of known transmembrane proteins that modulate the activity of FGFRs, SEF and XFLRT3 were shown to operate similarly to c-Answer during gastrulation and neurulation in *X. laevis* embryos and to interact with FGFRs to inhibit and promote their activity, respectively (Böttcher et al., 2004; Tsang et al., 2002). The authors speculated that SEF and XFLRT3 modulated FGFR signaling either directly or by recruiting some cytoplasmic cofactors. Which of these modes of activity are valid for the interactions of c-Answer with FGFR1–4 and P2ry1 should be investigated in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Manipulations with embryos and tadpoles
- METHOD DETAILS
  - Bioinformatic analysis, raw score identification, orthologs definition
  - Species used to identify lost genes
  - DNA constructs, luciferase assay
  - Synthetic mRNA and in situ hybridization
  - Morpholino oligonucleotides and analysis of malformations in regenerating tadpoles
  - qRT-PCR
  - Immunoprecipitation and antibodies
  - Confocal and fluorescent microscopy, detection of Case12 reporter protein signal
  - Cryosectioning
  - CRISPR/Cas9 knockout and genotyping
  - Single-cell data description
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.09.038.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aboitiz, F., and Montiel, J. (2007). Co-option of signaling mechanisms from neural induction to telencephalic patterning. Rev. Neurosci. 18, 311–342.

Bayramov, A.V., Eroshkin, F.M., Martynova, N.Y., Ermakova, G.V., Solovieva, E.A., and Zaraisky, A.G. (2011). Novel functions of Noggin proteins: inhibition of Activin/Nodal and Wnt signaling. Development 138, 5345–5356.

Böttcher, R.T., Pollet, N., Delius, H., and Niehrs, C. (2004). The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. Nat. Cell Biol. 6, 38–44.

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, eaar5780.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Danesin, C., and Houart, C. (2012). A Fox stops the Wnt: implications for forebrain development and diseases. Curr. Opin. Genet. Dev. 22, 323–330.

Dattani, M.T., Martinez-Barbera, J.P., Thomas, P.Q., Brickman, J.M., Gupta, R., Martensson, I.L., Toresson, H., Fox, M., Wales, J.K., Hindmarsh, P.C., et al. (1998). Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. Nat. Genet. 19, 125–133.

Ermakova, G.V., Solovieva, E.A., Martynova, N.Y., and Zaraisky, A.G. (2007). The homeodomain factor Xanf represses expression of genes in the presumptive rostral forebrain that specify more caudal brain regions. Dev. Biol. 307, 483–497.

Gorsic, M., Majdic, G., and Komel, R. (2008). Identification of differentially expressed genes in 4-day axolotl limb blastema by suppression subtractive hybridization. J. Physiol. Biochem. 64, 37–50.

Greig, A.V., James, S.E., McGrouther, D.A., Terenghi, G., and Burnstock, G. (2003). Purinergic receptor expression in the regeneration epidermis in a rat model of normal and delayed wound healing. Exp. Dermatol. 12, 860–871.

Harata, A., Nishida, H., and Nishihara, A. (2016). Purinergic P2Y receptors are involved in Xenopus head formation. CellBio 5, 49–65.

Harland, R.M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. Methods Cell Biol. 36, 685–695.

Houart, C., Westerfield, M., and Wilson, S.W. (1998). A small population of anterior cells patterns the forebrain during zebrafish gastrulation. Nature *391*, 788–792.

Ivanova, A.S., Tereshina, M.B., Ermakova, G.V., Belousov, V.V., and Zaraisky, A.G. (2013). Agr genes, missing in amniotes, are involved in the body appendages regeneration in frog tadpoles. Sci. Rep. *3*, 1279.

Ivanova, A.S., Shandarin, I.N., Ermakova, G.V., Minin, A.A., Tereshina, M.B., and Zaraisky, A.G. (2015). The secreted factor Ag1 missing in higher vertebrates regulates fins regeneration in Danio rerio. Sci. Rep. *5*, 8123.

Ivanova, A.S., Korotkova, D.D., Ermakova, G.V., Martynova, N.Y., Zaraisky, A.G., and Tereshina, M.B. (2018). Ras-dva small GTPases lost during evolution of amniotes regulate regeneration in anamniotes. Sci. Rep. *8*, 13035.

