Figures and figure supplements

Large-scale deorphanization of *Nematostella vectensis* neuropeptide G protein-coupled receptors supports the independent expansion of bilaterian and cnidarian peptidergic systems

Daniel Thiel and Luis Alfonso Yañez Guerra *et al*.

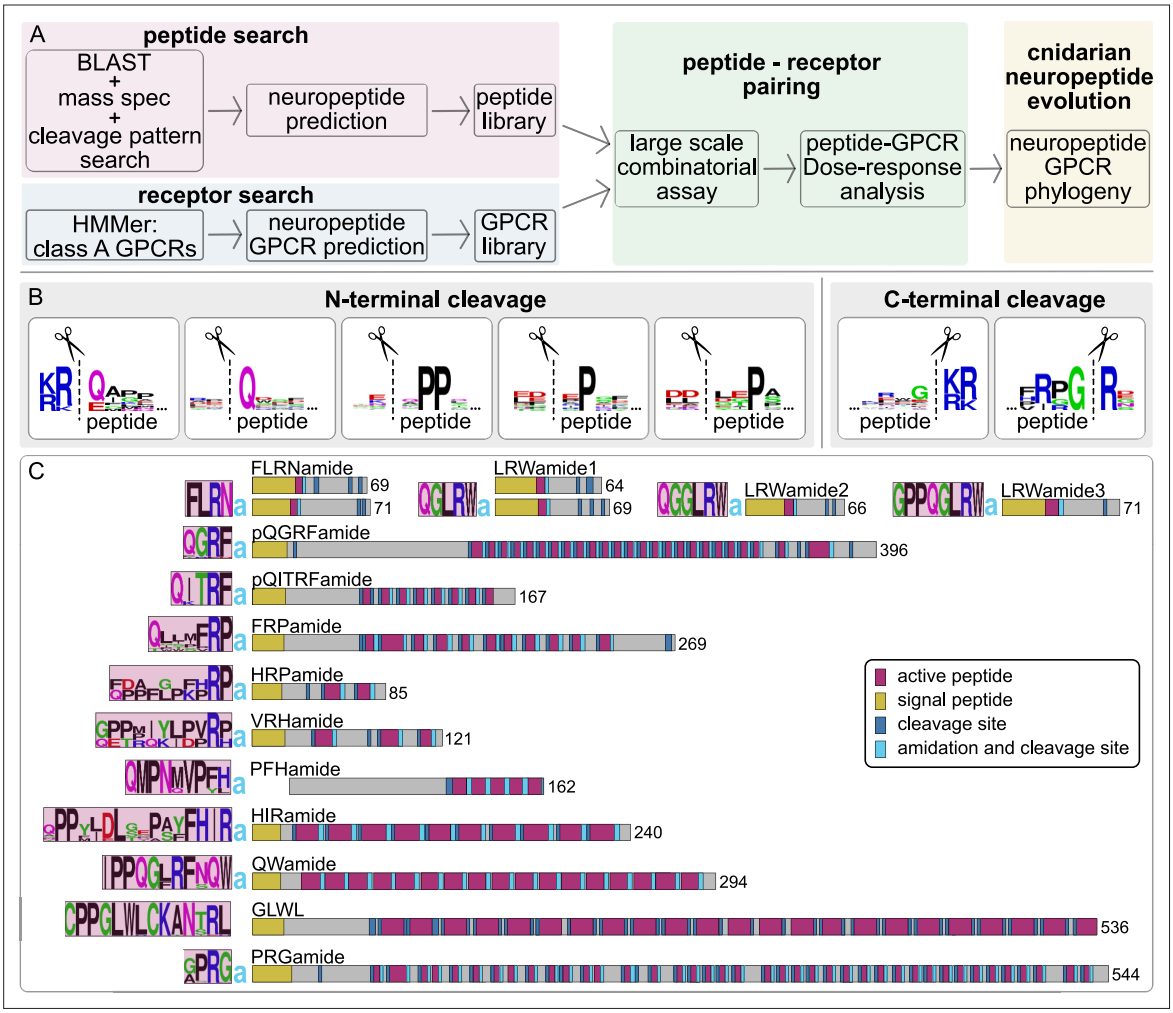


Figure 1. Identification of *N. vectensis* neuropeptides. (**A**) Pipeline to identify neuropeptides and their receptors and to reconstruct the evolution of cnidarian peptidergic signaling. (**B**) Peptide sequence logos of N-terminal and C-terminal peptide cleavage sites based on peptides detected by LC-MS/MS. Cleavage occurs at the dashed lines. (**C**) *N. vectensis* neuropeptide precursor schemes of peptides for which we identified a receptor, with sequence logos of the encoded peptide(s) on the left and length of precursor on the right. a=amide.

