1 **Title: A novel regulatory gene promotes novel cell fate by suppressing ancestral fate in**

2 **the sea anemone *Nematostella vectensis***

3

4 **Authors:** Leslie S Babonis[1,2], Camille Enjolras[1], Joseph F Ryan[1,3] and Mark Q Martindale[1,3]

5

6 **Affiliations:**

7 1. Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL 32080

8 2. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca NY 14853

9   (current address)

10 3. Department of Biology, University of Florida, Gainesville, FL 32611

11

12 **Corresponding Author:** Leslie S Babonis, Department of Ecology and Evolutionary Biology,

13 Cornell University, E145 Corson Hall, Ithaca NY 14853; lsb257@cornell.edu; (808) 347-6562

14

15 **Authors Contributions:** LSB, CE, JFR, and MQM collected data and performed analyses; LSB

16 conceived of the study and wrote the manuscript; CE, JFR, and MQM edited and approved the

17 manuscript.

18

19 **Competing Interests:** The authors declare no competing interests.

20

21 **Classification:** Biological Science, Developmental Biology

22 **Keywords:** cell identity, novelty, evolution

23

24 **This PDF file includes:**

25 Main Text

26 Figures 1 to 4

27   **Abstract:** Cnidocytes ("stinging cells") are an unequivocally novel cell type used by cnidarians

28   (corals, jellyfish, and their kin) to immobilize prey. Although they are known to share a common

29   evolutionary origin with neurons, the developmental program that promoted the emergence of

30   cnidocyte fate is not known. Using functional genomics in the sea anemone, *Nematostella*

31   *vectensis*, we show that cnidocytes evolved by suppression of neural fate in a subset of neurons

32   expressing RFamide. We further show that a single regulatory gene, a C_2H_2-type zinc finger

33   transcription factor (ZNF845), coordinates both the gain of novel (cnidocyte-specific) traits and

34   the inhibition of ancestral (neural) traits during cnidocyte development and that this gene arose

35   by domain shuffling in the stem cnidarian. Thus, we uncover a mechanism by which a truly

36   novel regulatory gene (ZNF845) promoted the origin of a truly novel cell type (cnidocyte)

37   through duplication of an ancestral cell lineage (neuron) and inhibition of its ancestral identity

38   (RFamide).

39

40   **Significance:** In this study, we demonstrate how new cell types can arise in animals through

41   duplication of an ancestral (old) cell type followed by functional divergence of the new daughter

42   cell. Specifically, we show that stinging cells in cnidarians (jellyfish and corals) evolved by

43   duplication of an ancestral neuron followed by inhibition of the RFamide neuropeptide it once

44   secreted. This is the first evidence that stinging cells evolved from a specific subtype of neurons

45   and suggests some neurons may be easier to co-opt for novel functions than others.

46

47   **Main Text**

48   **Introduction**

49   Understanding the mechanisms driving cell type diversification persists as one of the key

50   challenges in evolutionary biology (*1*). The gain of new adaptive cell functions requires either

51   the advent of novel genes (*2, 3*). the modification of existing gene regulatory networks (*4*), or

52   some combination of these two processes (*5*). Alone, this additive model, focused simply on the

53   emergence of novel gene interactions, is insufficient to explain expansion of cell identity as new

54   cell types would arise in place of ancestral cell types. In a process analogous to gene

55   duplication and divergence (*6*), new instances of cell division during embryogenesis could lead

56   to duplication of a cell lineage, providing the opportunity for one lineage to retain an ancestral

57   function and the other to acquire new functions. Cell type diversification, therefore, requires both

58   novel gene interactions and a novel cell lineage in which to express these traits.

59

60   Cnidarians are an unparalleled model for studying the evolution of novelty because the defining

61   synapomorphy of this group (the cnidocyte or "stinging cell") is an unequivocally novel cell type.

62   During embryogenesis, cnidocytes differentiate from a progenitor cell that also gives rise to

63   neurons, reflecting a common evolutionary origin for these two cell types (*7*, *8*). Two key

64   features were necessary for the transition away from neural fate in early cnidocytes: the

65   development of the explosive secretory organelle (the cnido<u>cyst</u>) from which cnidocytes derive

66   their "sting," and the suppression of neural phenotype (axons, synaptic signaling molecules,

67   etc). Studies tracking the synthesis of the cnidocyst-specific protein minicollagen have revealed

68   key steps leading to the origin of this novel organelle (*9*, *10*); the mechanisms driving the

69   suppression of neural phenotype in cnidocytes remain unknown. Here, we show that a single

70   novel transcription factor, ZNF845, both promotes cnidocyte fate and suppresses neural fate

71   during development of the sea anemone *Nematostella vectensis*. We further show that the six-

72   domain topology of ZNF845 arose through domain shuffling in the last common ancestor of

73   cnidarians, making this a clear example of a cnidarian-specific gene driving development of a

74   cnidarian-specific trait.

75

76   **Results**

77   **ZNF845 promotes cnidocyte fate**

78   Beginning at the blastula stage, ZNF845 is expressed throughout the embryo in cells that are

79   actively undergoing DNA synthesis, as labeled by EdU (**Fig. 1A**). ZNF845 is co-expressed with

80   SoxB2 in a subset of embryonic neural progenitor cells at the blastula stage (**Fig. 1B**) and is

81   also co-expressed in a subset of PaxA-expressing developing cnidocytes at the gastrula stage

82   (**Fig. 1C**). ZNF845 continues to be expressed through metamorphosis in a pattern reminiscent

83   of cnidocyte development (Supplementary Material Fig. S1). To understand the role of ZNF845

84   in neural/cnidocyte differentiation, we knocked down SoxB2 and PaxA using previously

85   published morpholinos (MOs) (*8*, *11*). ZNF845 expression was downregulated in embryos

86   injected with the SoxB2 MO but unaffected in embryos injected with PaxA MO (**Fig. 1D**). Paired

87   with the co-expression results, these data suggest that ZNF845 is part of the cnidogenesis

88   pathway acting upstream of PaxA in *N. vectensis*.