Iwanaga, K., Murata, T., Hori, M., and Ozaki, H. (2013). Purinergic P2Y1 receptor signaling mediates wound stimuli-induced cyclooxygenase-2 expression in intestinal subepithelial myofibroblasts. Eur. J. Pharmacol. *702*, 158–164.

Johnson, D.E., Lee, P.L., Lu, J., and Williams, L.T. (1990). Diverse forms of a receptor for acidic and basic fibroblast growth factors. Mol. Cell. Biol. *10*, 4728–4736.

Jun, J., Mandoiu, I.I., and Nelson, C.E. (2009). Identification of mammalian orthologs using local synteny. BMC Genomics *10*, 630.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. *47* (*W1*), W171–W174.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117–1134.

Lin, G., and Slack, J.M. (2008). Requirement for Wnt and FGF signaling in Xenopus tadpole tail regeneration. Dev. Biol. *316*, 323–335.

Lupo, G., Harris, W.A., Barsacchi, G., and Vignali, R. (2002). Induction and patterning of the telencephalon in Xenopus laevis. Development *129*, 5421–5436.

Lyubetsky, V.A., Seliverstov, A.V., and Zverkov, O.A. (2013). Transcription regulation of plastid genes involved in sulfate transport in Viridiplantae. Bio-Med Res. Int. *2013*, 413450.

Lyubetsky, V., Gershgorin, R., and Gorbunov, K. (2017). Chromosome structures: reduction of certain problems with unequal gene content and gene paralogs to integer linear programming. BMC Bioinformatics *18*, 537.

Martynova, N.Y., Eroshkin, F.M., Ermolina, L.V., Ermakova, G.V., Korotaeva, A.L., Smurova, K.M., Gyoeva, F.K., and Zaraisky, A.G. (2008). The LIM-domain protein Zyxin binds the homeodomain factor Xanf1/Hesx1 and modulates its activity in the anterior neural plate of Xenopus laevis embryo. Dev. Dyn. *237*, 736–749.

Massé, K., Bhamra, S., Eason, R., Dale, N., and Jones, E.A. (2007). Purine-mediated signalling triggers eye development. Nature *449*, 1058–1062.

Nakayama, T., Blitz, I.L., Fish, M.B., Odeleye, A.O., Manohar, S., Cho, K.W., and Grainger, R.M. (2014). Cas9-based genome editing in Xenopus tropicalis. Methods Enzymol. *546*, 355–375.

Pogozheva, I.D., and Lomize, A.L. (2018). Evolution and adaptation of single-pass transmembrane proteins. Biochim Biophys Acta Biomembr *1860*, 364–377.

Rodríguez-Trelles, F., Tarrío, R., and Ayala, F.J. (2003). Evolution of cis-regulatory regions versus codifying regions. Int. J. Dev. Biol. *47*, 665–673.

Rubanov, L.I., Seliverstov, A.V., Zverkov, O.A., and Lyubetsky, V.A. (2016). A method for identification of highly conserved elements and evolutionary analysis of superphylum Alveolata. BMC Bioinformatics *17*, 385.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Shimamura, K., and Rubenstein, J.L. (1997). Inductive interactions direct early regionalization of the mouse forebrain. Development *124*, 2709–2718.

Sive, H.L., Grainger, R.M., and Harland, R.M. (2010). Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

Tereshina, M.B., Ermakova, G.V., Ivanova, A.S., and Zaraisky, A.G. (2014). Ras-dva1 small GTPase regulates telencephalon development in Xenopus laevis embryos by controlling Fgf8 and Agr signaling at the anterior border of the neural plate. Biol. Open *3*, 192–203.

Tsang, M., Friesel, R., Kudoh, T., and Dawid, I.B. (2002). Identification of Sef, a novel modulator of FGF signalling. Nat. Cell Biol. *4*, 165–169.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327–335.

Wallis, D.E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E.H., Rommens, J., and Muenke, M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. Nature Genet. *22*, 196–198.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Xanthos, J.B., Kofron, M., Tao, Q., Schaible, K., Wylie, C., and Heasman, J. (2002). The roles of three signaling pathways in the formation and function of the Spemann Organizer. Development *129*, 4027–4043.

Yanai, I., Peshkin, L., Jorgensen, P., and Kirschner, M.W. (2011). Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev. Cell *20*, 483–496.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46* (*D1*), D754–D761.

Zverkov, O.A., Lyubetsky, V.A., and Seliverstov, A.V. (2012). Plastid-encoded protein families specific for narrow taxonomic groups of algae and protozoa. Mol. Biol. *46*, 717–726.