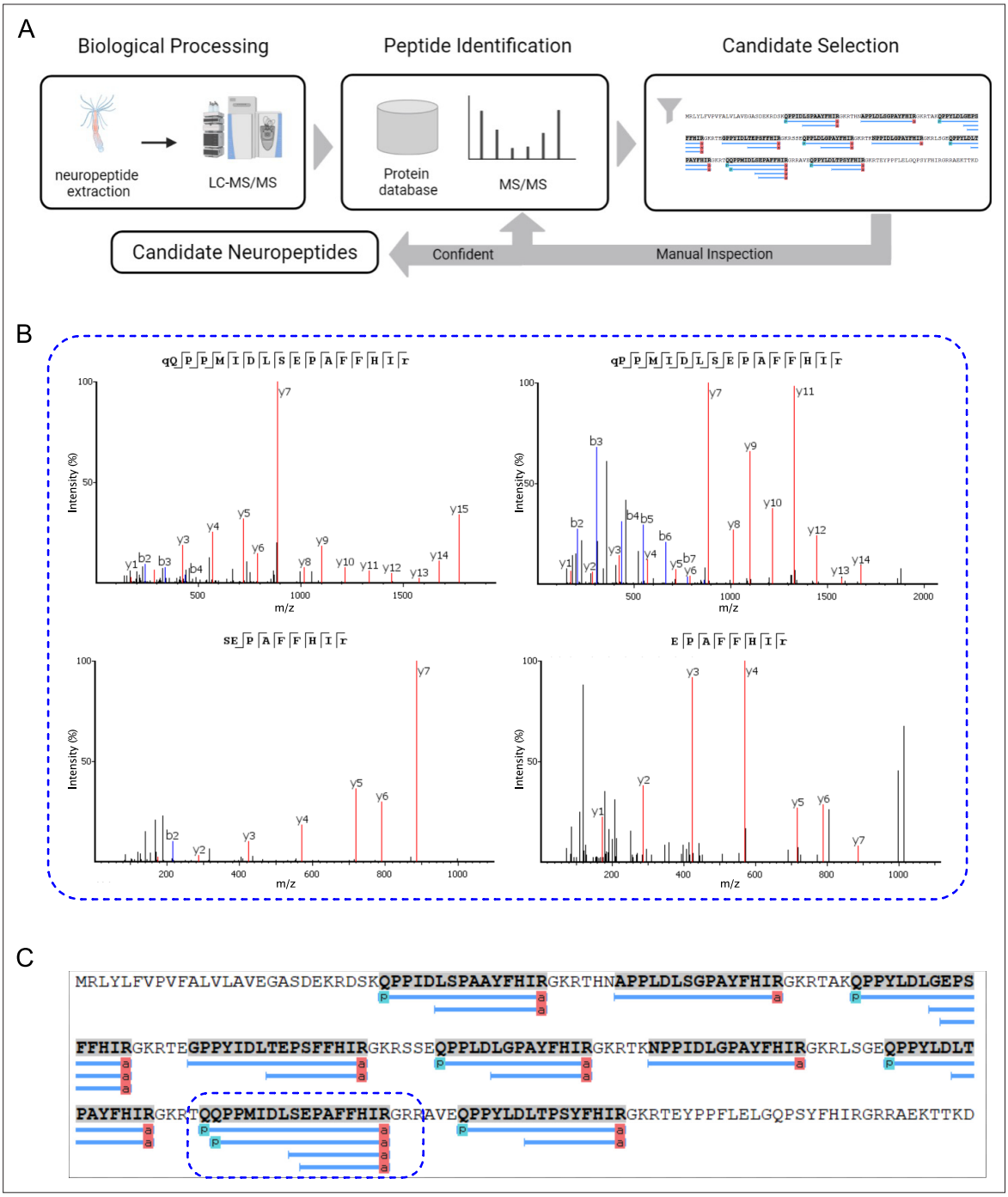


Figure 1—figure supplement 1. Mass spectrometry pipeline. (**A**) Pipeline for mass spectrometry identification of neuropeptide candidates. Created with BioRender.com. (**B**) Example spectra of detected HIRamide peptides with different lengths that originate from the same peptide on the HIRamide precursor. The shown spectra are from the encircled HIRamide peptide copy in (**C**). Figures within Panel B exported from PEAKS Studio X+ (v10.5 Build number 20200219, ***Bioinformatics Solutions, 2023***).