89

90   To further explore the role of ZNF845, we knocked down ZNF845 using small-hairpin RNAs

91   (shRNAs) (*12*) and assayed the effects on cnidocyte development. Knockdown of ZNF845 was

92   effective through the late planula stage (Supplementary Material Fig. S1) and resulted in nearly

93   complete loss of cnidocytes throughout the ectoderm of the planula (**Fig. 1E**). Using an antibody

94   directed against the cnidocyte-specific protein minicollagen4 (α-Mcol4) (*13*) we demonstrate a

95   significant loss of cnidocytes: from 10% of total cells in wildtype (WT) and control (ctrl) shRNA-

96   injected embryos to less than 3% when ZNF845 was knocked down. We recovered identical

97   results in embryos injected with a splice-blocking ZNF845 MO, relative to those injected with a

98   standard control MO (Supplementary Material Fig. S2).

99

100   To determine if ZNF845 acts upstream of PaxA and other markers specific to the cnidogenesis

101   pathway, we examined the effect of ZNF845 knockdown on known markers of neural and

102   cnidocyte differentiation in *N. vectensis* using quantitative PCR (**Fig. 1F, Table S2**). Knockdown

103   of ZNF845 resulted in significant downregulation of PaxA, and all of its known targets (GGT,

104   Ngal, Mcol1, Mcol3, Mcol4, and Mef2IV) (*11*) as well as three cnidocyte-expressed transcription

105   factors identified from single-cell RNA-Seq analysis (TEA/Scalloped, FoxL2, Pou4) (*14*).

106   Conversely, ZNF845 knockdown did not affect the expression of SoxB2 or the neuron-specific

107   regulatory genes Atonal and ELAV (*15*). While ZNF845 knockdown caused a statistically

108   significant increase in the expression of neural markers AshA, GLWamide, LWamide, and PaxC

109   assayed by qPCR, the response of these genes was minor relative to the large, significant

110   upregulation of RFamide. To spatially characterize these results, we performed *in situ*

111   hybridization for RFamide and PaxC in embryos injected with control- or ZNF845 shRNAs and

112   counted the number of cells expressing these neural markers at the gastrula stage (**Fig. 1G**).

113   Knockdown of ZNF845 significantly increased the number of RFamide-expressing cells from

114   2.23 (± 1.95; mean ± SD) in control embryos to 23.47 (± 17.85; mean ± SD) in ZNF845

115   knockdowns. Cell counts for PaxC expression revealed a small, non-significant increase in the

116   number of PaxC-expressing cells from 8.9 (± 4.4; mean ± SD) in controls to 14.3 (± 18.5; mean

117   ± SD) in ZNF845 knockdowns. Because PaxC expression during embryogenesis is reminiscent

118   of RFamide expression, we hypothesized that PaxC might be an upstream regulator of

119   RFamide neuron differentiation. On the contrary, we found that PaxC and RFamide are not co-

120   expressed (**Fig. 1H**) and that knockdown of PaxC using a previously published MO (*11*) did not

121   affect the number or distribution of RFamide-expressing cells (**Fig. 1I**).

122

123   These results suggest that there is a closer evolutionary relationship between cnidocytes and

124   RFamide-expressing neurons than between cnidocytes and other neural subtypes (**Fig. 1J**).

125   This is congruent with a recent study showing that the selector gene Pou4 regulates terminal

126   cell identity in cnidocytes and RFamide-expressing neurons in *N. vectensis* but not in other cell

127   types (*16*). Second, we demonstrate that a single transcription factor (ZNF845) upregulates both

128   the genes necessary to promote cnidocyte identity and the genes necessary to inhibit RFamide

129   neuron identity. To understand the mechanism by which ZNF845 suppresses RFamide

130   expression during cnidocyte differentiation, we searched for inhibitory transcription factors that

131   were co-expressed with cnidocyte specific genes.

132

133   **ZNF845 inhibits neural fate through NR12**

134   Nuclear receptors in the COUP-TF family (NR2F) are known to play an inhibitory role in neural

135   cell fate decisions in both cnidarians (*17*) and bilaterians (*18*). In *N. vectensis,* there are five

136   NR2F paralogs: NR10-14, most of which appear to have originated through lineage-specific

137   duplication in cnidarians (*19*). We examined the expression of all five NR2F genes

138   (Supplementary Material Fig. S3) and found that three of them (NR11,12,13) were expressed in

139   the ectoderm during early embryogenesis and were also downregulated in embryos injected

140   with ZNF845 shRNA (**Fig. 2A**). All three were also co-expressed with Mcol4 in differentiating

141   cnidocytes (**Fig. 2B-D**) and yet we found no evidence of that any of the NR2F paralogs were co-

142   expressed together in the same cell (**Fig. 2E,F**) suggesting these NR2F paralogs are expressed

143   uniquely in different subpopulations of cnidocytes.

144

145   The cnidocyte-specific expression of NR12 that we observed was also supported by single cell

146   RNA-Seq in *N. vectensis* (*14*). We examined NR12 further and found that it is co-expressed in a

147   subset of ZNF845-expressing cells at the gastrula stage (**Fig. 2G**) and in a subset of PaxA-

148   expressing cells at the early planula stage (**Fig. 2H**). We then examined the effects of NR12 on

149   cnidocyte development using shRNAs (Supplementary Material Fig. S4, Table S1) and found

150   that knockdown of NR12 resulted in a four-fold increase in the number of RFamide-expressing

151   cells but had no effect on expression of PaxC (**Fig. 2I**) or on the specification of cnidocytes (**Fig.**

152   **2J**). Knockdown of NR12 also had no effect on the number or distribution of NR11- or NR13-

153   expressing cells (**Fig. 2K**).

154

155   Modularity in the regulation of neural phenotype would have enabled cnidocytes to retain

156   beneficial aspects of the ancestral phenotype while silencing others through selective inhibition,

157   as has been shown for neural subtype specification in C. elegans (*18*). The upregulation of

158   RFamide following NR12 knockdown suggests that RFamide expression may be coupled to the

159   expression of another trait that had adaptive value during the evolution of cnidocytes (e.g.,

160   secretory vesicles) (**Fig. 2L**). The independent expression of NR11 and NR13 in non-

161   overlapping populations of developing cnidocytes suggests that multiple cnidocyte subtypes

162   may be specified through NR2F-mediated inhibition of neural traits in other (non-RFamide)

163   neural subtypes.

164

165   **ZNF845 is a novel transcription factor**

166   Transcription factors with Cys2-His2 zinc finger (ZF-C_2H_2) domains (PF00096) represent one of

167   the largest families of transcription factors across animals (*20*). Numerous evolutionary

168   processes have contributed to diversification in this gene family, including

169   duplication/divergence, gain/loss of ZF-C_2H_2 domains, and gain/loss of accessory domains (*21*).