Zverkov, O.A., Seliverstov, A.V., and Lyubetsky, V.A. (2015). A Database of Plastid Protein Families from Red Algae and Apicomplexa and Expression Regulation of the moeB Gene. BioMed Res. Int. *2015*, 510598.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
EZview Red Anti-c-Myc Affinity Gel	Sigma-Aldrich	Cat#E6654; RRID:AB_10093201
EZview Red ANTI-FLAG M2 Affinity Gel	Sigma-Aldrich	Cat#F2426; RRID:AB_2616449
Bacterial and Virus Strains		
DH5α competent *E. coli*	Thermo Fisher Scientific	Cat#18265017
Biological Samples		
Apical cap cells from gastrula stage of embryo	this paper	N/A
Chemicals, Peptides, and Recombinant Proteins		
Tricaine	Sigma-Aldrich	Cat#E10521-10G
Fluorescein Lysine Dextran (FLD) 40 kD	Invitrogen	Cat#D1845
3xMYC peptide	Sigma	Cat#M2435
3xFLAG peptide	Sigma	Cat#F4799
Neg-50 (Richard-Allan Scientific)	Thermo Scientific	Cat#6502
Critical Commercial Assays		
GeneArtTM PlatinumTM Cas9 Nuclease Ready-to-transfect wild-type Cas9 protein for performing CRISPR/Cas9-mediated genome editing	Invitrogen	Cat#B25640, B25641
Dual-Luciferase Reporter Assay System	Promega	Cat# E1910
mMESSAGE SP6 Transcription Kit	Ambion	Cat#AM1340
Experimental Models: Organisms/Strains		
Xenopus laevis: Juvenile male and female	Nasco	Cat#LM00453MX
Oligonucleotides		
Morpholino antisense oligonucleotides *c-answer* **MO1** 5′-ACAGGCACAACACCGCCTTCATTTT	GeneTools	81178
forward primer WT c-Answer - EcoRI 5′-ATC**GAATTC**TACCAGCGAACAACAAA For **WT c-Answer cloning**	This paper	N/A
reverse primer WT c-Answer - XhoI: 5′- C**CTCGAG**GT**TTA**CATGACCAGGGGCTGTG For **WT c-Answer cloning**	This paper	N/A
forward WT-c-Answer- EcoRI: 5′- ATC**GAATTC**GCCACC**ATG**AAGGCGGTGTTG For **Extracellular c-Answer cloning**	This paper	N/A
reverse primer Extracellular c-Answer – XhoI: 5′-AATT**CTCGAGCTA**ATCTCCGAGACTCAGGC For **Extracellular c-Answer cloning**	This paper	N/A
forward WT-c-Answer- AgeI: 5′-AATC**ACCGGT**CCTGCTGTCAGCCTTATTTTGCTGGCCTGAGTCTCG For **deltaN-c- Answer cloning**	This paper	N/A
reverse primer WT c-Answer - XhoI: 5′- C**CTCGAG**GT**TTA**CATGACCAGGGGCTGTG For **deltaN-c- Answer cloning**	This paper	N/A
forward WT-c-Answer- EcoRI: 5′- ATC**GAATTC**GCCACC**ATG**AAGGCGGTGTTG For **deltaC-c-Answer cloning**	This paper	N/A
reverse primer deltaC c-Answer – XhoI: 5′-GAA**CTCGAGTTA**TTGTGGTTTCTTCATTTGCCG For **deltaC-c-Answer cloning**	This paper	N/A

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pCase12-mem vector	Evrogen	Cat#FP99
SRE-Luc pGL4.33[luc2P/SRE/Hygro]	Promega	Cat#9PIE134
AP-1-Luc pGL4.44[luc2P/AP1 RE/Hygro]	Promega	Cat#9PIE411
vector p-mKate2-mem	Evrogen	Cat#FP186
Software and Algorithms		
Bioinformatics approach for prediction of gene losses and gains between several groups	This paper	http://lab6.iitp.ru/en/lossgainrsl/
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
Tool for selecting target sites for CRISPR/Cas9	Labun et al., 2019	http://chopchop.cbu.uib.no/
Ensembl Compara database	Vilella et al., 2009	https://useast.ensembl.org/info/docs/api/compara/index.html
BLAST	Camacho et al., 2009	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Xenopus time series data	Briggs et al., 2018	https://kleintools.hms.harvard.edu/tools/currentDatasetsList_xenopus_v2.html
Other		
See Table S1 for all recombinant DNAs and oligonucleotides used		N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andrey Zaraisky (azaraisky@yahoo.com). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Manipulations with embryos and tadpoles