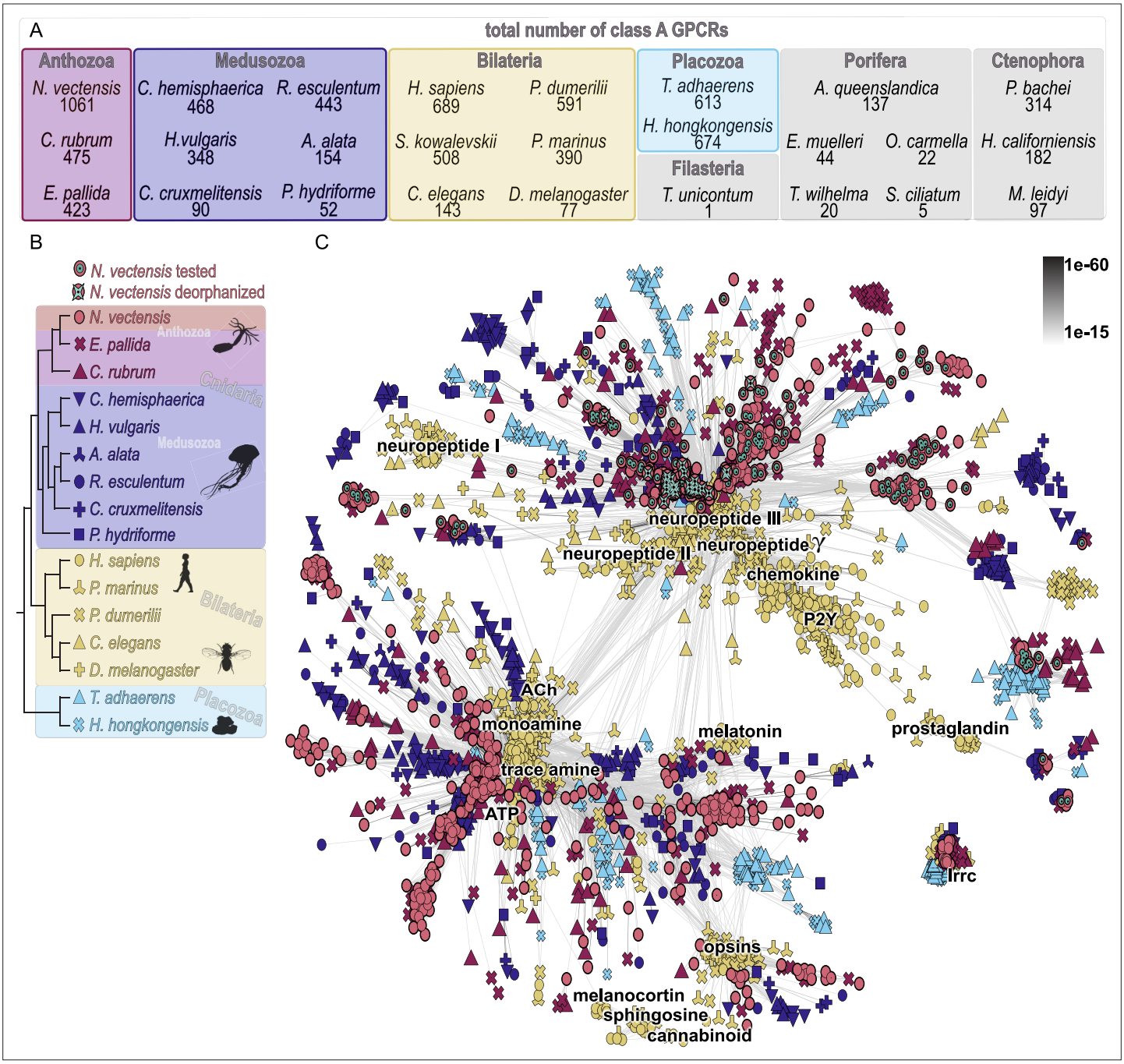


Figure 2. Cluster map of selected class A G protein-coupled receptors (GPCRs). (**A**) Number of class A GPCRs identified by HMMer search in the different investigated species. (**B**) Relationship of species used for cluster analysis in C. (**C**) Cluster analysis of major class A GPCR groups from cnidarian, bilaterian, and placozoan species. Each dot represents a GPCR sequence with color-coding and symbols according to the phylogeny in B. Connecting lines between single sequences show similarity with p-values indicated in the top right. Cluster annotations are based on deorphanized bilaterian class A GPCRs. Abbreviations in C: ACh = acetylcholine, lrrc = leucine-rich-repeat containing, P2Y=purinergic P2Y receptor. Silhouette images in B were taken from phylopic.org.

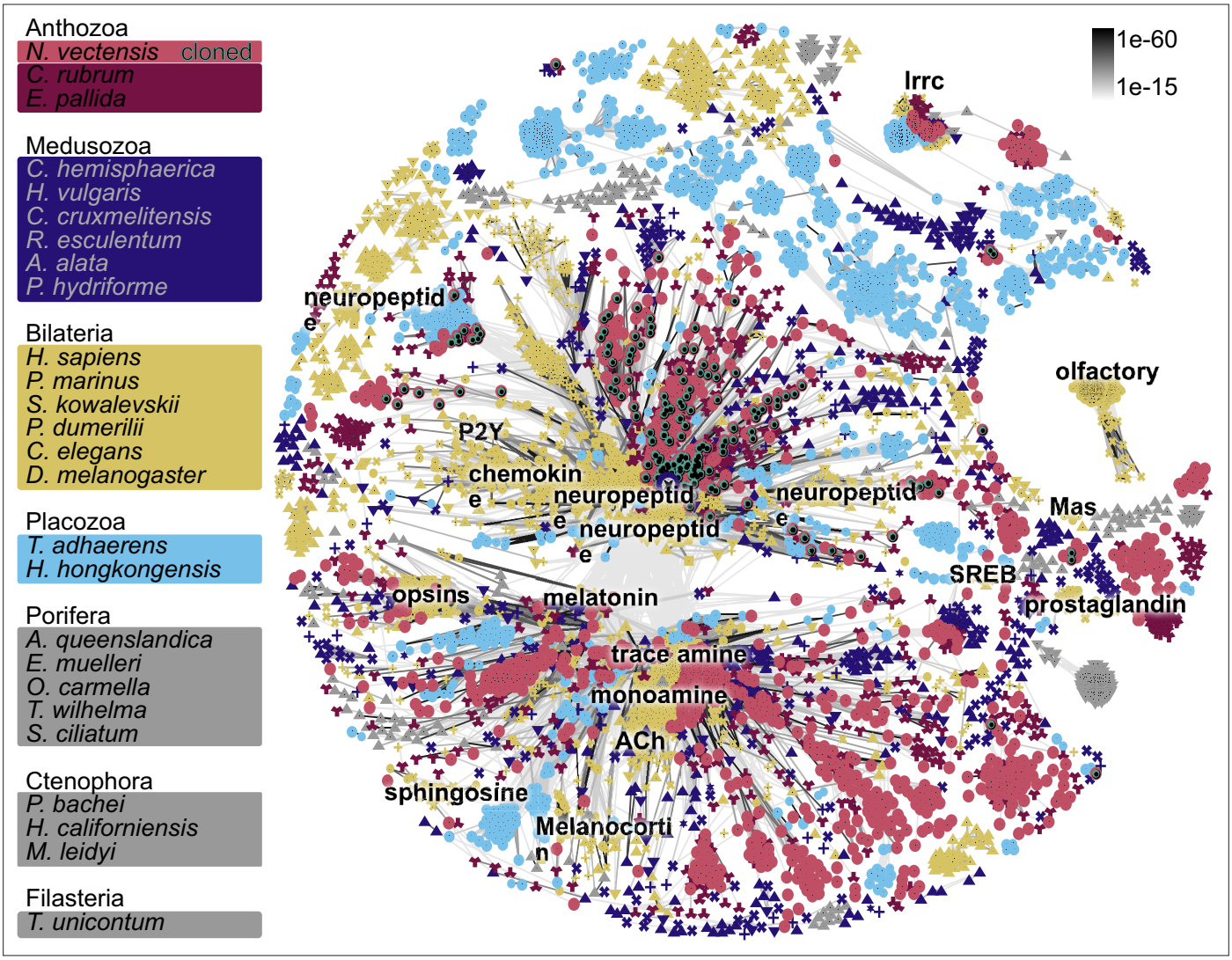


Figure 2—figure supplement 1. Cluster map of metazoan class A G protein-coupled receptors (GPCRs). Each dot represents a GPCR sequence with color-coding according to species list on the left. Connecting lines between single sequences show similarity with p-values indicated in top right. Annotations are based on sequences of deorphanized bilaterian class A GPCRs. Abbreviations in C: ACh = acetylcholine, lrrc = leucine-rich-repeat containing, P2Y = purinergic P2Y receptor.

