



# A novel regulatory gene promotes novel cell fate by suppressing ancestral fate in the sea anemone *Nematostella vectensis*

Leslie S. Babonis<sup>a,1,2</sup> , Camille Enjolras<sup>a</sup>, Joseph F. Ryan<sup>a,b</sup> , and Mark Q. Martindale<sup>a,b</sup>

Edited by John Pringle, Stanford University School of Medicine, Stanford, CA; received July 25, 2021; accepted March 25, 2022

Cnidocytes (i.e., stinging cells) are an unequivocally novel cell type used by cnidarians (i.e., corals, jellyfish, and their kin) to immobilize prey. Although they are known to share a common evolutionary origin with neurons, the developmental program that promoted the emergence of cnidocyte fate is not known. Using functional genomics in the sea anemone, *Nematostella vectensis*, we show that cnidocytes develop by suppression of neural fate in a subset of neurons expressing RFamide. We further show that a single regulatory gene, a C<sub>2</sub>H<sub>2</sub>-type zinc finger transcription factor (ZNF845), coordinates both the gain of novel (cnidocyte-specific) traits and the inhibition of ancestral (neural) traits during cnidocyte development and that this gene arose by domain shuffling in the stem cnidarian. Thus, we report a mechanism by which a truly novel regulatory gene (ZNF845) promotes the development of a truly novel cell type (cnidocyte) through duplication of an ancestral cell lineage (neuron) and inhibition of its ancestral identity (RFamide).

cell identity | novelty | evolution | cnidarians

Understanding the mechanisms driving cell-type diversification persists as one of the key challenges in evolutionary biology (1). The gain of new adaptive cell functions requires either the advent of novel genes (2, 3), the modification of existing gene regulatory networks (4), or some combination of these two processes (5). Alone, this additive model, focused simply on the emergence of novel gene interactions, is insufficient to explain expansion of cell identity, as new cell types would arise in place of ancestral cell types. In a process analogous to gene duplication and divergence (6), new instances of cell division during embryogenesis could lead to duplication of a cell lineage, providing the opportunity for one cell lineage to retain an ancestral function and the other to acquire new functions (7). Cell-type individuation, therefore, requires both novel gene interactions and a novel cell lineage in which to express these traits (8). Applying a phylogenetic framework like that of gene duplication to the diversification of cell types is a powerful way to make testable predictions about the evolution of development and the emergence of evolutionary novelty (9, 10).

Cnidarians are an unparalleled model for studying the evolution of cellular novelty, because the defining synapomorphy of this group (i.e., the cnidocyte or “stinging cell”) is an unequivocally novel cell type. During embryogenesis, cnidocytes differentiate from a progenitor cell that also gives rise to neurons, reflecting a common evolutionary origin for these two cell types (11, 12). Two key features were necessary for the transition away from neural fate in early cnidocytes: the development of the explosive secretory organelle (the cnidocyst), from which cnidocytes derive their “sting”; and the suppression of neural phenotype (e.g., axons, synaptic signaling molecules). Studies tracking the synthesis of the cnidocyst-specific, structural molecule minicollagen have revealed key steps leading to the origin of this novel organelle (13, 14); the mechanisms driving the suppression of neural phenotype in cnidocytes remain unknown. Here, we show that a single transcription factor—a six-domain C<sub>2</sub>H<sub>2</sub> zinc finger transcription factor—both promotes cnidocyte fate and suppresses neural fate during development of the sea anemone *Nematostella vectensis*. We further show that the six-domain topology of this transcription factor arose through domain shuffling in the last common ancestor of cnidarians, making this a clear example of a cnidarian-specific gene driving development of a cnidarian-specific trait.

## Results

**ZNF845 Promotes Cnidocyte Fate.** A previous study of stem cell dynamics in *Hydra* identified a six-domain C<sub>2</sub>H<sub>2</sub> zinc finger transcription factor (called *ZNF845*) as a general marker of the multipotent interstitial stem cell lineage (15). To determine if this gene

## Significance

In this study, we demonstrate how a new cell type can arise through duplication of an ancestral cell type followed by functional divergence of the new daughter cell. Specifically, we show that stinging cells in a cnidarian (namely, a sea anemone) emerged by duplication of an ancestral neuron followed by inhibition of the RFamide neuropeptide it once secreted. This finding is evidence that stinging cells evolved from a specific subtype of neurons and suggests other neuronal subtypes may have been coopted for other novel secretory functions.

Author affiliations: <sup>a</sup>Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL 32080; and <sup>b</sup>Department of Biology, University of Florida, Gainesville, FL 32611

Author contributions: L.S.B. designed research; L.S.B., C.E., J.F.R., and M.Q.M. performed research; L.S.B. contributed new reagents/analytic tools; L.S.B., C.E., J.F.R., and M.Q.M. analyzed data; and L.S.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>Present address: Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853.

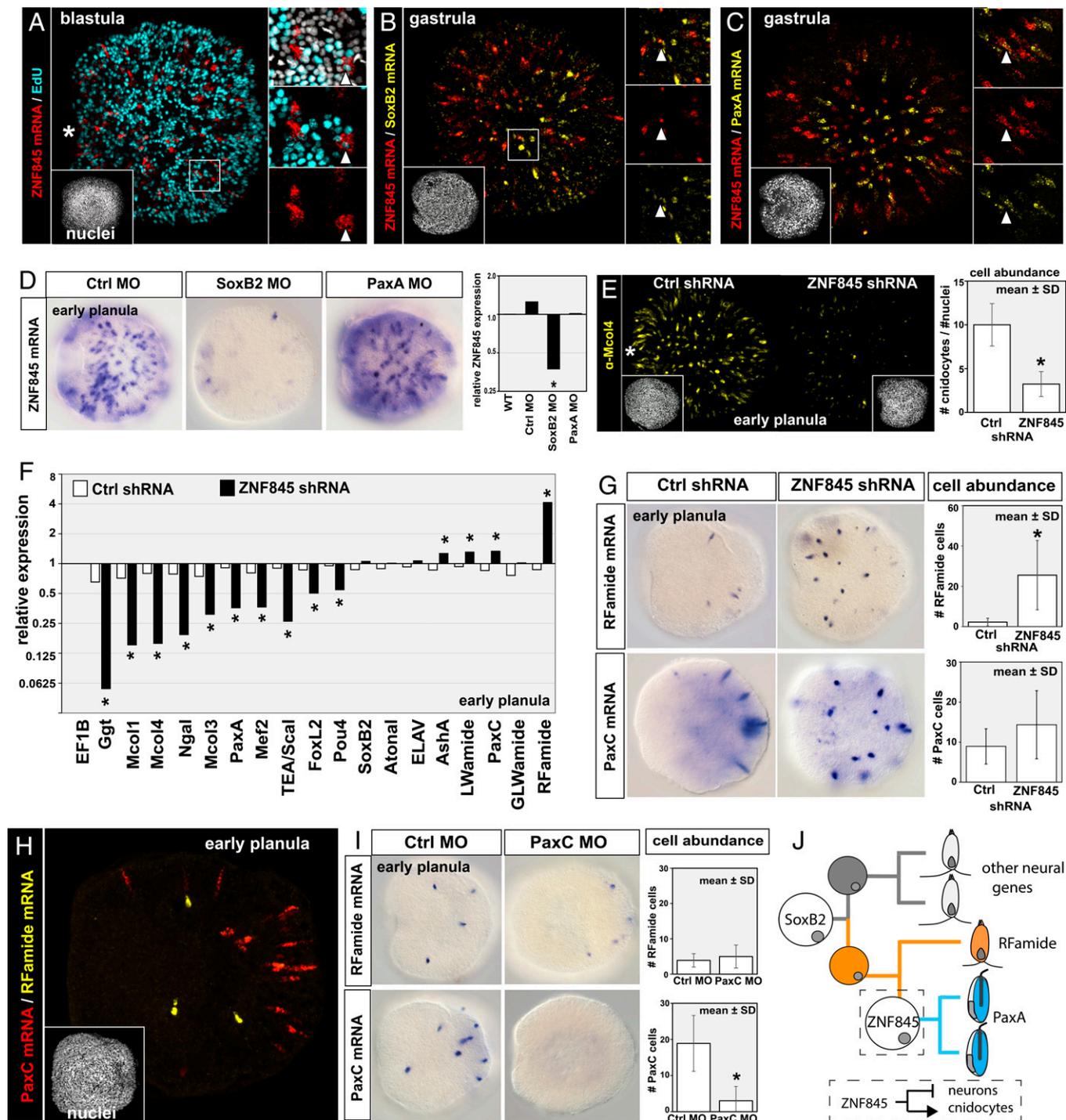
<sup>2</sup>To whom correspondence may be addressed. Email: lsb257@cornell.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113701119/-DCSupplemental>.