170   Among the oldest members of this clade are ZNF proteins with multiple tandem C_2H_2 domains

171   but no other conserved functional domains (*22*). ZNF845 has six tandem ZF-C_2H_2 domains; to

172   understand how these six domains came together in a single protein, we generated a maximum

173   likelihood phylogeny of all ZF-C_2H_2 domains extracted from predicted proteomes for three

174   bilaterian taxa (*Homo sapiens*, *Drosophila melanogaster*, and *Caenorhabditis elegans*) and four

175   cnidarians (two anthozoans: *N. vectensis* and *Acropora digitifera*, and two medusozoans: *Hydra*

176   *magnipapillata* and *Nemopilema nomurai*). The full phylogeny contains over 11,000 branch tips;

177   the FASTA alignment is provided in Supplementary Material Data S1.

178

179   *Nematostella vectensis* has 218 predicted proteins with one or more ZF-C_2H_2 domains, sixteen

180   of which, including ZNF845, encode six tandem ZF-C_2H_2 domains. We examined the

181   evolutionary relationships of the ZF-C_2H_2 domains from ZNF845 (JGI protein ID: 81344) and

182   compared the results with an analysis of another six-domain ZNF protein, Growth Factor

183   Independence 1B (Gfi1B; JGI PID: 112378). Gfi1B is a potent regulator of cell differentiation in

184   vertebrates (*23*), but the function of this protein has not been characterized in *N. vectensis*.

185   Each of the six domains from the *N. vectensis* ortholog of ZNF845 groups with ZF-C_2H_2 domains

186   from the orthologous ZNF845 protein in other cnidarians: Hmag_XP_002158383.2 (originally

187   named ZNF845 by Hemmrich et al. (*24*)), Nnom_10870, and Nnom_3755 (**Fig. 3A** and

188   Supplementary Material Fig. S5). These relationships suggest the anthozoan and medusozoan

189   ZNF845 proteins descended from a common ancestor that already had six ZF-C_2H_2 domains in

190   the stem cnidarian. Additionally, two of these domains (domains 1 and 5) appear to have arisen

191   by tandem duplication as these domains form a clade that lacks other ZF-C_2H_2 domains. Each

192   of the six ZF-C_2H_2 domains in ZNF845 shares some level of homology with a bilaterian ZF-C_2H_2

193   domain, but none of these bilaterian domains are found in the same protein.

194

195   Examination of the ZF-C_2H_2 domains from Gfi1B suggests that, unlike ZNF845, this six-domain

196   transcription factor emerged in its current form before cnidarians and bilaterians diverged from

197   their common ancestor (**Fig. 3B**, Supplementary Material Fig. S6). Each of the ZF-C_2H_2

198   domains from the *N. vectensis* ortholog of Gfi1B grouped with the syntenic ZF-C_2H_2 domain

199   from the Gfi1B ortholog in each of the bilaterian taxa examined (Homo_NP_001120687.1,

200   Homo_XP_006717360.1, Dmel_Q9VM77, Cele_5376). Together, these observations suggest

201   that ZNF845 arose as a novel 6-domain protein in the stem cnidarian and that both domain

202   shuffling and domain duplication/divergence were important for the emergence of this protein

203   (**Fig. 3C**).

204

205   **Modeling cell type expansion**

206   Gene duplication and divergence is an important source for novel gene function (*6, 25*).

207   Extending this concept, we propose a model for the emergence of novel cell types through cell

208   lineage duplication and divergence (**Fig. 4**). New cells are generated from progenitors during

209   growth and tissue repair (**Fig. 4A**). Co-option of a progenitor cell to produce a novel daughter

210   would not increase the number of differentiated cell types if the novel cell fate simply replaced

211   its ancestor (**Fig. 4B**). Duplication of the progenitor cell first would allow for maintenance of

212   ancestral cell identity and the origin of novel identity, increasing cell type diversity (**Fig. 4C**).

213

214   These basic principles can be extended to explain the emergence of cnidocytes from a neural

215   precursor (**Fig. 4D,E**). At least one SoxB gene was present in the ancestor of all animals (*26*).

216   Analyses of animal neural diversity have suggested RFamide-like neurons arose early in animal

217   evolution as well (*27, 28*). To generate a novel cell lineage, the progenitor cell would first have to

218   duplicate to generate a new progenitor lineage. Absent additional changes, duplication of a

219   neural progenitor cell would have doubled the number of RFamide-expressing daughters,

220   potentially leading to aberrant synapse formation. If progenitor cell duplication were coupled to

221   the origin of a transcription factor that inhibited RFamide expression, the original number of

222   RFamide-expressing daughters would be maintained and RFamide neurons would now be

223   sister to a second lineage of "neurons" lacking a synaptic payload. Key to the emergence of

224   cnidocytes, therefore, was the origin of two transcription factors: one (ZNF845) that could

225   maintain progenitor cell fate in a newly duplicated cell lineage and another (NR12) that could

226   manipulate the identity of the secreted payload in the duplicated daughter cell. Without the

227   selection pressure to maintain synaptic signaling in this new daughter cell, these cells may have

228   experienced relaxed selection for the maintenance of axons, allowing the secretory vesicles to

229   relocate to the cell body. Subsequent mutations resulting in the emergence of a novel payload

230   (e.g., minicollagen) would have further promoted daughter cell divergence following duplication

231   in this cell lineage.

232

Discussion

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 (*29–31*), cnidocytes have always provided a clear example of an adaptive role for novel effector genes (e.g., minicollagen) in driving the evolution of a novel cell phenotype (*9*). In the present study, we extend this adaptive role for novel genes up the cnidocyte gene regulatory network by showing that the emergence of a new transcription factor assembled through domain shuffling in the stem cnidarian (ZNF845) was essential for the origin of cnidocytes.

Second, we demonstrate support for the hypothesis that modularity is an important driver of phenotypic evolution (*32, 33*). During the early divergence of cnidocytes, modularity in neural cell phenotype allowed for selective inhibition of certain traits (e.g., vesicular payload, axons) and retention of others (e.g., secretory vesicle). In this scenario, a single mutation that allowed for inhibition of RFamide expression could rapidly change the selection pressures affecting the recently duplicated sister cells, promoting retention of ancestral phenotype in one and allowing additional mutations to arise in the other.