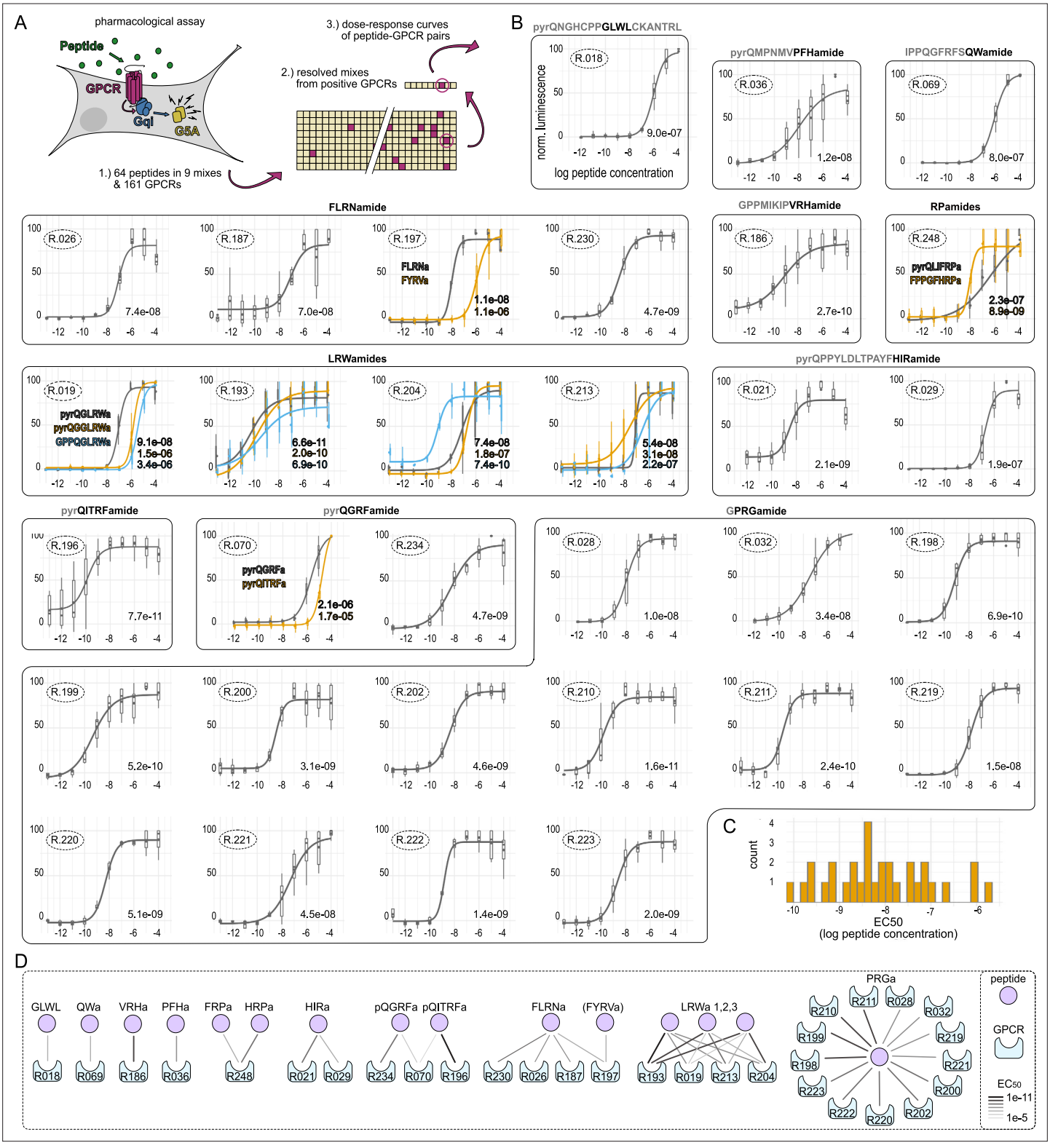


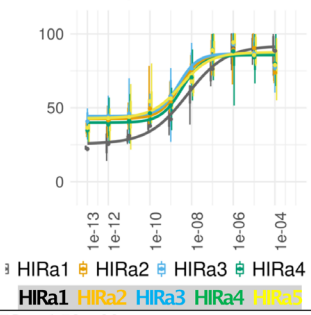
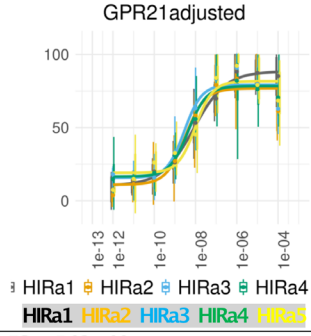
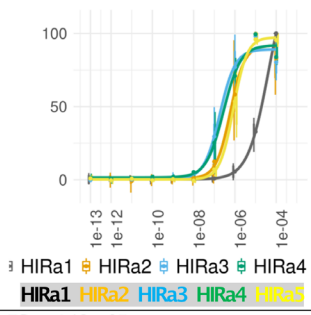
Figure 3. Dose-response curves of *Nematostella* neuropeptide G protein-coupled receptor (GPCR) pairs. (**A**) Pharmacological assay and pipeline to identify peptide-GPCR pairs. (**B**) Dose-response curves of peptide-GPCR pairs with log peptide concentration plotted against normalized luminescence. GPCRs that are activated by the same peptide(s) are grouped together with peptide sequence shown above and peptide name highlighted in black. If several peptides activate the same receptor, peptide sequences are shown within the graph. Receptor identification number is encircled in the upper left of each curve, EC_{50} values are indicated in the lower right. Sample size per datapoint = 9. Error bars show distribution of datapoints with box

Figure 3 continued on next page

Figure 3 continued

indicating upper and lower quartile. (**C**) Histogram of EC_{50} values of peptide-GPCR pairs, showing only the lowest EC_{50} per GPCR. (**D**) Peptide-receptor pairings showing number of receptors activated by the different peptides. Connection strength indicates EC_{50} values.

A

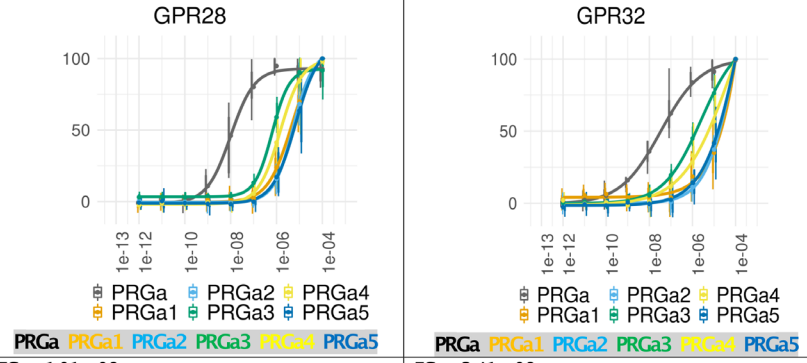
	GPR21 (original readings)	GPR21 (1e−13 set to zero)	GPR29
			