Published May 2, 2022.

played a role in marking multipotent stem cells in other cnidarians, we examined the spatial expression and function of the *ZNF845* ortholog in the sea anemone *N. vectensis*. Consistent with a stem cell function, *ZNF845* was found to be expressed throughout the early embryo in cells that are actively undergoing DNA synthesis, as labeled by EdU (Fig. 1A). However, further investigation revealed that *ZNF845* was partially coexpressed with *SoxB2*, a marker of embryonic neural progenitor cells (12)

(Fig. 1B) and partially coexpressed with *PaxA*, an early marker of differentiating cnidocytes (16) (Fig. 1C). Moreover, *ZNF845* continues to be expressed through metamorphosis in *N. vectensis* in a pattern reminiscent of cnidocyte development (*SI Appendix, Fig. S1*). To understand the role of *ZNF845* in cnidocyte differentiation more fully, we knocked down *SoxB2* and *PaxA* using previously published morpholinos (MOs) (12, 16). *ZNF845* expression was down-regulated in embryos injected with the



**Fig. 1.** *ZNF845* specifies cnidocyte identity in *N. vectensis*. (A–C) *ZNF845* is partially coexpressed with: (A) EdU (proliferating cells), (B) *SoxB2* mRNA (neural progenitor cells), and (C) *PaxA* mRNA (cnidocytes). Insets: white arrowheads show coexpression; nuclei are white (DAPI stain). (D) *ZNF845* expression after knockdown of *SoxB2* and *PaxA* (by MO) assayed by in situ hybridization and qPCR. \* $P < 1E-02$  for *SoxB2* MO or *PaxA* MO versus Ctrl MO. (E) Cnidocyte differentiation ( $\alpha$ -Mcol4 antibody) after *ZNF845* knockdown (shRNA). (F) qPCR of target gene expression after *ZNF845* knockdown; fold-change relative to housekeeping gene *EF1B*. (G) Response of *RFamide*- and *PaxC*-expressing cells to *ZNF845* knockdown. (H) *RFamide* and *PaxC* are not coexpressed. (I) Expression of *RFamide* and *PaxC* after knockdown of *PaxC* (MO). (J) Model for *ZNF845*-mediated specification of cnidocyte identity from an *RFamide*-expressing ancestral cell lineage. \* $P < 1E-02$  (*SI Appendix, Table S2*).

*SoxB2* MO but unaffected in embryos injected with *PaxA* MO (Fig. 1D). Paired with the coexpression results, these data suggest that *ZNF845* is part of the cnidogenesis pathway acting upstream of *PaxA*. In support of this hypothesis, a recent study using single-cell RNA sequencing (RNA-seq) in the hydrozoan *Clytia hemisphaerica* revealed *ZNF845* was expressed specifically in the subpopulation of interstitial stem cells that goes on to become cnidocytes (17). Together, these results suggest that *ZNF845* plays a similar role in the early patterning of cnidocytes in two distantly related cnidarians and may not be a general marker of stem cell fate, as once thought.

To further explore the role of *ZNF845*, we knocked down *ZNF845* using short hairpin RNAs (shRNAs) (18, 19) and assayed the effects on cnidocyte development in *N. vectensis*. Knockdown of *ZNF845* was effective up to the late planula stage (SI Appendix, Fig. S1) and resulted in nearly complete loss of cnidocytes throughout the ectoderm of the planula larva (Fig. 1E). Using an antibody directed against the cnidocyte-specific molecule Minicollagen-4 ( $\alpha$ -Mcol4) (20), we demonstrate a significant loss of cnidocytes: from 10% of total cells in wild-type (WT) and control shRNA-injected embryos (mean  $\pm$  SD: WT,  $10.6 \pm 2.01$ ; control,  $10.0 \pm 2.42$ ) to only 3.2% ( $\pm 1.43$ ; mean  $\pm$  SD) when *ZNF845* was knocked down (SI Appendix, Table S2). We recovered identical results in embryos injected with a splice-blocking *ZNF845* MO, relative to those injected with a standard control MO (SI Appendix, Fig. S2 and Table S2).

To determine if *ZNF845* acts upstream of *PaxA* and other genes specific to the cnidogenesis pathway, we used qPCR to examine the effect of *ZNF845* knockdown on known markers of neural and cnidocyte differentiation in *N. vectensis* (Fig. 1F and SI Appendix, Table S2). Knockdown of *ZNF845* resulted in significant down-regulation of *PaxA* and all of its known targets (*GGT*, *Ngal*, *Mcol1*, *Mcol3*, *Mcol4*, and *Mef2IV*) (16), as well as three cnidocyte-expressed transcription factors identified from previous analyses (*TEA/Scalloped*, *FoxL2*, and *Pou4*) (21, 22). Conversely, *ZNF845* knockdown did not affect the expression of *SoxB2* or the neuron-specific regulatory genes *Atonal* and *ELAV* (23). While *ZNF845* knockdown caused a statistically significant increase in the expression of neural markers *AshA*, *LWamide*, and *PaxC*, assayed by qPCR, the response of these genes was minor relative to the large, significant up-regulation of *RFamide*. To spatially characterize these results, we performed in situ hybridization for *RFamide* and *PaxC* in embryos injected with control or *ZNF845* shRNAs and counted the number of cells expressing these neural markers at the early planula stage (Fig. 1G). Knockdown of *ZNF845* significantly increased the number of *RFamide*-expressing cells from  $2.23 (\pm 1.95)$ ; mean  $\pm$  SD) in control embryos to  $23.47 (\pm 17.85)$ ; mean  $\pm$  SD) in *ZNF845* knockdowns. Cell counts for *PaxC* expression revealed a small, nonsignificant increase in the number of *PaxC*-expressing cells from  $8.9 (\pm 4.4)$ ; mean  $\pm$  SD) in controls to  $14.3 (\pm 18.5)$ ; mean  $\pm$  SD) in *ZNF845* knockdowns. Because *PaxC* expression during embryogenesis is reminiscent of *RFamide* expression, we hypothesized that *PaxC* might be an upstream regulator of *RFamide* neuron differentiation. On the contrary, we found that *PaxC* and *RFamide* are not coexpressed (Fig. 1H) and that knockdown of *PaxC* using a previously published MO (16) did not affect the number or distribution of *RFamide*-expressing cells (Fig. 1J).