For the study *Xenopus laevis* clawed frog was used as a model object. Embryos at stages 10-18 and tadpoles at stages 40-46 of both sexes were used for experiments. Sex in *Xenopus laevis* is not shown to affect the processes we were addressing to in this study. Animal experiments were performed in accordance with guidelines approved by the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry (Moscow, Russia) Animal Committee and handled in accordance with the Animals (Scientific Procedures) Act 1986 and Helsinki Declaration. Embryos were collected and processed for microinjections and whole-mount *in situ* hybridization as described previously (Bayramov et al., 2011). *X. laevis* tadpoles were obtained, amputated and harvested as we described previously (Ivanova et al., 2018). The amputations of *X. laevis* tails were performed using Vannas microscissors after anesthesia with 0,1% tricaine (Sigma-Aldrich).

METHOD DETAILS

Bioinformatic analysis, raw score identification, orthologs definition

We define the raw score for any two genes *i* and *j* as the maximum of the BLAST raw scores for all ordered pairs of proteins: one protein for *i* and the other for *j*. The novel score between given genes *m* and *n* is defined by the formula:

$$\widehat{S}_{mn} = S_{mn} \cdot \sum_{\substack{i \neq m, \\ j \neq n}}' \frac{2S_{ij}}{S_{ii} + S_{jj}} \cdot \exp\left\{ -\frac{[r(i,m) + r(j,n)]}{R} \right\},$$

where R is the normalization parameter, and *r* is the distance between the genes *i* and *m*, *j* and *n*. The summation is carried out over all pairs *i* and *j* of genes in the same contigs (*i* for *m* and *j* for *n*). Moreover, the sum includes not all summands (as indicated by the prime mark). Sorting all summands in descending order gives us a numeric sequence. The first summand in the equation has the index $\{i_0, j_0\}$, now let us remove of members with the indexes i_0 or j_0 from the sequence; then the second summand is the first member of the resulting sequence, etc. The equation allows that $m = n$.

For the definition (*i*) of homologs in section Results, the BLAST program (Camacho et al., 2009) and the new score are used. Specifically, given X gene in the basic species, its homolog X* in another species is chosen from genes, for which the E-value corresponding to the BLAST-made alignment of X and X* is below the specified threshold, so that the raw score is at its maximum (X* is

the best hit for *X*, BH). In addition, Ч can similarly correspond to Ч* using the same score (Ч* and Ч are bidirectional best hits, BBH). Routinely, BH and BBH are used for upper and lower species, respectively; although other variants are permissible. For the new score the choice of *X** is postponed until all raw scores have been calculated, which allows calculating the new scores and then acting as above. The definition (Jun et al., 2009) considers genes as orthologous and paralogous (with no account of synteny) if their proteins belong to the same cluster generated using the new score. Proteins of all considered species are clustered using an original program described in Zverkov et al. (2012) and extensively approved for the raw score elsewhere (Lyubetsky et al., 2013; Rubanov et al., 2016; Zverkov et al., 2012, 2015). The definition (*iii*) is based on the reconciliation of the gene trees against the species tree, which allows to distinguish duplication and speciation events, thus inferring orthologs and paralogs (also with no account of synteny). This approach was used in the Ensembl Compara database (Vilella et al., 2009) from which we obtain the tables of orthologs and paralogs for the available genes. An original distinction of such genes proposed in Lyubetsky et al. (2017) and Rubanov et al. (2016) can be used as an alternative. Several orthology inference methods including the local synteny consideration were compared on five mammalian genomes in Jun et al. (2009).

The new score can take the phylogenetic positions of the considered species into account. Highly conserved elements are important to be considered as witnesses, an algorithm and program for their identification have been proposed in Rubanov et al. (2016). In a wider context, the gene loss in our approach is considered as a combination of considerable change in nucleotide composition and exon-intron structure as well as significant changes in the related synteny and tissue-specific expression. However, these aspects are ignored in the calculations performed in this study. The implementation of main points of the method can be found at http://lab6. iitp.ru/en/lossgainrsl/. The program is deeply parallelized and can operate on a supercomputer, which is essential if a lot of complete genomes are jointly considered or synteny blocks consist of many neighboring genes. The calculations were carried out on an MVS-10P computer at the Joint Supercomputer Center of the Russian Academy of Sciences.