HIRa1 (pyro–QPPYLDLGEPSFFHIR–amide)	EC_{50}=1.59e–09	EC_{50}=5.93e–09	EC_{50}=4.12e–05
HIRa2 (YPPFLELGQPSYFHIR–amide)	EC_{50}=1.15e–11	EC_{50}=3.1e–09	EC_{50}=5.96e–07
HIRa3 (pyro–QPPYLDLTPAYFHIR–amide)	EC_{50}=7.64e–12	EC_{50}=2.07e–09	EC_{50}=1.86e–07
HIRa4 (pyro–QPPLDLGPAYFHIR–amide)	EC_{50}=3.37e–11	EC_{50}=3.43e–09	EC_{50}=2.78e–07
HIRa5 (NPPIDLGPAYFHIR–amide)	EC_{50}=1.08e–11	EC_{50}=2.96e–09	EC_{50}=8.48e–07

B

MRLYLFVPVFALVLAVEGASDEKRDSKQPPIDLSPAAYFHIRGKRTYNAPPLDLSGPAYFHIRGKRTAKQPPYLDLGEPSFFHIRGKRTEGPPYIDLTEPSFFHIRGKRSSEQPPLDLGPAYFHIRGK
RTKNPPIDLGPAYFHIRGKRLSGEQPPYLDLTPAYFHIRGKRTQQPPMIDLSEPAFFHIRGRRAVEQPPYLDLTPSYFHIRGKRTEYPPFLELGQPSYFHIRGRRAEKTTKD

Figure 3—figure supplement 1. Dose-response curves with EC_{50} values and peptide precursor of different HIRamide peptides. (**A**) Dose-response curves of HIRamide peptide and GPR 21 + 29. Dose-response curves show activation of receptors GPR21 and GPR29 with different HIRamide peptide versions from the same precursor. Sequences and EC_{50} values for different HIRamides are shown below the dose-response curves. The graph in ***Figure 3*** shows HIRamide3 as the peptide with the lowest EC_{50} values. GPR21 showed already at low concentrations a high base activation compared to the negative control, but did not show a clear increase in intensity at concentrations between 1E-13 and 1E-11. We therefore set the minimum to the value measured at a concentration of 1E-13 instead of the negative control and show a comparison of the corresponding graphs and EC_{50} values here. (**B**) HIRamide precursor. The N-terminal signal peptide of the precursors sequence is not highlighted, cleavage + amidation sites are highlighted in a darker gray. Tested peptides are underlined and colored as shown underneath the dose-response curves.

A

	GPR28	GPR32
		
PRGa (GPRG-amide)	EC_{50}=1.01e-08	EC_{50}=3.41e-08
PRGa1 (APRGGRSLLDLLAAPRG-amide)	EC_{50}=4.69e-06	EC_{50}=7.99e-05
PRGa2 (GPRGGREVNLEGPRG-amide)	EC_{50}=5.34e-06	EC_{50}=9.51e-05
PRGa3 (GPRGGRATEFGPRG-amide)	EC_{50}=6.29e-07	EC_{50}=2.49e-06
PRGa4 (GPRGGRSYDGPRG-amide)	EC_{50}=1.43e-06	EC_{50}=2.15e-05
PRGa5 (APRGGRDAPRG-amide)	EC_{50}=6.69e-06	EC_{50}=1.34e-04

B

MAYFKWTLCAFVLAVLCLSTNYVAGEKNEEKSEKSTLLQELSKKDFDKELAADEAIVLELLKEKADDLDEESLLDLLAPRGGRDAPRGGRSLLDAPRGGRSLLDLLAAPRGGRDAPRGGRSLTELLNA
PRRGRSLADLLDAPRGGRSLAELLDAPRGGRSIEAPRGGRSLAELLEAPRGGRSLIELLEAPRGGRSVEAPRGGRSILELLTAPRGGRSAPRGGRSAEKSDAIHKEKKAPRGGRKRRSLPEEEMDGPR
GGRSAVSGRSSEDGPRGGRAVYGPRGGRSFDGSRGGRSYDGPRGGRSMENGPRGGRATEFGPRGGRSYEGSRGGREIDGPRGGRSFKEGPRGGRAVEGPRGGRDLYEEGPRGGREVYLEGPRGGRDLV
LEGPRGGRDLYEEGPRGGREVNLEGPRGGRDLYEEGPRGGRSYDGPRGGRSAVYGPRGGRDANDGPRGGRDANDGPRGGRDANEGPRGGRDANEGPRGGRDANKGPRGGRNTNTVGAQ
QADARQGQRGGRSTGMGPRGGRSVSSEKDNRH

Figure 3—figure supplement 2. Dose-response curves with EC_{50} values and peptide precursor of different PRGamide peptide versions for two of the PRGamide receptors. (**A**) Dose-response curves show activation of the receptors GPR28 and GPR32 with different PRGamide peptide versions from the same precursor. Sequences and EC_{50} values for different PRGamides are shown below the dose-response curves. The graph in ***Figure 3*** shows the shortest PRGamide (GPRGamide) as the peptide with the lowest EC_{50} values, which is likely the fully processed version. (**B**) PRGamide precursor. The N-terminal signal peptide of the precursor sequence is not highlighted, cleavage and amidation sites are highlighted in a darker gray. The tested peptides are colored to match the color of the dose-response curves and the GPRG peptide is underlined.