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 (*34, 35*). 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 (*36–38*), could also explain the

258   rapid expansion of neural function in early bilaterians. Studies of neural fate in bilaterian model

259   systems have demonstrated that discrimination of specific subtypes within a neural lineage

260   relies on inhibition of the effector genes that define other subtypes in the lineage (*39*), and that

261   this inhibition of sister cell fate is mediated through the actions of conserved transcription factors

262   including orthologs of NR2F/COUP-TF (*18*). The data presented here indicate that this role for

263   NR2F-mediated inhibition of cell fate may extend back to the common ancestor of cnidarians

264   and bilaterians, suggesting the divergence-through-inhibition regulatory logic was already

265   driving the expansion of cell fate nearly 700 million years ago (*40*).

266

267   **Materials and Methods**

268   **Gene knockdowns**: To assess the influence of ZNF845 and NR12 on cnidocyte fate, we

269   performed mRNA knockdown by microinjection of small-hairpin RNAS (shRNAs) following the

270   protocol of He et al. (*12*). shRNAs used in this study were synthesized *in vitro* (primer

271   sequences provided in Table S1), diluted to 800ng/ul in nuclease-free water (Ambion AM9937)

272   with a final concentration of 0.2 mg/ml Alexa-555 RNAse free dextran (Invitrogen D34679) to

273   facilitate injection. To account for non-specific effects, control embryos were injected with an

274   shRNA (Table S1; 800ng/ul) that was not complimentary to any part of the genome (*42)*.

275   Embryos were raised to the early planula stage (72 hours post fertilization) at 16C and effects of

276   knockdowns were assayed via immunofluorescence, *in situ* hybridization, or quantitative PCR

277   (qPCR). Independent confirmation of the effect of ZNF845 knockdown on cnidocyte

278   specification was assayed using a splice-blocking morpholino (GeneTools; Table S1) as

279   previously described (*11*). Briefly, lyophilized morpholinos were reconstituted in nuclease-free

280   water to 1 mM following the manufacturer's instructions. Before each use, the stock was heated

281   to 60C for 5 mins and centrifuged for ~1 minute before being diluted to a final working

282   concentration of 0.3 mM in nuclease free water with 0.2 mg/ml RNase-free dextran. To control

283   for non-specific effects, ZNF845-injected embryos were compared to embryos injected with a

284   standard control morpholino (GeneTools; Table S1) prepared the same way and injected at the

285   same concentration as the ZNF845 MO. Splicing defects were confirmed using PCR and gel

286   electrophoresis as described previously (*11*).

287

288   **Cell and tissue analysis:** To assay effects of gene knockdown on cnidocyte specification,

289   developing cnidocytes were labeled, imaged, and counted as described in a similar study (*11*)

290   using an antibody directed against minicollagen4 (Mcol4) (*13*). Additional targets of ZNF845 and

291   NR12 shRNAs were assessed using *in situ* hybridization as previously described (*11*). For

292   qPCR analysis, embryos injected with control shRNA and embryos injected with ZNF845

293   shRNA were both compared to uninjected embryos raised under the same conditions. Five

294   replicates of each condition (ZNFshRNA, control shRNA, and uninjected/WT) representing five

295   independent injections performed on different days were compared using the delta-delta CT

296   method and the PCR package (43) in the R statistical computing environment (*44*). ZNF845

297   expression in Figure 1D are presented as fold-change, relative to ZNF845 expression in

298   uninjected embryos and statistical significance for qPCR in Figure 1D was calculated from the

299   comparison of SoxB2 MO and PaxA MO-injected embryos relative to control MO-injected

300   embryos. mRNA expression values in Figures 1F and 2A are presented as fold-change relative

301   to expression of EF1B in the uninjected embryos (arbitrarily set to 1) and significance was

302   calculated from the comparison of ZNF845 shRNA-injected embryos to control shRNA-injected

303   embryos. Cell counts in Figures 1E, 1G, 1I, 2I, 2J, and 2K were analyzed with a Mann-Whitney

304   U nonparametric test for two-way comparisons (control vs target shRNA/MO) and are presented

305   as mean ± standard deviation. Statistical significance for all quantitative comparisons is

306   indicated (*) where p<1E-02.

307

308   **Maximum likelihood phylogeny:** Phylogenetic analysis of ZF-C_2H_2 domains was performed

309   using a modification of a previously published protocol (*45*). In brief, an alignment was

310   generated using a custom script (hmm2aln, available at github.com/josephryan) and the ZF-

311   C_2H_2 HMM (PF00096) in predicted proteomes from all target taxa. For cnidarians, we sampled

312   two anthozoans: *Nematostella vectensis* and *Acropora digitifera*, and two medusozoans: *Hydra*

313   *magnipapillata* and *Nemopilema nomurai* (*46*). For bilaterians, we sampled *Homo sapiens*,

314   *Drosophila melanogaster*, and *Caenorhabditis elegans.* Download information for each taxon is

315   provided in Table S3. The alignment contains over 11,000 ZF-C2H2 domains from these

316   combined taxa and is provided in Data S1. To generate the phylogeny, we first used the model

317   finder function (-MF) with IQTREE to determine the best substitution model (VT+R8) and then

318   generated a single tree and applied 500 bootstraps using fast bootstrapping.

319

Acknowledgements and Funding

326

References

328   1.   D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev,

329   G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell

330   types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

331   2.   J. Zhang, Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292–298

332   (2003).

333   3.   S. B. V. Oss, A.-R. Carvunis, De novo gene birth. *PLOS Genet.* **15**, e1008160 (2019).

334 4. D. H. Erwin, E. H. Davidson, The evolution of hierarchical gene regulatory networks. *Nat.*

335   *Rev. Genet.* **10**, 141–148 (2009).

336 5. G. P. Wagner, V. J. Lynch, The gene regulatory logic of transcription factor evolution.

337   *Trends Ecol. Evol.* **23**, 377–385 (2008).

338 6. A. Force, M. Lynch, F. B. Pickett, A. Amores, Y. Yan, J. Postlethwait, Preservation of

339   Duplicate Genes by Complementary, Degenerative Mutations. *Genetics.* **151**, 1531–1545

340   (1999).

341 7. A. L. Burnett, A Model of Growth and Cell Differentiation in Hydra. *Am. Nat.* **100**, 165–189

342   (1966).

343 8. G. S. Richards, F. Rentzsch, Transgenic analysis of a *SoxB* gene reveals neural progenitor

344   cells in the cnidarian *Nematostella vectensis*. *Development*. **141**, 4681–4689 (2014).