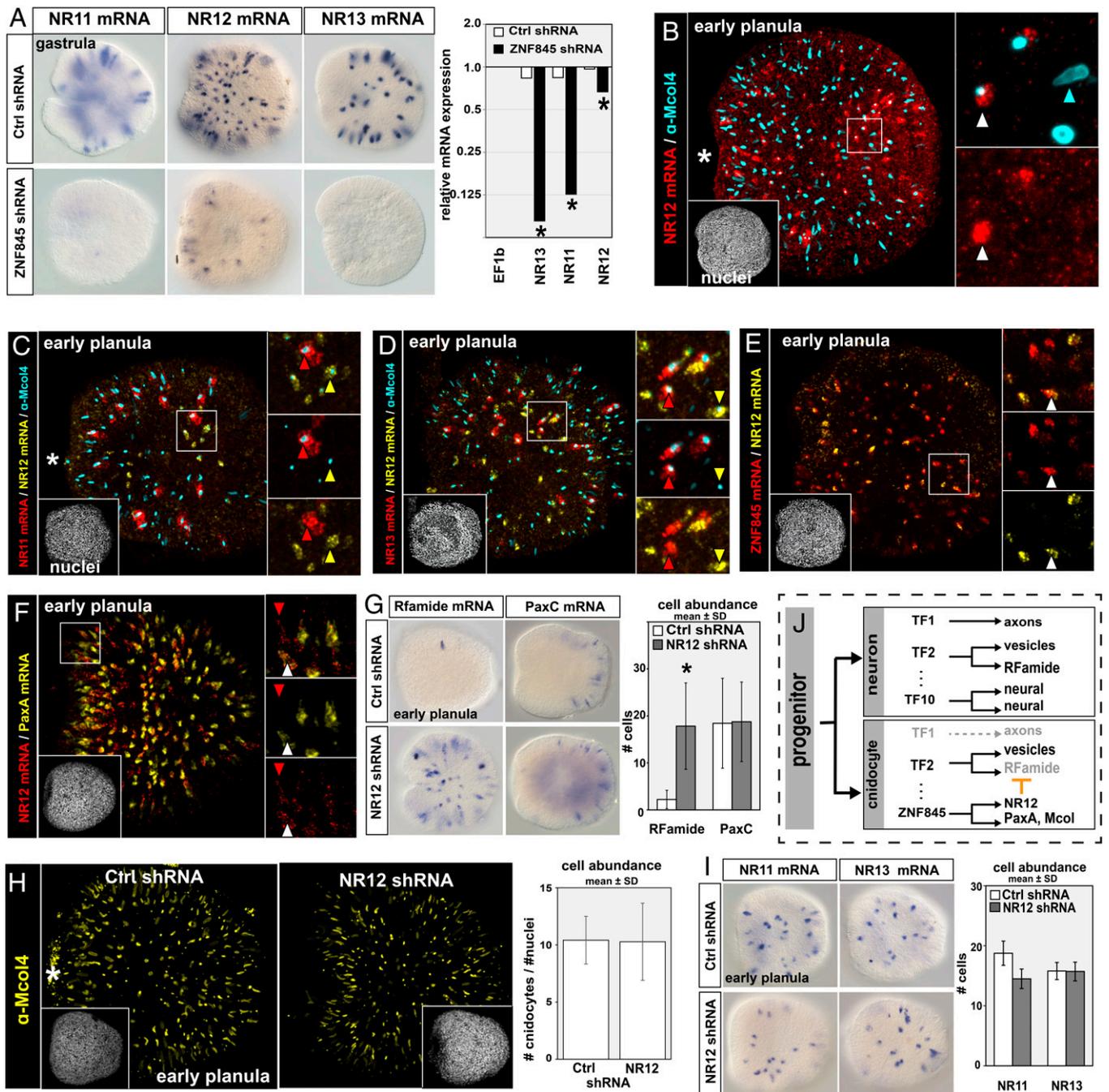
These results suggest that there is a closer evolutionary relationship between cnidocytes and *RFamide*-expressing neurons than between cnidocytes and other neural subtypes (Fig. 1J). This is congruent with results from a recent study showing that the selector gene *Pou4* regulates terminal cell identity in

cnidocytes and *RFamide*-expressing neurons in *N. vectensis* but not in other cell types (22). Further, we demonstrate that a single transcription factor (*ZNF845*) up-regulates both the genes necessary to promote cnidocyte identity and the genes necessary to inhibit *RFamide* neuron identity. To understand the mechanism by which *ZNF845* suppresses *RFamide* expression during cnidocyte differentiation, we searched for inhibitory transcription factors that were coexpressed with cnidocyte-specific genes.

**ZNF845 Inhibits Neural Fate through NR12.** Nuclear receptors in the COUP-TF family (NR2F) are known to play an inhibitory role in neural cell-fate decisions in both cnidarians (24) and bilaterians (25). Given that an NR2F ortholog has been shown to be expressed during cnidocyte differentiation in a hydrozoan cnidarian (26), we examined the NR2F genes for a potential role in early cnidogenesis. In *N. vectensis*, there are five NR2F paralogs: NR10 through 14, most of which appear to have originated through lineage-specific duplication in cnidarians (27). We examined the expression of all five NR2F genes (SI Appendix, Fig. S3) and found that three of them (NR11, NR12, and NR13) were expressed in the ectoderm during early embryogenesis and were also down-regulated in embryos injected with *ZNF845* shRNA (Fig. 2A). All three were also coexpressed with *Mcol4* in differentiating cnidocytes (Fig. 2B–D), and yet we found no evidence of that any of the NR2F paralogs were coexpressed together in the same cell (Fig. 2C and D), suggesting these NR2F paralogs are expressed uniquely in different subpopulations of cnidocytes.