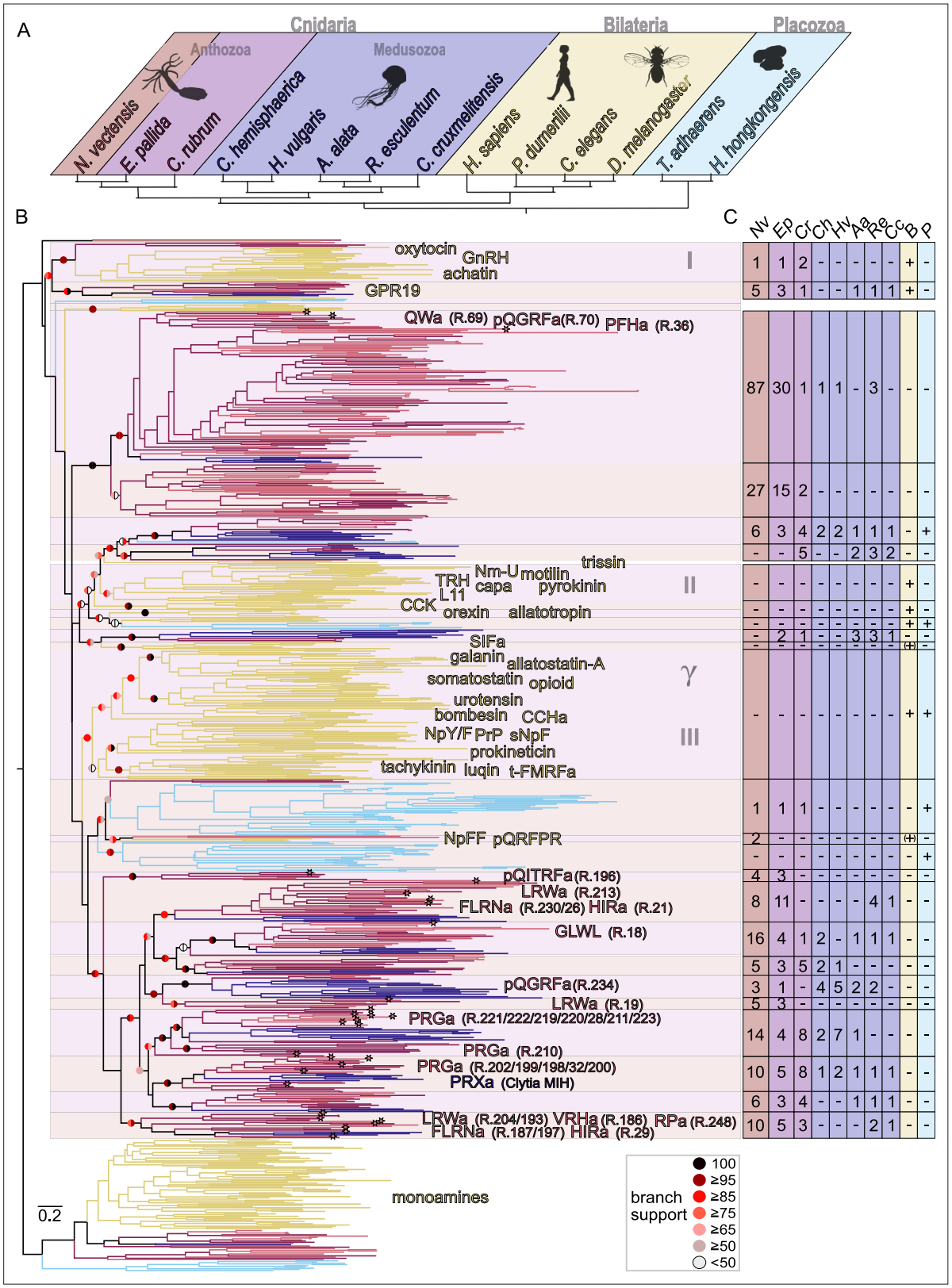


Figure 4. Phylogeny of metazoan class A neuropeptide G protein-coupled receptors (GPCRs). (**A**) Phylogeny of species used in B. (**B**) Phylogeny of neuropeptide GPCRs with names of ligands. Branches are color-coded according to A. Branches of deorphanized *Nematostella* GPCRs end in an asterisk. Alternating shades behind the tree branches highlight different monophyletic groups. Roman numbers 1–3 and Greek symbol gamma indicate approximate neuropeptide clusters shown in ***Figure 2***. Left half circle of branch support indicates aBayes and the right half circle aLRT-SH-like support

Figure 4 continued on next page

Figure 4 continued

values. Detailed annotations in ***Supplementary file 11***. (**C**) Table with number of receptors per group as highlighted in receptor phylogeny with a straight line indicating no receptor present. Two-letter abbreviations on top correspond to species in A. Abbreviations: a=amide, B=Bilateria, CCK = cholecystokinin, GnRH = gonadotropin releasing hormone, MIH = maturation-inducing hormone, Nm-U=neuromedin U, NpFF = neuropeptide FF, NpY/F=neuropeptide Y/neuropeptide F, P=Placozoa, PrP = prolactin releasing peptide, R.#=*Nematostella* GPCR number, sNpF = short neuropeptide F, t-FMRFa=trochozoan FMRFamide, TRH = thyrotropin releasing hormone.