345 9. C. N. David, S. Özbek, P. Adamczyk, S. Meier, B. Pauly, J. Chapman, J. S. Hwang, T.

346   Gojobori, T. W. Holstein, Evolution of complex structures: minicollagens shape the

347   cnidarian nematocyst. *Trends Genet.* **24**, 431–438 (2008).

348 10. J. S. Hwang, H. Ohyanagi, S. Hayakawa, N. Osato, C. Nishimiya-Fujisawa, K. Ikeo, C. N.

349   David, T. Fujisawa, T. Gojobori, The evolutionary emergence of cell type-specific genes

350   inferred from the gene expression analysis of Hydra. *Proc. Natl. Acad. Sci.* **104**, 14735–

351   14740 (2007).

352 11. L. S. Babonis, M. Q. Martindale, PaxA, but not PaxC, is required for cnidocyte

353   development in the sea anemone *Nematostella vectensis*. *EvoDevo*. **8**, 14 (2017).

12. S. He, F. del Viso, C.-Y. Chen, A. Ikmi, A. E. Kroesen, M. C. Gibson, An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*. *Science*. **361**, 1377–1380 (2018).

13. C. Zenkert, T. Takahashi, M.-O. Diesner, S. Özbek, Morphological and Molecular Analysis of the *Nematostella vectensis* Cnidom. *PLoS ONE*. **6**, e22725 (2011).

14. A. Sebé-Pedrós, B. Saudemont, E. Chomsky, F. Plessier, M.-P. Mailhé, J. Renno, Y. Loe-Mie, A. Lifshitz, Z. Mukamel, S. Schmutz, S. Novault, P. R. H. Steinmetz, F. Spitz, A. Tanay, H. Marlow, Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell*. **173**, 1520-1534.e20 (2018).

15. F. Rentzsch, M. Layden, M. Manuel, The cellular and molecular basis of cnidarian neurogenesis: Cnidarian neurogenesis: cellular and molecular basis. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e257 (2017).

16. O. Tournière, D. Dolan, G. S. Richards, K. Sunagar, Y. Y. Columbus-Shenkar, Y. Moran, F. Rentzsch, NvPOU4/Brain3 Functions as a Terminal Selector Gene in the Nervous System of the Cnidarian *Nematostella vectensis*. *Cell Rep.* **30**, 4473-4489.e5 (2020).

17. D. Gauchat, H. Escriva, M. Miljkovic-Licina, S. Chera, M.-C. Langlois, A. Begue, V. Laudet, B. Galliot, The orphan COUP-TF nuclear receptors are markers for neurogenesis from cnidarians to vertebrates. *Dev. Biol.* **275**, 104–123 (2004).

18. 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).

374   19.   A. M. Reitzel, A. M. Tarrant, Nuclear receptor complement of the cnidarian *Nematostella*

375   *vectensis*: phylogenetic relationships and developmental expression patterns. *BMC Evol.*

376   *Biol.* **9**, 230 (2009).

377   20.   J. F. Schmitz, F. Zimmer, E. Bornberg-Bauer, Mechanisms of transcription factor evolution

378   in Metazoa. *Nucleic Acids Res.* **44**, 6287–6297 (2016).

379   21.   H. D. Tadepally, G. Burger, M. Aubry, Evolution of C2H2-zinc finger genes and subfamilies

380   in mammals: Species-specific duplication and loss of clusters, genes and effector domains.

381   *BMC Evol. Biol.* **8**, 176 (2008).

382   22.   R. Mackeh, A. K. Marr, A. Fadda, T. Kino, C2H2-Type Zinc Finger Proteins: Evolutionarily

383   Old and New Partners of the Nuclear Hormone Receptors. *Nucl. Recept. Signal.* **15**,

384   1550762918801071 (2018).

385   23.   A. Kazanjian, E. A. Gross, H. L. Grimes, The growth factor independence-1 transcription

386   factor: New functions and new insights. *Crit. Rev. Oncol. Hematol.* **59**, 85–97 (2006).

387   24.   G. Hemmrich, K. Khalturin, A.-M. Boehm, M. Puchert, F. Anton-Erxleben, J. Wittlieb, U. C.

388   Klostermeier, P. Rosenstiel, H.-H. Oberg, T. Domazet-Lošo, T. Sugimoto, H. Niwa, T. C.

389   G. Bosch, Molecular Signatures of the Three Stem Cell Lineages in Hydra and the

390   Emergence of Stem Cell Function at the Base of Multicellularity. *Mol. Biol. Evol.* **29**, 3267–

391   3280 (2012).

392   25.   M. Lynch, A. Force, The Probability of Duplicate Gene Preservation by

393   Subfunctionalization. *Genetics*. **154**, 459–473 (2000).

394   26.   H. Flici, C. E. Schnitzler, R. C. Millane, G. Govinden, A. Houlihan, S. D. Boomkamp, S.

395   Shen, A. D. Baxevanis, U. Frank, An Evolutionarily Conserved SoxB-Hdac2 Crosstalk

396   Regulates Neurogenesis in a Cnidarian. *Cell Rep.* **18**, 1395–1409 (2017).

397   27.   M. Jager, R. Chiori, A. Alié, C. Dayraud, E. Quéinnec, M. Manuel, New insights on

398   ctenophore neural anatomy: Immunofluorescence study in Pleurobrachia pileus (Müller,

399   1776). *J. Exp. Zoolog. B Mol. Dev. Evol.* **316B**, 171–187 (2011).

400   28.   A. Senatore, T. S. Reese, C. L. Smith, Neuropeptidergic integration of behavior in

401   *Trichoplax adhaerens* , an animal without synapses. *J. Exp. Biol.* **220**, 3381–3390 (2017).

402   29.   H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**,

403   1313–1326 (2010).

404   30.   L. S. Babonis, M. Q. Martindale, J. F. Ryan, Do novel genes drive morphological novelty?

405   An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC*

406   *Evol. Biol.* **16**, 114 (2016).

407   31.   J. A. Capra, K. S. Pollard, M. Singh, Novel genes exhibit distinct patterns of function

408   acquisition and network integration. *Genome Biol.* **11**, R127 (2010).

409   32.   G. P. Wagner, M. Pavlicev, J. M. Cheverud, The road to modularity. *Nat. Rev. Genet.* **8**,

410   921–931 (2007).