Species used to identify lost genes

The used species are well represented in the Ensembl database (Zerbino et al., 2018). The assembly name is specified for each genome below (with the GenBank accession number in parentheses if available). <u>Fish</u>: agnathous fish – *Petromyzon marinus* (lamprey) Pmarinus_7.0 (GCA_000148955.1); gnathostomatous fish – *Astyanax mexicanus* (cave fish) AstMex102 (GCA_000372685.1), *Danio rerio* (zebrafish) GRCz10 (GCA_000002035.3), *Gasterosteus aculeatus* (stickleback) BROAD S1 (GCA_000180675.1), *Latimeria chalumnae* (coelacanth) LatCha1 (GCA_000225785.1), *Lepisosteus oculatus* (spotted gar) LepOcu1 (GCA_000242695.1), *Oreochromis niloticus* (tilapia) Orenil1.0 (GCA_000188235.1), *Oryzias latipes* (medaka) HdrR, *Poecilia formosa* (amazon molly) Poecilia_formosa-5.1.2 (GCA_000485575.1), *Takifugu rubripes* (fugu) FUGU 4.0 (GCF_000180615.1, FUGU5), *Tetraodon nigroviridis* (tetraodon) TETRAODON 8.0 (GCA_000180735.1), *Xiphophorus maculatus* (platyfish) Xipmac4.4.2 (GCA_000241075.1); <u>amphibians</u>: *Xenopus tropicalis* (clawed frog) JGI_4.2 (GCA_000004195.1); <u>reptiles</u>: *Anolis carolinensis* (anole lizard) AnoCar2.0 (GCA_000090745.1), *Pelodiscus sinensis* (chinese softshell turtle) PelSin_1.0 (GCA_000230535.1); <u>birds</u>: *Gallus gallus* (chicken) Gallus_gallus-5.0 (GCA_000002315.3), *Meleagris gallopavo* (turkey) Turkey_2.01 (GCA_000146605.1), *Taeniopygia guttata* (zebra finch) taeGut3.2.4 (GCA_000151805.1), *Anas platyrhynchos* (duck) BGI_duck_1.0 (GCA_000355885.1), *Ficedula albicollis* (flycatcher) FicAlb_1.4 (GCA_000247815.1); <u>primitive mammals</u> – *Ornithorhynchus anatinus* (platypus) OANA5 (GCF_000002275.2), *Sarcophilus harrisii* (Tasmanian devil) DEVIL7.0 (GCA_000189315.1), *Monodelphis domestica* (opossum) monDom5 (GCF_000002295.2), and <u>placental mammals</u> – *Homo sapiens* (human) GRCh38.p10 (GCA_000001405.25), *Mus musculus* (mouse) GRCm38.p5 (GCA_000001635.7), and *Cavia porcellus* (Guinea pig) cavPor3 (GCF_000151735.1). Our algorithm predicts the same set of lost genes for more recent GenBank sequencing data.

DNA constructs, luciferase assay

Cloning strategies are described in Table S1. Luciferase assay was performed as described (Bayramov et al., 2011). Embryos were injected at 2–4 cell stage with synthetic c-*answer* + *fgf8* mRNAs or solely *fgf8* mRNA (3-4 nL of 100ng/μl mRNA water solution per embryo) mixed with one of the luciferase reporter plasmids: AP-1-Luc pGL4.44[luc2P/AP1 RE/Hygro] (Promega) sensitive to the stress-activated MAPK/JNK pathway, or SRE-Luc pGL4.33[luc2P/SRE/Hygro] (Promega) sensitive to MAP/ERK pathway, and the reference pRenilla plasmid (50 pg/embryo of each reporter plasmid). Animal cap explants were excised from the injected embryos at stage 10, cultured until stage 11, selected in three replicate samples by 10 explants in each and processed for luciferase analysis according to Promega protocol.