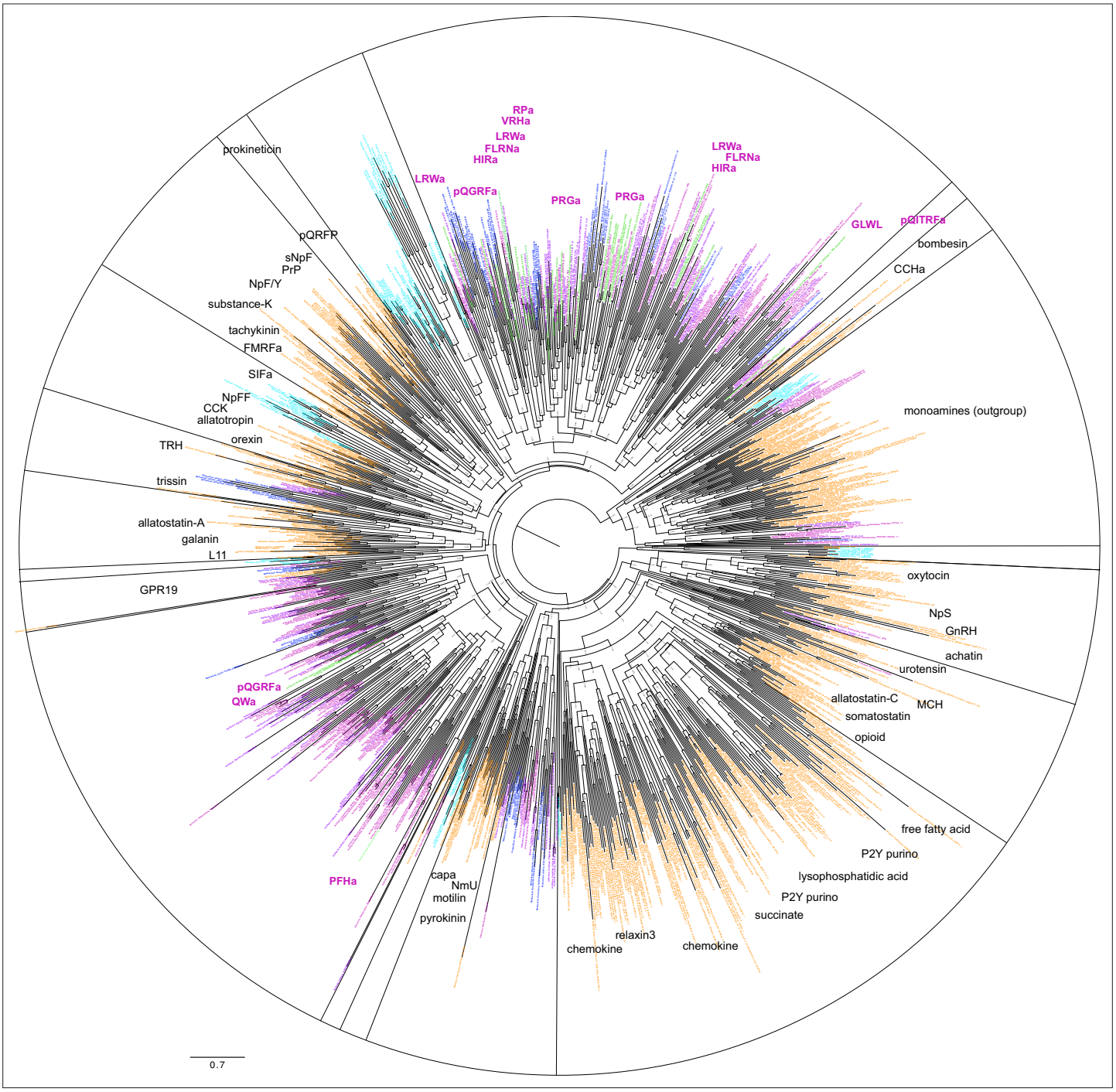


Figure 4—figure supplement 1. Tree (FastTree) of neuropeptide G protein-coupled receptors (GPCRs) with bilaterian chemokine and related receptors. Bilaterian sequences are shown in yellow, cnidarian sequences are shown in dark blue (Medusozoa) and magenta (Anthozoa), placozoan sequences are shown in light blue. Deorphanized *N. vectensis* sequences are shown in green. Sequences were aligned using muscle, alignment was trimmed with the gappyout option of trimal, tree was calculated using FastTree. Abbreviations: a = amide, CCK = cholecystokinin, GnRH = gonadotropn releasing hormone, GPR19 = G protein-coupled receptor 19, L11 = elevenin, MCH = melanin concentrating hormone, Nm-U = neuromedin U, NpFF = neuropeptide FF, NpS = neuropeptide S, NpY/F = neuropeptide Y/neuropeptide F, PrP = prolactin releasing peptide, sNpF = short neuropeptide F, TRH = thyrotropin releasing hormone.