411   33.   K. H. ten Tusscher, P. Hogeweg, Evolution of Networks for Body Plan Patterning; Interplay

412   of Modularity, Robustness and Evolvability. *PLOS Comput. Biol.* **7**, e1002208 (2011).

413   34.   L. S. Babonis, M. B. DeBiasse, W. R. Francis, L. M. Christianson, A. G. Moss, S. H. D.

414   Haddock, M. Q. Martindale, J. F. Ryan, Integrating embryonic development and

415   evolutionary history to characterize tentacle-specific cell types in a ctenophore. *Mol. Biol.*

416   *Evol.* (2018), doi:10.1093/molbev/msy171.

417 35. A. Brückner, J. Parker, Molecular evolution of gland cell types and chemical interactions in

418   animals. *J. Exp. Biol.* **223** (2020), doi:10.1242/jeb.211938.

419 36. C. H. Hoyle, in *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, Ed. (John Wiley &

420   Sons, Ltd, Chichester, UK, 2008;

421   http://doi.wiley.com/10.1002/9780470015902.a0006150.pub2), p. a0006150.pub2.

422 37. T. L. Koch, C. J. P. Grimmelikhuijzen, Global Neuropeptide Annotations From the

423   Genomes and Transcriptomes of Cubozoa, Scyphozoa, Staurozoa (Cnidaria: Medusozoa),

424   and Octocorallia (Cnidaria: Anthozoa). *Front. Endocrinol.* **10**, 831 (2019).

425 38. T. Takahashi, Comparative Aspects of Structure and Function of Cnidarian Neuropeptides.

426   *Front. Endocrinol.* **11** (2020), doi:10.3389/fendo.2020.00339.

427 39. O. Hobert, P. Kratsios, Neuronal identity control by terminal selectors in worms, flies, and

428   chordates. *Curr. Opin. Neurobiol.* **56**, 97–105 (2019).

429 40. M. Dohrmann, G. Wörheide, Dating early animal evolution using phylogenomic data. *Sci.*

430   *Rep.* **7**, 3599 (2017).

431 41. M. Y. Sachkova, E.-L. Nordmann, J. J. Soto-Àngel, Y. Meeda, B. Górski, B. Naumann, D.

432   Dondorp, M. Chatzigeorgiou, M. Kittelmann, P. Burkhardt, *bioRxiv*, in press,

433   doi:10.1101/2021.03.31.437758.

434 42. A. Karabulut, S. He, C.-Y. Chen, S. A. McKinney, M. C. Gibson, Electroporation of short

435   hairpin RNAs for rapid and efficient gene knockdown in the starlet sea anemone,

436   *Nematostella vectensis*. *Dev. Biol.* **448**, 7–15 (2019).

437 43. M. Ahmed, D. R. Kim, pcr: an R package for quality assessment, analysis and testing of

438   qPCR data. *PeerJ*. **6**, e4473 (2018).

439 44. R Core Team, R: A language and environment for statistical computing. *R Found. Stat.*

440   *Comput. Vienna Austria* (2020) (available at https://www.R-project.org/).

441 45. L. S. Babonis, J. F. Ryan, C. Enjolras, M. Q. Martindale, Genomic analysis of the tryptome

442   reveals molecular mechanisms of gland cell evolution. *EvoDevo*. **10**, 23 (2019).

443 46. H.-M. Kim, J. A. Weber, N. Lee, S. G. Park, Y. S. Cho, Y. Bhak, N. Lee, Y. Jeon, S. Jeon,

444   V. Luria, A. Karger, M. W. Kirschner, Y. J. Jo, S. Woo, K. Shin, O. Chung, J.-C. Ryu, H.-S.

445   Yim, J.-H. Lee, J. S. Edwards, A. Manica, J. Bhak, S. Yum, The genome of the giant

446   Nomura's jellyfish sheds light on the early evolution of active predation. *BMC Biol.* **17**, 28

447   (2019).

448

449

Figures and Figure Legends

451
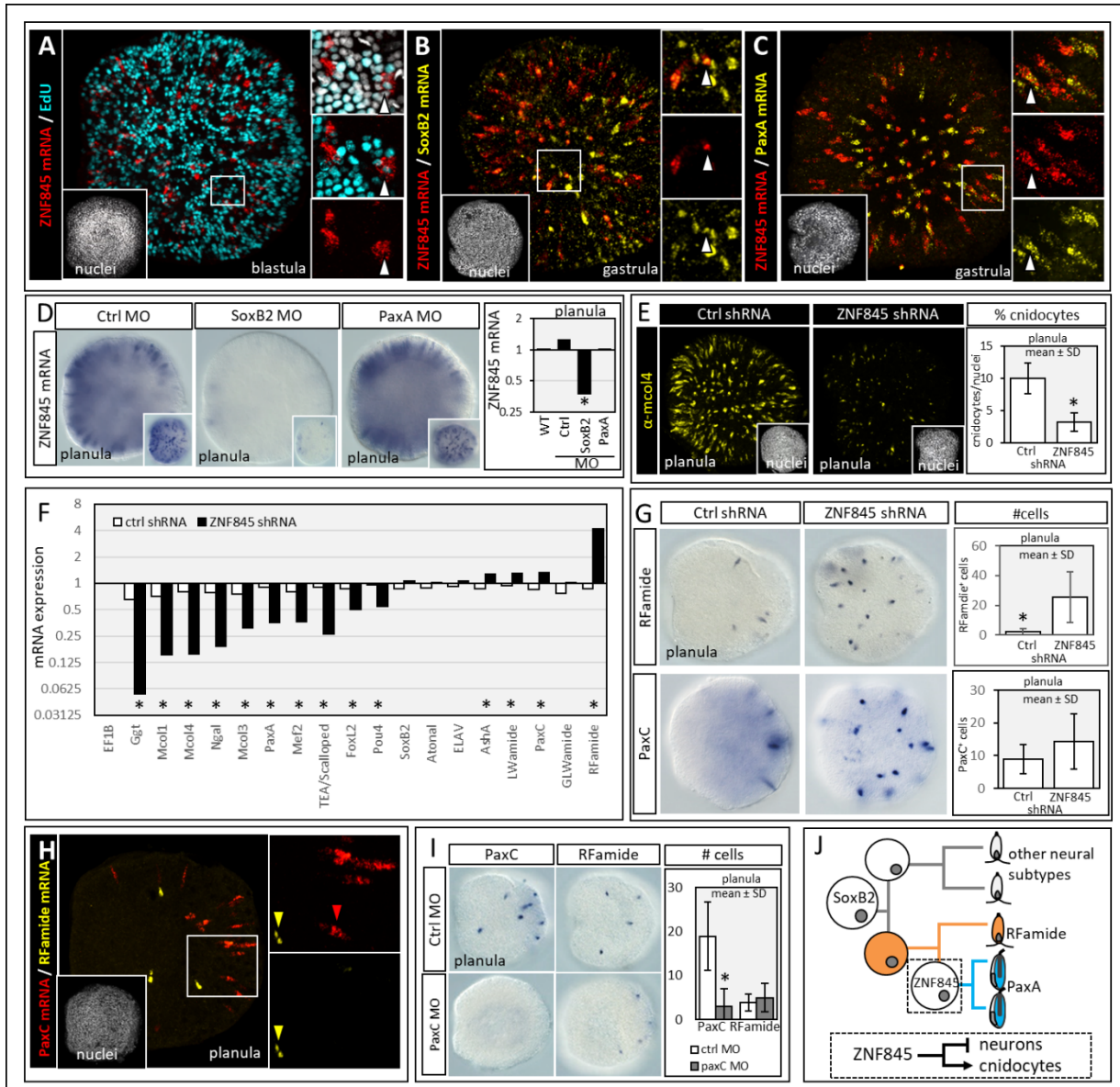
Fig. 1. ZNF845 specifies cnidocyte identity in *N. vectensis*. (A-C) ZNF845 is partially co-expressed with: (A) EdU (proliferating cells), (B) SoxB2 mRNA (neural progenitor cells), and (C) paxA mRNA (cnidocytes). Insets: white arrowheads show co-expression; nuclei are white (DAPI). (D) ZNF845 expression after knockdown of SoxB2 and PaxA (by morpholino, MO) assayed by *in situ* hybridization and qPCR; significance (*) is indicated for SoxB2MO or PaxA MO vs Ctrl MO. (E) Cnidocyte differentiation (α-Mcol4 antibody) after ZNF845 knockdown (shRNA). (F) qPCR of target gene expression after ZNF845 knockdown; fold-change relative to