The cnidocyte-specific expression of *NR12* that we observed was also supported by a previous study using single-cell RNA-seq in *N. vectensis* (21), which revealed the expression of *NR12* (“*Oasis1*”) in cell clusters that appear to be differentiating cnidocytes. We examined *NR12* further and found that it is coexpressed in a subset of *ZNF845*-expressing cells (Fig. 2E) and in a subset of *PaxA*-expressing cells at the early planula stage (Fig. 2F). We examined the effects of *NR12* on cnidocyte development using shRNAs (SI Appendix, Fig. S4 and Table S1) and found that knockdown of *NR12* resulted in a four-fold increase in the number of *RFamide*-expressing cells but had no effect on the expression of *PaxC* (Fig. 2G) or on the specification of cnidocytes (Fig. 2H). Knockdown of *NR12* also had no effect on the number or distribution of *NR11*- or *NR13*-expressing cells (Fig. 2I).

Modularity in the regulation of neural phenotype would have enabled cnidocytes to retain beneficial aspects of the ancestral phenotype while silencing others through selective inhibition, as has been shown for neural subtype specification in *Caenorhabditis elegans* (25). The up-regulation of *RFamide* following *NR12* knockdown suggests that *RFamide* expression is actively suppressed in cnidocytes. One explanation for this observation is that it could reflect the coupling of *RFamide* expression to the expression of another trait that had adaptive value during the evolution of cnidocytes (e.g., secretory vesicles) (Fig. 2J). Up-regulation of this gene regulatory module would enable early evolving cnidocytes to develop a secretory vesicle, using the module already in place in the ancestral neuron; subsequent inhibition of the ancestral payload (*RFamide*) would provide the opportunity for a novel secretory phenotype to emerge. Thus, active inhibition of *RFamide* in cnidocytes may be interpreted simply as a vestige of the shared ancestry of neurons and cnidocytes (28). The independent expression of *NR11* and *NR13* in nonoverlapping populations of developing cnidocytes suggests that multiple cnidocyte subtypes may be specified through NR2F-mediated inhibition of neurosecretory products in other (non-*RFamide*) neural subtypes.



**Fig. 2.** RFamide identity is inhibited by *NR12* in developing cnidocytes. (A) *NR11*, *NR12*, and *NR13* are down-regulated in *ZNF845* knockdowns (shRNA) as assayed by in situ hybridization and qPCR. (B–D) *NR11*, *NR12*, and *NR13* are coexpressed with *Mcol4* (cnidocytes; cyan), but not with each other (C and D). (E and F) *NR12* is partially coexpressed with (E) *ZNF845* and (F) *PaxA*. (G) Response of *Rfamide*- and *PaxC*-expressing cells after *NR12* knockdown. (H and I) There is no effect of *NR12* knockdown on (H) cnidocyte differentiation (*Mcol4*) or (I) expression of *NR11* or *NR13*. (J) Modular regulation of neural traits (e.g., axons, vesicles plus neuropeptides) explains how some traits can be lost (e.g., axons) and others retained but inhibited (RFamide) in developing cnidocytes. \* $P < 1E-02$ . See *SI Appendix, Table S2* for detailed information. TF, transcription factor.

**ZNF845 Is a Novel Transcription Factor.** Transcription factors with Cys2-His2 zinc finger (ZF-C<sub>2</sub>H<sub>2</sub>) domains (PF00096) represent one of the largest families of transcription factors across animals (29). Numerous evolutionary processes have contributed to diversification in this gene family, including duplication and divergence, gain and loss of ZF-C<sub>2</sub>H<sub>2</sub> domains, and gain and loss of accessory domains (30). Among the oldest members of this clade are proteins with multiple, tandem ZF-C<sub>2</sub>H<sub>2</sub> domains but no other conserved functional domains (31). The *N. vectensis* ortholog of *ZNF845* has six tandem ZF-C<sub>2</sub>H<sub>2</sub> domains. To understand how these six domains came together in a single

protein, we generated a maximum likelihood phylogeny of all ZF-C<sub>2</sub>H<sub>2</sub> domains extracted from predicted proteomes for three bilaterian taxa (*Homo sapiens*, *Drosophila melanogaster*, and *C. elegans*) and four cnidarians (two anthozoans: *N. vectensis* and *Acropora digitifera*; and two medusozoans: *Hydra magnipapillata* and *Nemopilema nomurai*). The full phylogeny contains more than 11,000 branch tips; the FASTA alignment is provided (Dataset S1). To support this initial investigation, an additional analysis was performed using the same approach with all ZF-C<sub>2</sub>H<sub>2</sub> domains ( $N > 5,000$ ) from *N. vectensis* and two representatives of the Lophotrochozoa: *Capitella teleta* and *Lottia*

*gigantea*; the FASTA alignment for this analysis, containing more than 5,000 domains, is provided (Dataset S2).

*N. vectensis* has 218 predicted proteins with one or more ZF-C<sub>2</sub>H<sub>2</sub> domains, 16 of which, including ZNF845, encode six tandem ZF-C<sub>2</sub>H<sub>2</sub> domains. We examined the evolutionary relationships of the ZF-C<sub>2</sub>H<sub>2</sub> domains from ZNF845 (Joint Genome Institute [JGI] Protein Identification [ID]: 81344) and compared the results with an analysis of another six-domain ZNF protein, an ortholog of growth factor independence 1B (GFI1B) (JGI Protein ID: 112378). GFI1B is a potent regulator of cell differentiation in vertebrates (32), but the function of this protein has not been characterized in *N. vectensis*. Each of the six domains from the *N. vectensis* ortholog of ZNF845 groups with ZF-C<sub>2</sub>H<sub>2</sub> domains from the orthologous ZNF845 protein in other cnidarians: Hmag\_XP\_002158383.2 [originally named ZNF845 by Hemmrich et al. (15)], Nnom\_10870, and Nnom\_3755 (Fig. 3A and SI Appendix, Fig. S5). These relationships suggest the anthozoan and medusozoan ZNF845 proteins descended from a common ancestor that already had six ZF-C<sub>2</sub>H<sub>2</sub> domains in the stem cnidarian. Additionally, some evidence suggests that two of the

domains from ZNF845 in cnidarians (domains 1 and 5) may have arisen by tandem duplication, as these domains form a clade that lacks other ZF-C<sub>2</sub>H<sub>2</sub> domains in the initial tree; however, these two domains were found in more distantly related lineages in the analysis that included lophotrochozoans. In both analyses, all six of the ZF-C<sub>2</sub>H<sub>2</sub> domains in ZNF845 were found to share some level of homology with a bilaterian ZF-C<sub>2</sub>H<sub>2</sub> domain, but none of these bilaterian domains are found in the same protein (SI Appendix, Figs. S5 and S7).