Synthetic mRNA and in situ hybridization

Synthetic mRNAs (see Table S1) were prepared with mMessage Machine SP6 Kit (Ambion) after linearization of pCS2-based plasmids with NotI and injected into 2-4 cell stage embryo (3-4 nL of 100ng/μl mRNA water solution per embryo) either into one half of embryo or into the whole embryo or into a particular blastomere in dependence of experiment design. Whole-mount *in situ* hybridization with antisense probes to c-*answer*, *rax*, *pax6*, *six3*, *engrailed1* (Ermakova et al., 2007) was performed as described (Harland, 1991). All mRNAs were mixed with Fluorescein Lysine Dextran (FLD) (Invitrogen, 40 kD, 5 μg/μl) before injections.

Morpholino oligonucleotides and analysis of malformations in regenerating tadpoles

Morpholino antisense oligonucleotides (GeneTools) to *c-answer* (*c-answer* MO) and the control, mismatched, variant of *c-answer* MO (see Table S1) were injected in water solution with MO final concentration 0.25 mM in volume of 3–4 nL per embryo at 2-4-cell stage. All MOs were mixed with Fluorescein Lysine Dextran (FLD) (Invitrogen, 40 kD, 5 μg/μl) before injections.

For regeneration experiments *c-answer* vivoMO was injected into tail stumps of stage 40-42 tadpoles immediately after amputation according to the protocol described earlier (Ivanova et al., 2018). *Mis-ras-dva*-vivoMO (Ivanova et al., 2018) was used as a control to estimate possible effects of the vivoMO as exogenous chemical. Or *wild-type c-answer* mRNA / *extracellular-c-answer* mRNA / *deltaC-c-answer* mRNA / *deltaN-c-answer* mRNA was into embryo into 4 blastomeres on 4-cell stage or sgRNA *c-answer* + Cas9 were injected into not cleaved egg 20' after fertilization. Embryos were incubated until stage 40-42, at which their tails were inspected using fluorescent stereomicroscope Leica M205 and amputated with micro-scissors (Gills-Vannas scissors). On 7-8 dpa tadpoles with both normally and abnormally regenerated tails were counted. Statistical significance was determined with the Student t test and was set p < 0,01. In sum, 248 tadpoles were analyzed in three independent experiments for MO injections. From 100 to 300 tadpoles for *wild-type c-answer* mRNA / *extracellular-c-answer* mRNA / *deltaC-c-answer* mRNA / *deltaN-c-answer* mRNA / sgRNA *c-answer* + Cas9 injections.

qRT-PCR

Sample preparation and qRT-PCR procedure was performed as described (Ivanova et al., 2013; Xanthos et al., 2002). Regenerating after amputation tails of tadpoles were harvested on 0,1,2,6 dpa (0 dpa sample was the piece of stump harvested just after amputation considered as the basal control) and were subjected to qRT-PCR with primers to *c-answer, paramyosin-like*, *sfrpx, foxo1 homolog* and *pnhd* (see Table S1) and two housekeeping genes, *ef1alpha* and *odc*.

In another experiment *c-answer* MO1 or *c-answer* MO2 or *mis-c-answer* MO that served as control were injected into 2-cell embryo. Embryos were incubated until 41 stage and then their tails were amputated. Regenerating after amputation tails were harvested on 0 and 2 dpa and qRT-PCR was performed with primers to *fgf20*, *ag1*, *wnt5a*, *fgf8*, *msx1b*, *ras-dva1* (see Table S1) and two housekeeping genes, *ef1alpha* and *odc*.

Immunoprecipitation and antibodies

Lysates of embryos were prepared as described (Bayramov et al., 2011; Martynova et al., 2008). For immunoprecipitation, aliquots of lysates containing standard amount of tagged protein were mixed with EZview Red ANTI-MycAffinity Gel (Sigma E6654) or EZview Red ANTI-Flag Affinity Gel (Sigma F 2426) and incubated with rotation overnight at4C. After washing 5 times with IPB (immunoprecipitation buffer), protein complexes were eluted with 0.15 mg/ml 3xMYC peptide (Sigma) or 3xFLAG peptide (Sigma) accordingly and analyzed by blotting with monoclonal anti-FLAG Alkaline phosphatase conjugated antibodies or anti-MYC Alkaline phosphatase conjugated antibodies accordingly as described previously (Martynova et al., 2008). In each case, from five to seven replications of experiments performed on different clutches of embryos in different days were done. MYC-/FLAG-epitope-containing proteins were produced in *X. laevis* embryos by using constructs in pCS2 plasmids (see Table S1).