459   housekeeping gene EF1B. (G) Response of RFamide- and PaxC-expressing cells to ZNF845

460   knockdown. (H) RFamide and PaxC are not co-expressed. (I) Expression of RFamide and PaxC

461   after knockdown of PaxC (MO). (J) Model for ZNF845-mediated specification of cnidocyte

462   identity from and RFamide-expressing cell lineage. Significance (*) for all tests indicated as

463   p<1E-02. See Table S2 for supporting information.

464
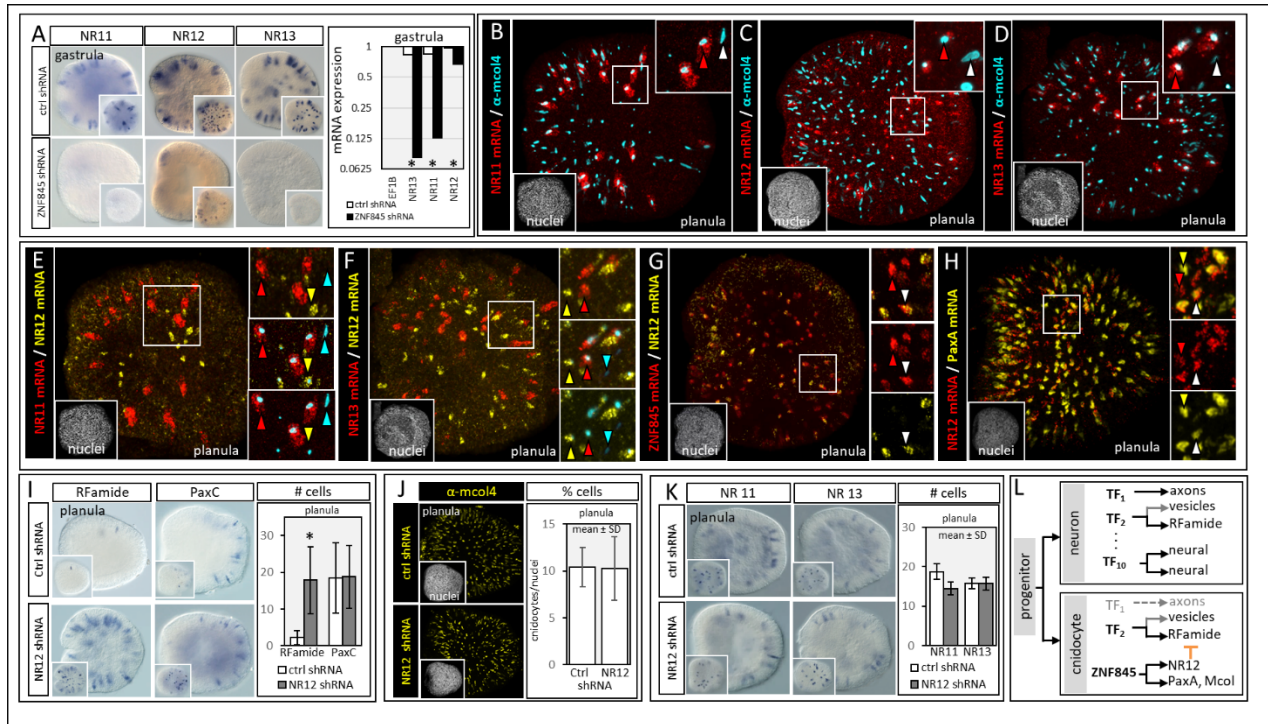
Fig. 2. RFamide identity is inhibited by NR12 in developing cnidocytes. (A) NR11, NR12, and NR13 are downregulated in ZNF845 knockdowns (shRNA) as assayed by *in situ* hybridization and qPCR. Insets: surface detail. (B-D) NR11, NR12, and NR13 are co-expressed with Mcol4 (cnidocytes), but not with each other (E,F). Images B and E show the same embryo, as do D and F. NR12 is partially co-expressed with (G) ZNF845 and (H) PaxA. (I) Response of RFamide- and PaxC-expressing cells after NR12 knockdown. No effect of NR12 knockdown on (J) cnidocyte differentiation (α-Mcol4) or (K) expression of NR11 or NR13. (L) Modular regulation of neural traits (e.g., axons, vesicles + neuropeptides) explains how some traits can be lost (e.g., axons) and others retained but inhibited (RFamide) in developing cnidocytes. Significance (*) for all tests indicated as p<1E-02. See Table S2 for supporting information.