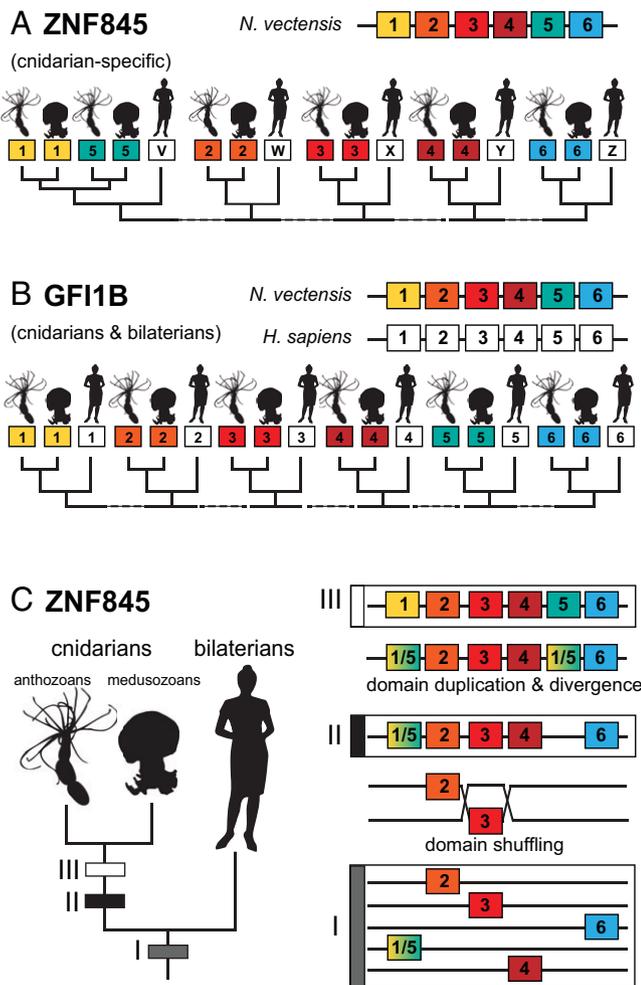
Examination of the ZF-C<sub>2</sub>H<sub>2</sub> domains from GFI1B suggests that, unlike ZNF845, this six-domain transcription factor emerged in its current form before cnidarians and bilaterians diverged from their common ancestor (Fig. 3B and SI Appendix, Figs. S6 and S8). Each of the ZF-C<sub>2</sub>H<sub>2</sub> domains from the *N. vectensis* ortholog of GFI1B grouped with the syntenic ZF-C<sub>2</sub>H<sub>2</sub> domain from the GFI1B ortholog in each of the bilaterian taxa examined (Hsap\_NP\_001120687.1; Hsap\_XP\_006717360.1; Dmel\_Q9VM77; Cele\_5376; Ctel\_45287; Lgig\_83709; and Lgig\_129344). Together, these observations suggest that ZNF845, but not GFI1B, arose as a novel six-domain protein in the stem cnidarian and that both domain shuffling and domain duplication and divergence were important for the emergence of this protein (Fig. 3C).

## Discussion

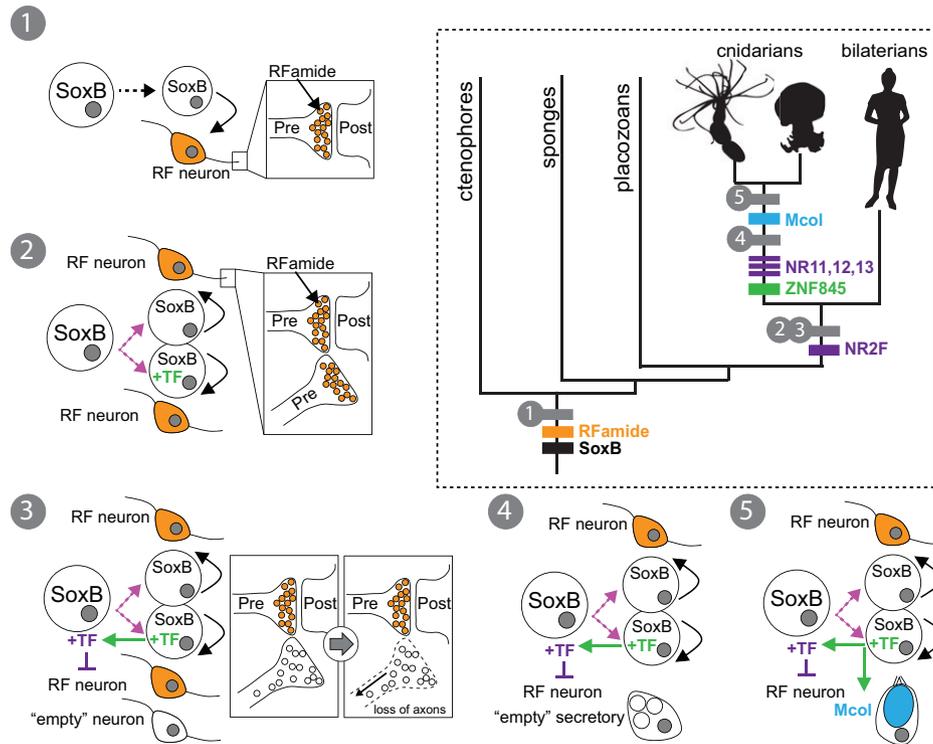
**Renaming Cnidarian ZNF845 to CnZNF1.** The cnidarian gene named *ZNF845* was originally identified in a screen of stem cell-specific genes in *Hydra* (15) and was likely named as such because it had significant Basic Local Alignment Search Tool (BLAST) similarity to the *ZNF845* gene found in humans (National Center for Biotechnology Information [NCBI] accession NP\_001308451). Here, we show explicitly that no bilaterian ZNF gene is the reciprocal best BLAST hit for cnidarian *ZNF845* (NCBI accession for *N. vectensis*: XP\_001641179.2). As such, we suggest changing the name of the cnidarian gene formerly called *ZNF845* (15, 17, 33–36) to *CnZNF1*. Renaming the gene this way accomplishes two things: first, the prefix “Cn” denotes that this is a cnidarian-specific transcription factor; second, this naming scheme reduces any confusion that derives from the fact that the name *ZNF845* is already associated with a human gene that has a distinct evolutionary history.

**Modeling Cell-Type Expansion.** That sister cell types should express a similar molecular “fingerprint” is not a new idea (10); indeed, this concept has inspired the development of numerous cell-type phylogenies to explain the functional diversification of multicellular organisms (37–40). Many of these studies take advantage of widely accessible observational techniques (e.g., RNA-seq) in distantly related taxa to reconstruct ancestral cell states and provide a roadmap for cell-type evolution through duplication and segregation of ancestral functions. Just as gene duplication and divergence are important sources for novel gene function (6, 41), this concept has been expanded to provide a framework for understanding the emergence of novel cell types through cell lineage duplication and divergence (7). Here, we demonstrate explicitly how gene expression became segregated between sister cell types to drive the emergence of cnidocytes from their common origin with neurons in *N. vectensis* (Fig. 4).