Confocal and fluorescent microscopy, detection of Case12 reporter protein signal

Distribution of EGFP-labeled cells of embryos injected with EGFP-c-Answer mRNA was detected on confocal microscope. All confocal images and FRAP experiments were performed with the confocal microscope "Leica DM IRE 2" using HCX PL APO 63x objective, Ar/Kr laser (488 nm) for excitation of EGFP-tagged proteins. The confocal imaging were obtained at the early-midgastrula stages (stages 10–11) embryos preliminary injected at 4-cell stage with synthetic RNA template (usually 70 pg/blastomere). The *in vivo* fluorescence detection was performed using the fluorescent stereomicroscope Leica M205 and photographed with Leica camera (DC 400F). To test effects of c-Answer on this signaling through P2ry1 receptor, embryos were injected with mRNA of Ca^{2+}-sensitive variant of green fluorescent protein (Case12 reporter protein) mixed with *p2ry1* and *c-answer* mRNAs or solely *p2ry1* mRNA and incubated until stage 11. Animal cap explants were excised from the injected embryos and dissociated on single cells in Calcium Magnesium Free Medium (CMF) (116 mM NaCl, 0.67 mM KCl, 4.6 mM Tris, 0.4 mM EDTA). Ca^{2+} influx upon the addition of P2ry1 agonist ADP (Sigma) to 300 mkM final concentration was recorded via measuring the fluorescence of Case12 reporter on the fluorescent stereomicroscope Leica M205 with Leica camera (DC 400F). The maximal value of signal produced after adding of ADP in each round of measurement was used for further statistical analysis. 40 embryos were analyzed in three independent experiments for each variant of injected mRNA. Statistical significance was determined with the Student t test, p ≤ 0,01.

Cryosectioning

After fixation in 4% PFA the *X. laevis* tail and hind limb bud samples were transferred to the melted warm (+47°C) 1,5% bacto-agar on 5% sucrose solution and were oriented until the sample curdled. The cube with the sample was left in 30% sucrose solution for 12-15 hours and then was bound by the Neg-50 (Richard-Allan Scientific) to the specimen holder and covered by Neg-50. Further, the holder with the sample was carefully inserted into a liquid nitrogen and then cryosectioned (20μ thick) on the Microm HM 525 (Thermo Scientific) and placed on superfrost plus microscope slides (Fisher Scientific, cat.# 12-550-15). Cell nuclei were then stained on these sections with DAPI solution.

CRISPR/Cas9 knockout and genotyping

Target sites within the 2nd and 6th exons of *c-answer* was designed in http://chopchop.cbu.uib.no/software, assuming the score given by this program (see Table S1) (Labun et al., 2019). Besides this, the target sites locations were chose by such a manner to minimize possible ''rescuing'' effect of the in-frame mutations: in the case of 2nd exon near its 5′ end, for the violation of the acceptor site at the 3′ end of the adjacent intron; and in case of 6th exon within the sequence encoding for the transmembrane domain, for to violate incorporation of the mutated c-Answer into the cell membrane. SgRNA template construction and template assembly by PCR was performed as described (Nakayama et al., 2014). *In vitro* transcription of sgRNA was carried out with kit GeneArtTM PlatinumTM Cas9 Nuclease Ready-to-transfect wild-type Cas9 protein for performing CRISPR/Cas9-mediated genome editing Catalog Numbers B25640, B25641 (Invitrogen) according to the guidelines enclosed. Cas9 protein from the kit was mixed with c-*Answer* sgRNA with final concentration 0.8ng and 400pg accordingly in water solution with Fluorescein Lysine Dextran (FLD) (Invitrogen, 40 kD, 5 μg/μl), incubated for 5 min at RT and injected into embryos 20min after fertilization. Embryos were incubated until stage 12-14 and then total DNA was extracted from 10 randomly selected embryos as described (Sive et al., 2010) for method validation. The cDNA of the region of putative mutation was obtained by PCR with direct and reverse specific exterior primers flanking the mutated region of *c-answer*. 50ng of the obtained PCR product was taken for another PCR (total volume 25 μl, 40 cycles) with the interior direct primer (1 μL of 100 pmol/μl), reverse adaptor primer (reverse interior primer with the adaptor part common for all barcode primers) (1 μL of 10 pmol/μl) and bar-code primer unique for each embryo (1 μL of 100 pmol/μl. For all primer sequences see Table S1. The obtained PCR products with unique bar-codes for each embryo were mixed together and genotyped by NGS (Illumina MiSeq).