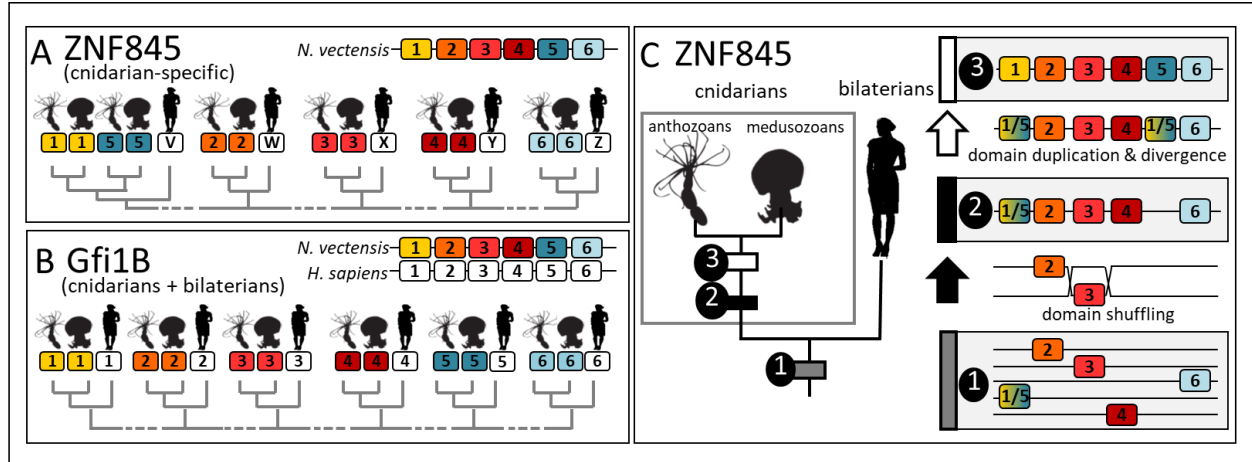
Fig. 3. Evolutionary history of two 6-domain ZNF proteins. Relationships among ZF-C_2H_2 domains from (A) ZNF845 (*N. vectensis* JGI PID 81344) and (B) Gfi1B (*N. vectensis* JGI PID: 112378). (C) Proposed model for the emergence of ZNF845: 1. In the common ancestor of cnidarians and bilaterians ZF-C_2H_2 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. Broken grey lines in A, B represent the locations of the other ~11,000 branch tips of the complete tree. See Supplementary Figs S5,S6 for supporting information.
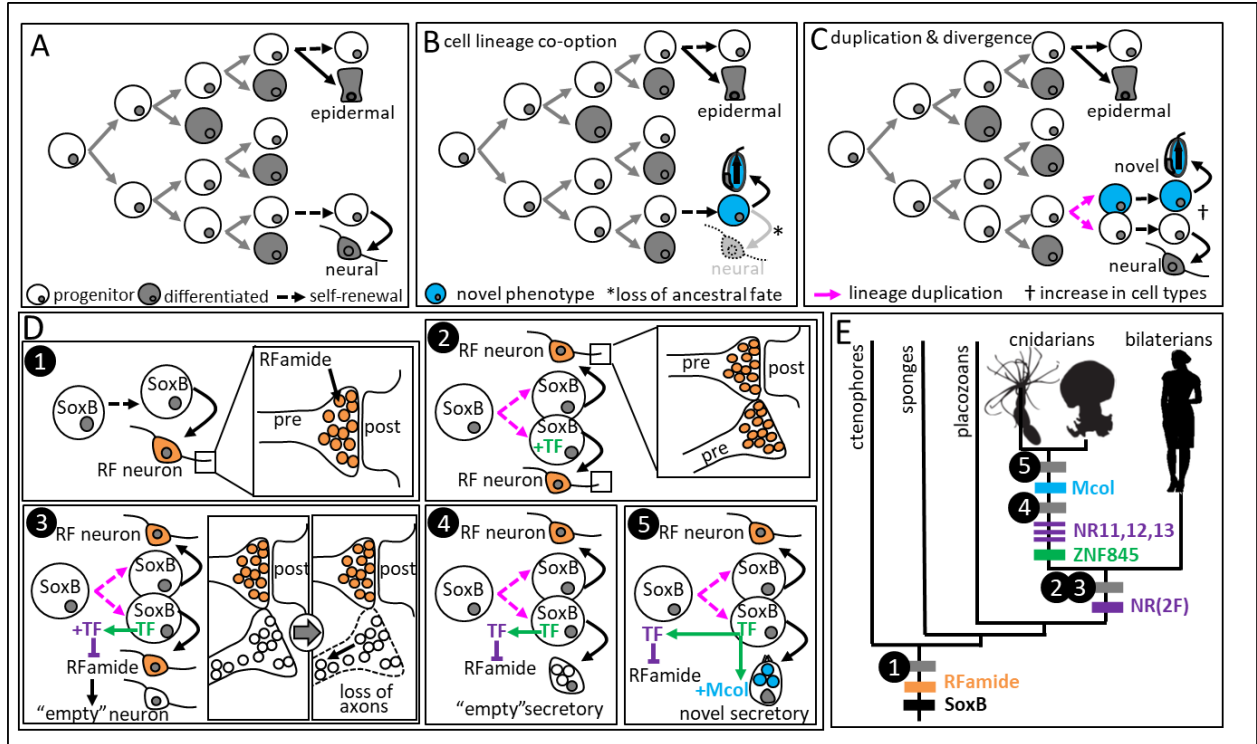
Fig. 4. Origin of novelty through lineage duplication and suppression of ancestral fate.
(A) Embryogenesis as a model for cell type evolution. (B) Co-option of a progenitor cell for novel fate does not increase cell type diversity. (C) Duplication of the progenitor cell prior to co-option allows for the expansion of differentiated cell types. (D) A stepwise model for the origin of cnidocytes by duplication and divergence of an RFamide neural lineage. (E) A proposal for the evolutionary timing of these events; similar duplication of a neural progenitor cell and inhibition of the neural payload (steps 2,3) may also have facilitated the expansion of neural subtypes in early bilaterians.