At least one *SoxB* gene was present in a neural progenitor cell in the ancestor of all animals prior to the duplication event that promoted the emergence of cnidocytes (42) (Fig. 4, step 1). Analyses of animal neural diversity have suggested *RFamide*-like



**Fig. 3.** Evolutionary history of two six-domain C<sub>2</sub>H<sub>2</sub> ZNF proteins. (A and B) Relationships among ZF-C<sub>2</sub>H<sub>2</sub> domains from (A) ZNF845 (*N. vectensis*; JGI Protein Identification [ID]: 81344) and (B) GFI1B (*N. vectensis* JGI Protein ID: 112378). Broken gray lines represent the locations of the other ~11,000 branch tips of the complete tree. (C) Proposed model for the emergence of ZNF845: 1. In the common ancestor of cnidarians and bilaterians, ZF-C<sub>2</sub>H<sub>2</sub> domains were in distinct proteins. 2. In the stem cnidarian, domain shuffling brought domains 1/5, 2, 3, 4 and 6 together in a single protein. 3. Domain 1/5 then duplicated and diverged to become two distinct domains (1 and 5) before the diversification of extant cnidarians (SI Appendix, Figs. S5–S8).



**Fig. 4.** Origin of novelty through lineage duplication and suppression of ancestral fate. A stepwise model for the origin of cnidocytes by duplication and divergence of an *RFamide* neural lineage and a proposal for the evolutionary timing of these events are presented.

neurons arose early in animal evolution as well (43, 44). To generate a novel cell lineage, this *SoxB*-expressing progenitor cell would first have to duplicate to generate a new progenitor lineage (Fig. 4, step 2, magenta arrows). Absent additional changes, duplication of a neural progenitor cell would have doubled the number of *RFamide*-expressing daughters, potentially leading to aberrant synapse formation (Fig. 4, step 2, inset). Individuation would have required the origin of a transcription factor (Fig. 4, step 3, green TF) that promoted independent regulation of the duplicated daughter cells (38) and regulatory independence would have enabled the establishment of a method for inhibition of *RFamide* expression in one daughter only (Fig. 4, step 2, purple TF). Thus, the original number of *RFamide*-expressing daughters would be maintained and *RFamide* neurons would now be sister to a second lineage of “neurons” lacking a synaptic payload (Fig. 4, step 3). Key to the emergence of cnidocytes, therefore, was the origin of two transcription factors: one (*ZNF845*) that could maintain progenitor cell fate in a newly duplicated cell lineage and another (*NR12*) that could manipulate the identity of the secreted payload in the duplicated daughter cell. Without the selection pressure to maintain synaptic signaling in the new daughter cell, the novel cell lineage may have experienced relaxed selection for the maintenance of axons, allowing the secretory vesicles to relocate to the cell body (Fig. 4, step 3, insets). Subsequent mutations resulting in the centralization and specialization of the secretory vesicle (Fig. 4, step 4) and the emergence of a novel payload (e.g., Minicollagen) would have further promoted divergence of this novel cell type (Fig. 4, step 5). Mapping these traits onto a phylogeny of animals suggests that *NR2F*-mediated payload inhibition may also drive secretory cell diversification in bilaterian taxa.

The stepwise model presented here illustrates three important paradigms for the evolution of novel cell types. First, although the role of novel genes in driving evolutionary innovation has been debated (45–47), cnidocytes have always

provided a clear example of an adaptive role for novel effector genes (e.g., Minicollagen) in driving the evolution of a new morphological character state (13). In the present study, we extended this adaptive role for novel genes up the cnidocyte-gene regulatory network by showing that a new transcription factor assembled through domain shuffling in the stem cnidarian (*ZNF845*) is essential for the development of cnidocytes in *N. vectensis*. Thus, we present a complete cnidocyte character identity network which includes the genes necessary for activation of new character traits and suppression of ancestral traits (48), and provide a mechanistic explanation for the development of a truly novel cell type. Given the similar expression of *ZNF845* in hydrozoans, in the progenitor cell population that gives rise to cnidocytes (17), we suggest the results presented here may reflect a broader phenomenon—that this cnidarian-specific gene (*ZNF845*) was essential for the emergence of a cnidarian-specific cell type (cnidocytes) in the ancestor of all cnidarians.

Second, we demonstrate support for the hypothesis that modularity is an important driver of phenotypic evolution (49, 50). During the early divergence of cnidocytes, modularity in neural cell phenotype could have allowed for selective inhibition of certain traits (e.g., vesicular payload, axons) and retention of others (e.g., secretory vesicle). An independent regulatory environment would have enabled the inhibition of *RFamide* expression in only one of the recently duplicated sister cells, promoting retention of ancestral phenotype in the other. In this example, *ZNF845* provided for regulatory independence between sister cell types, explicitly supporting the cell-type individuation model proposed previously (1, 38).

Finally, our results broadly support the hypothesis that the origin of a secretory-cell lineage was advantageous for the emergence of diverse, novel cell functions (51, 52). The ability to segregate gene products into a compartment within the cell and to target those products for delivery to the extracellular space

allows for the retention of novel gene products that may otherwise be deleterious if retained intracellularly (e.g., collagen fibers).

Applied broadly, this scenario of secretory-cell duplication coupled to payload inhibition combined with the early diversification of novel neuropeptides (53–55) could also explain the rapid expansion of neural function in early bilaterians. Studies of neural fate in bilaterian model systems have demonstrated that discrimination of specific subtypes within a neural lineage relies on inhibition of the effector genes that define other subtypes in the lineage (56) and that this inhibition of sister cell fate is mediated through the actions of conserved transcription factors, including orthologs of NR2F/COUP-TF (25). The data presented here indicate that this role for NR2F-mediated inhibition of cell fate may extend back to the common ancestor of cnidarians and bilaterians, suggesting the divergence-through-inhibition regulatory logic was already driving the expansion of cell fate nearly 700 Mya (57).

## Materials and Methods

**Gene Knockdowns.** To assess the influence of *ZNF845* and *NR12* on cnidocyte fate, we performed messenger RNA (mRNA) knockdown by microinjection of shRNAs following the protocol of He et al. (18). shRNAs used in this study were synthesized in vitro (primer sequences are provided in the *SI Appendix, Table S1*) and diluted to 800 ng/μL in nuclease-free water (Ambion AM9937) with a final concentration of 0.2 mg/mL Alexa-555 RNase-free dextran (Invitrogen D34679) to facilitate injection. To account for nonspecific effects, control embryos were injected with an shRNA that was not complementary to any part of the genome (*SI Appendix, Table S1*; 800 ng/μL) (19). Embryos were raised to the early planula stage (72 h postfertilization at 16 °C) and effects of knockdowns were assayed via immunofluorescence, in situ hybridization, or qPCR. Independent confirmation of the effect of *ZNF845* knockdown on cnidocyte specification was assayed using a splice-blocking MO (GeneTools; *SI Appendix, Table S1*) as previously described (16). Briefly, lyophilized MOs were reconstituted in nuclease-free water to 1 mM, following the manufacturer's instructions. Before each use, the stock was heated to 60 °C for 5 min and centrifuged for 1 min before being diluted to a final working concentration of 0.3 mM in nuclease-free water with 0.2 mg/mL RNase-free dextran. To control for nonspecific effects, *ZNF845* MO-injected embryos were compared with embryos injected with a standard control MO (*SI Appendix, Table S1*) prepared the same way and injected at the same concentration as the *ZNF845* MO. Splicing defects were confirmed using PCR and gel electrophoresis as described previously (16).