Single-cell data description

At stage 16 *c-answer* is expressed by 7% (1004 out of 13478) cells in range of 1-25 mRNA molecules per cell. At stage 18 *c-answer* is expressed by 11% (1442 out of 12432) cells in range of 1-17 mRNA molecules per cell. At stage 22 *c-answer* is expressed by 8% (3257 out of 37749) cells in range of 1-20 mRNA molecules per cell. We considered tissues to be enriched *c-answer* if their cells expressing *c-answe*r give more than 15% of all the cells belonging to the given tissue sub-type. All the tissue sub-types for these stages may be found on: https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/cAnswerSt16 for stage 16;

https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/cAnswerSt18 for stage 18;
https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/cAnswerSt22 for stage 22;

in order to obtain the percentage of cells per tissue sub-type expressing *c-answer*, gene name should be chosen as LOC100135223 and all the cells expressing it in the range of 1-max UMI per cell (Slider select 1-max) should be laid out (Layout, Move left). Then the cells expressing *c-answer* should be positive selected and option Celltype should be turned on.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

For experiments on the effect of *c-answer* inhibition and overexpression on regeneration, the injected embryos were incubated until stage 40-42, at which their tails were inspected under fluorescent stereomicroscope Leica M205. Tadpoles with fluorescence were selected and their tails were amputated with micro-scissors (Gills-Vannas scissors). On 7-8 dpa tadpoles with both normally and abnormally regenerated tails were counted. Statistical significance was determined with Student t test and was set p < 0,01. In sum, 248 tadpoles were analyzed in three independent experiments for MO injections, 120 for *deltaN-c-answer* mRNA injections, 130 for *deltaC-c-answer* mRNA. 700 and 600 tadpoles were analyzed for sgRNA *c-answer* + Cas9 protein.

In experiments on the effect of *c-answer* inhibition and overexpression on the forebrain size and expression zone markers we used ImageJ for measurements and then compered condition with the control with the Student t test, p ≤ 0,01 (Schneider et al., 2012).

For qRT-PCR experiments regenerating after amputation tales of tadpoles were harvested on 0,1,2,6 dpa (0 dpa sample was the piece of stump harvested just after amputation considered as the basal control) and were subjected to qRT-PCR with primers to *c-answer, paramyosin-like*, *sfrpx*, *foxo1 homolog* and *pnhd* or *fgf20*, *ag1*, *wnt5a*, *fgf8*, *msx1b*, *ras-dva1* (see Table S1) and two housekeeping genes, *ef1alpha* and *odc*. The experiment had three biological replicates. In each experiment the results were normalized to *ef1alpha* and *odc*. To compare condition with the control we used Student t test and statistical significance was set p < 0,05.

For luciferase assay animal cap explants were excised from the injected embryos at stage 10, cultured until stage 11, selected in three replicate samples by 10 explants in each and processed for luciferase analysis according to Promega protocol. To compare condition with the control we used Student t test, statistical significance was determined p ≤ 0,05.

Ca^{2+} influx upon the addition of P2ry1 agonist ADP (Sigma) was recorded via measuring the fluorescence of Case12 reporter in green channel on the fluorescent stereomicroscope Leica M205 with Leica camera (DC 400F). The maximal value of signal in green channel produced after adding of ADP in each round of measurement was used for further statistical analysis. It was normalized to the signal in red channel that showed the distribution of the injected material and did not change during the experiment. 40 embryos were

analyzed in three independent experiments for each variant of injected mRNA. Statistical significance was determined with the Student t test, p ≤ 0,01.

All data on the figures represent mean ± standard deviation.

DATA AND CODE AVAILABILITY

The bioinformatics approach generated in this paper for prediction of gene losses and gains between several groups of species is freely available at http://lab6.iitp.ru/en/lossgainrsl/.

Links to datasets with single cell data expression on stages 16, 18, 20, 22, where embryonic cells are arranged by their expression profile similarity (Briggs et al., 2018), clustered and colored according to the tissue subtype, where they belong (See Figure S3):

https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/S16
https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/S18
https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/S20
https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/xenopus_embryo_timecourse_v2/S22. In order to obtain the percentage of cells per tissue sub-type expressing *c-answer*, gene name should be chosen as LOC100135223 and all the cells expressing it in the range of 1-max UMI per cell (Slider select 1-max) should be laid out (Layout, Move left). Then the cells expressing *c-answer* should be positive selected and option Celltype should be turned on.