**Cell and Tissue Analysis.** To assay effects of gene knockdown on cnidocyte specification, developing cnidocytes were labeled, imaged, and counted as described in a similar study (16), using an antibody directed against Minicollagen-4 (α-Mcol4) (13). Additional targets of *ZNF845* and *NR12* shRNAs were assessed using in situ hybridization as previously described (16). Embryonic stages used for in situ hybridization, cell counts, and qPCR analysis are

noted throughout the article. All stages are associated with age (in hours postfertilization [hpf]) in *SI Appendix, Fig. S1*; briefly, blastula: 24 hpf; gastrula: 48 hpf; early planula: 72 hpf; late planula, 120 hpf; tentacle bud, 240 hpf; and primary polyp, 2 wk. For qPCR analysis, embryos injected with control shRNA and embryos injected with *ZNF845* shRNA were both compared with uninjected embryos raised under the same conditions to 72 hpf. Approximately 300 embryos were collected for each replicate of each condition, homogenized in Tri-Reagent (Sigma T9424), and stored at –80C for further processing. All samples were processed at the same time for RNA extraction and complementary DNA synthesis, as previously described (16). Five replicates of each condition (*ZNF845* shRNA, control shRNA, and uninjected) representing five independent injections performed on different days were compared using the comparative cycle threshold method ( $\Delta\Delta$  CT) and the PCR package in the R statistical computing environment (58, 59). *ZNF845* expression in Fig. 1D is presented as fold-change, relative to *ZNF845* expression in uninjected embryos, and statistical significance for qPCR in Fig. 1D was calculated from the comparison of *SoxB2* MO- and *PaxA* MO-injected embryos relative to control MO-injected embryos. mRNA expression values in Figs. 1F and 2A are presented as fold-change relative to expression of *EF1B* in the uninjected embryos (arbitrarily set to 1) and significance was calculated from the comparison of *ZNF845* shRNA-injected embryos to control shRNA-injected embryos. Cell counts in Figs. 1E, G, and I and 2G, H, and I were analyzed with a Mann-Whitney *U* nonparametric test for two-way comparisons (control vs. target shRNA/MO) and are presented as mean  $\pm$  SD. Statistical significance for all quantitative comparisons is indicated where  $P < 1E-02$ .

**Maximum-Likelihood Phylogeny.** Phylogenetic analysis of ZF-C<sub>2</sub>H<sub>2</sub> domains was performed using a modification of a previously published protocol (60). In brief, an alignment was generated using a custom script (hmm2aln, available at <https://github.com/josephryan>) and the ZF-C<sub>2</sub>H<sub>2</sub> HMM (PF00096) in predicted proteomes from all target taxa. For cnidarians, we sampled two anthozoans: *N. vectensis* and *A. digitifera*, and two medusozoans: *H. magnipapillata* and *Nemopilema nomurai* (61). For bilaterians, we sampled *Homo sapiens* (Hsap), *D. melanogaster* (Dmel), and *C. elegans* (Cele) initially and followed up with additional analysis including two lophotrochozoans: *Capitella teleta* (Ctel) and *L. gigantea* (Lgig) (62). Download information for each taxon is provided in the *SI Appendix, Table S3*. The first alignment contains more than 11,000 ZF-C<sub>2</sub>H<sub>2</sub> domains from these combined taxa and is provided in *Dataset S1*; the alignment for *N. vectensis* and the lophotrochozoans is provided in *Dataset S2*. To generate the phylogeny, we first used the model-finder function (-MF) with IQTREE to determine the best substitution model (VT+R8) and then generated a single tree and applied 500 bootstraps using fast bootstrapping.

**Data Availability.** All study data are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** We are grateful to Christine Schnitzler, Danielle deJong, Namrata Ajuha, and Malcolm Moses for their research assistance. This work was funded by NASA (grant NNX14AG70G to M.Q.M.) and the NSF (grant 1542597 to J.F.R.).

1. D. Arendt et al., The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
2. J. Zhang, Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18**, 292–298 (2003).
3. A. B. Van Oss, A.-R. Carvunis, De novo gene birth. *PLoS Genet.* **15**, e1008160 (2019).
4. D. H. Erwin, E. H. Davidson, The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* **10**, 141–148 (2009).
5. G. P. Wagner, V. J. Lynch, The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol.* **23**, 377–385 (2008).
6. A. Force et al., Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
7. D. Arendt, The evolution of cell types in animals: Emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
8. G. P. Wagner, *Homology, Genes, and Evolutionary Innovation* (Princeton University Press, Princeton, NJ, 2014). <https://www.jstor.org/stable/j.ctt6wpzfz>.
9. J. M. Serb, T. H. Oakley, Hierarchical phylogenetics as a quantitative analytical framework for evolutionary developmental biology. *BioEssays* **27**, 1158–1166 (2005).
10. D. Arendt, Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* **47**, 563–571 (2003).
11. A. L. Burnett, A model of growth and cell differentiation in hydra. *Am. Nat.* **100**, 165–189 (1966).
12. G. S. Richards, F. Rentsch, Transgenic analysis of a *SoxB* gene reveals neural progenitor cells in the cnidarian *Nematostella vectensis*. *Development* **141**, 4681–4689 (2014).
13. C. N. David et al., Evolution of complex structures: Minicollagens shape the cnidarian nematocyst. *Trends Genet.* **24**, 431–438 (2008).
14. J. S. Hwang et al., The evolutionary emergence of cell type-specific genes inferred from the gene expression analysis of *Hydra*. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14735–14740 (2007).
15. G. Hemmrich et al., Molecular signatures of the three stem cell lineages in *Hydra* and the emergence of stem cell function at the base of multicellularity. *Mol. Biol. Evol.* **29**, 3267–3280 (2012).
16. L. S. Babonis, M. Q. Martindale, *PaxA*, but not *PaxC*, is required for cnidocyte development in the sea anemone *Nematostella vectensis*. *Evodevo* **8**, 14 (2017).
17. T. Chari et al., Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia* medusa cell types. *Sci. Adv.* **7**, eabh1683 (2021).
18. S. He et al., An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*. *Science* **361**, 1377–1380 (2018).
19. A. Karabulut, S. He, C.-Y. Chen, S. A. McKinney, M. C. Gibson, Electroporation of short hairpin RNAs for rapid and efficient gene knockdown in the starlet sea anemone, *Nematostella vectensis*. *Dev. Biol.* **448**, 7–15 (2019).
20. C. Zenkert, T. Takahashi, M.-O. Diesner, S. Özbek, Morphological and molecular analysis of the *Nematostella vectensis* cnidom. *PLoS One* **6**, e22725 (2011).
21. A. Sebé-Pedrós et al., Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534.e20 (2018).
22. O. Tournière et al., NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian *Nematostella vectensis*. *Cell Rep.* **30**, 4473–4489.e5 (2020).
23. F. Rentsch, M. Layden, M. Manuel, The cellular and molecular basis of cnidarian neurogenesis. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e257 (2017).

24. D. Gauchat *et al.*, The orphan COUP-TF nuclear receptors are markers for neurogenesis from cnidarians to vertebrates. *Dev. Biol.* **275**, 104–123 (2004).
25. S. Y. Kerk, P. Kratsios, M. Hart, R. Mourao, O. Hobert, Diversification of *C. elegans* motor neuron identity via selective effector gene repression. *Neuron* **93**, 80–98 (2017).
26. D. J. Duffy, U. Frank, Modulation of COUP-TF expression in a cnidarian by ectopic Wnt signalling and allrecognition. *PLoS One* **6**, e19443 (2011).
27. A. M. Reitzel, A. M. Tarrant, Nuclear receptor complement of the cnidarian *Nematostella vectensis*: Phylogenetic relationships and developmental expression patterns. *BMC Evol. Biol.* **9**, 230 (2009).
28. D. Arendt, H. Hausen, G. Purschke, The 'division of labour' model of eye evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2809–2817 (2009).
29. J. F. Schmitz, F. Zimmer, E. Bornberg-Bauer, Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res.* **44**, 6287–6297 (2016).
30. H. D. Tadepally, G. Burger, M. Aubry, Evolution of C2H2-zinc finger genes and subfamilies in mammals: Species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol. Biol.* **8**, 176 (2008).
31. R. Mackeh, A. K. Marr, A. Fadda, T. Kino, C2H2-type zinc finger proteins: Evolutionarily old and new partners of the nuclear hormone receptors. *Nucl. Recept. Signal.* **15**, 1550762918801071 (2018).
32. A. Kazanjian, E. A. Gross, H. L. Grimes, The growth factor independence-1 transcription factor: New functions and new insights. *Crit. Rev. Oncol. Hematol.* **59**, 85–97 (2006).
33. P. Lapébie *et al.*, Differential responses to Wnt and PCP disruption predict expression and developmental function of conserved and novel genes in a cnidarian. *PLoS Genet.* **10**, e1004590 (2014).
34. Y. Yasuoka, C. Shinzato, N. Satoh, The mesoderm-forming gene brachyury regulates ectoderm-endoderm demarcation in the coral *Acropora digitifera*. *Curr. Biol.* **26**, 2885–2892 (2016).
35. Y. Wenger, W. Buzgariu, B. Galliot, Loss of neurogenesis in *Hydra* leads to compensatory regulation of neurogenic and neurotransmission genes in epithelial cells. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150040 (2016).
36. S. Tomczyk *et al.*, Deficient autophagy in epithelial stem cells drives aging in the freshwater cnidarian *Hydra*. *Development* **147**, dev177840 (2020).
37. E. M. Erkenbrack, J. R. Thompson, Cell type phylogenetics informs the evolutionary origin of echinoderm larval skeletogenic cell identity. *Commun. Biol.* **2**, 160 (2019).
38. J. M. Musser, G. P. Wagner, Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called "species signal". *J. Exp. Zool. B Mol. Dev. Evol.* **324**, 588–604 (2015).
39. K. Kin, M. C. Nnamani, V. J. Lynch, E. Michaelides, G. P. Wagner, Cell-type phylogenetics and the origin of endometrial stromal cells. *Cell Rep.* **10**, 1398–1409 (2015).
40. N. U. Nair *et al.*, Study of cell differentiation by phylogenetic analysis using histone modification data. *BMC Bioinformatics* **15**, 269 (2014).
41. M. Lynch, A. Force, The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).
42. H. Flici *et al.*, An evolutionarily conserved SoxB-Hdac2 crosstalk regulates neurogenesis in a cnidarian. *Cell Rep.* **18**, 1395–1409 (2017).
43. M. Jager *et al.*, New insights on ctenophore neural anatomy: Immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *J. Exp. Zool. B Mol. Dev. Evol.* **316B**, 171–187 (2011).
44. A. Senatore, T. S. Reese, C. L. Smith, Neuropeptidergic integration of behavior in *Trichoplax adhaerens*, an animal without synapses. *J. Exp. Biol.* **220**, 3381–3390 (2017).
45. H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
46. L. S. Babonis, M. Q. Martindale, J. F. Ryan, Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC Evol. Biol.* **16**, 114 (2016).
47. J. A. Capra, K. S. Pollard, M. Singh, Novel genes exhibit distinct patterns of function acquisition and network integration. *Genome Biol.* **11**, R127 (2010).
48. G. P. Wagner, Evolutionary innovations and novelties: Let us get down to business! *Zool. Anz.* **256**, 75–81 (2015).
49. G. P. Wagner, M. Pavlicev, J. M. Cheverud, The road to modularity. *Nat. Rev. Genet.* **8**, 921–931 (2007).
50. K. H. Ten Tusscher, P. Hogeweg, Evolution of networks for body plan patterning: Interplay of modularity, robustness and evolvability. *PLoS Comput. Biol.* **7**, e1002208 (2011).
51. L. S. Babonis *et al.*, Integrating embryonic development and evolutionary history to characterize tentacle-specific cell types in a ctenophore. *Mol. Biol. Evol.* **35**, 2940–2956 (2018).
52. A. Brückner, J. Parker, Molecular evolution of gland cell types and chemical interactions in animals. *J. Exp. Biol.* **223**, jeb211938 (2020).
53. C. H. Hoyle, "Neuropeptides and their receptors: evolution" In *Encyclopedia of Life Sciences*, E. A. Williams, Ed. (John Wiley & Sons, Ltd, Chichester, UK, 2008), pp. 1–14.
54. T. L. Koch, C. J. P. Grimmelikhuijzen, Global neuropeptide annotations from the genomes and transcriptomes of Cubozoa, Scyphozoa, Staurozoa (Cnidaria: Medusozoa), and Octocorallia (Cnidaria: Anthozoa). *Front. Endocrinol. (Lausanne)* **10**, 831 (2019).
55. T. Takahashi, Comparative aspects of structure and function of cnidarian neuropeptides. *Front. Endocrinol. (Lausanne)* **11**, 339 (2020).
56. O. Hobert, P. Kratsios, Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr. Opin. Neurobiol.* **56**, 97–105 (2019).
57. M. Dohrmann, G. Wörheide, Dating early animal evolution using phylogenomic data. *Sci. Rep.* **7**, 3599 (2017).
58. M. Ahmed, D. R. Kim, pcr: An R package for quality assessment, analysis and testing of qPCR data. *PeerJ* **6**, e4473 (2018).
59. R Core Team, "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna, Austria, 2020). <https://www.R-project.org/>. Accessed 12 January 2020.
60. L. S. Babonis, J. F. Ryan, C. Enjolras, M. Q. Martindale, Genomic analysis of the tryptome reveals molecular mechanisms of gland cell evolution. *Evodevo* **10**, 23 (2019).
61. H.-M. Kim *et al.*, The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. *BMC Biol.* **17**, 28 (2019).
62. O. Simakov *et al.*, Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).