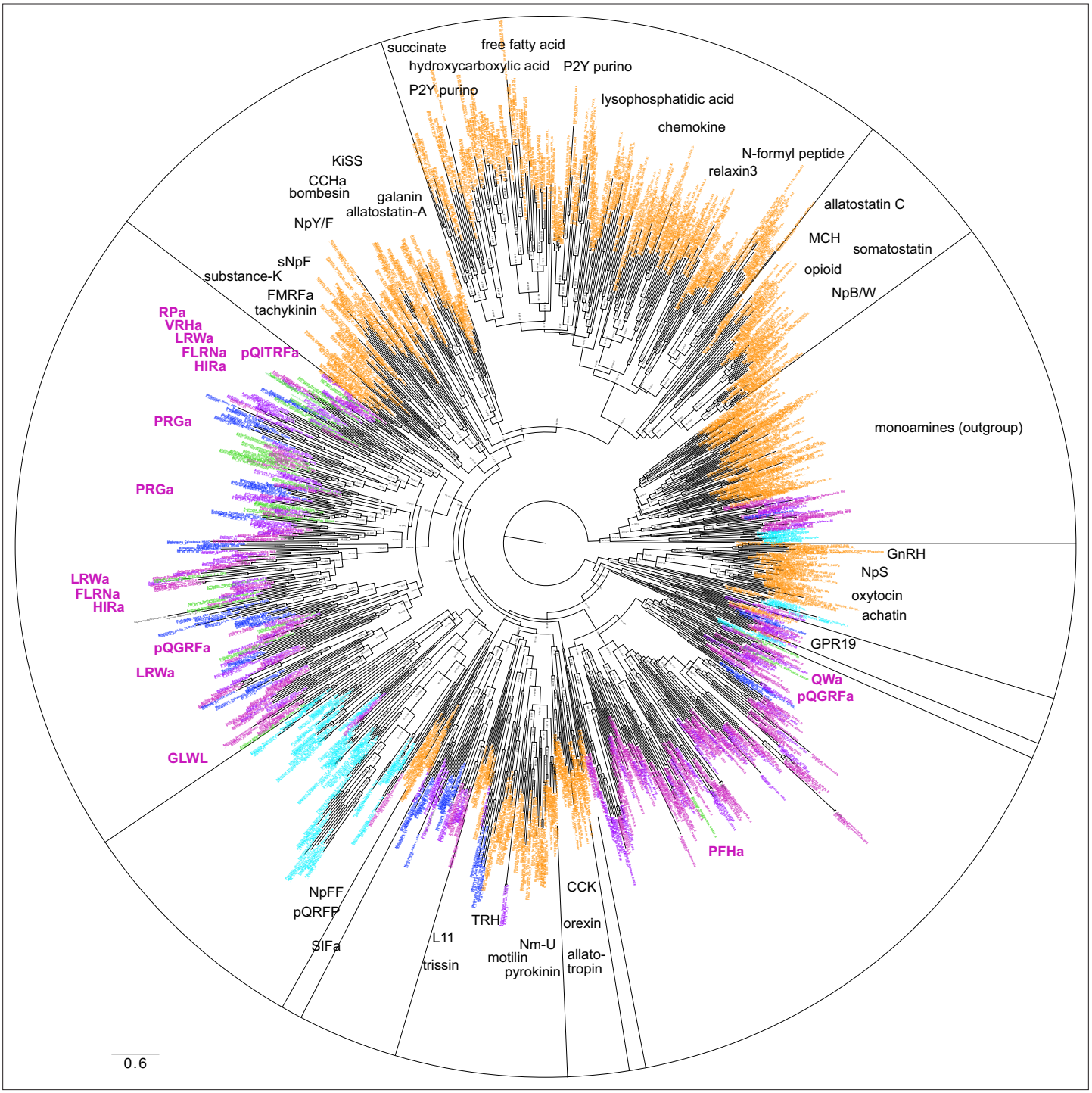


Figure 4—figure supplement 2. Tree (IQtree) of neuropeptide G protein-coupled receptors (GPCRs) with bilaterian chemokine and related receptors. Bilaterian sequences are shown in yellow, cnidarian sequences are shown in dark blue (Medusozoa) and magenta (Anthozoa), placozoan sequences are shown in light blue. Deorphanized *N. vectensis* sequences are shown in green. Sequences were aligned using mafft, alignment was trimmed with the gappyout option of trimal, tree was calculated using IQtree. Abbreviations: a = amide, CCK = cholecystokinin, GnRH = gonadotropin releasing hormone, GPR19 = G protein-coupled receptor 19, L11 = elevenin, MCH = melanin concentrating hormone, Nm-U = neuromedin U, NpB/W = neuropeptide B/neuropeptide W, NpFF = neuropeptide FF, NpS = neuropeptide S, NpY/F = neuropeptide Y/neuropeptide F, sNpF = short neuropeptide F.

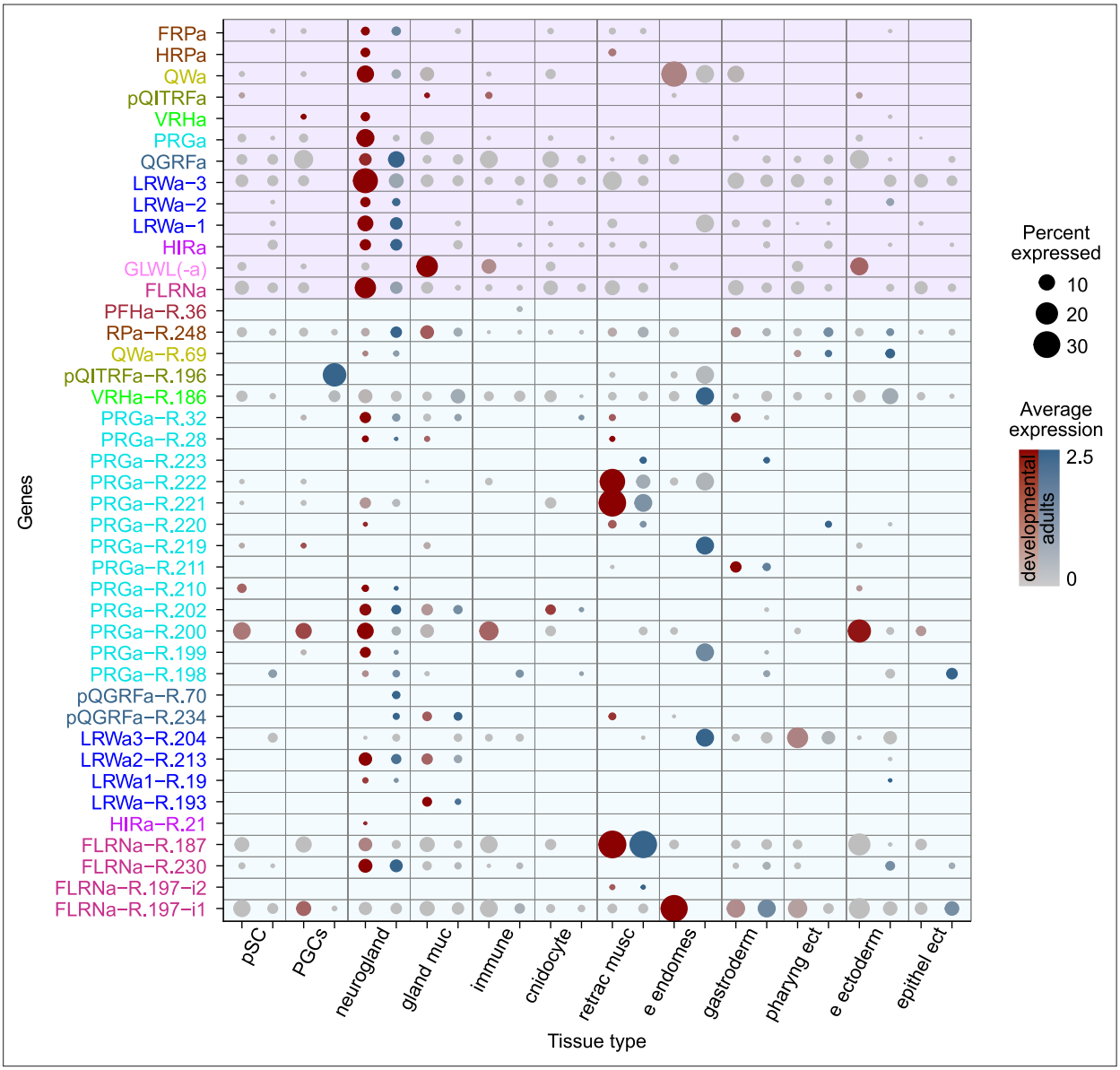


Figure 5. Tissue-specific expression of neuropeptide precursors and receptors (G protein-coupled receptors [GPCRs]) in *N. vectensis*. Dotplot for tissue-specific expression of peptide precursors and GPCRs. Red dots indicate expression in the developmental dataset, blue dots indicate expression in the adult dataset. Abbreviations: a=amide, e=embryonic, ect = ectoderm, endomes = endomesoderm, gland = glandular, muc = mucous, musc = muscle, neurogland = neuroglandular, PGCs = primary germ cells, pharyng = pharyngeal, pSC = putative stem cells, R=receptor (GPCR), retrac = retractor.

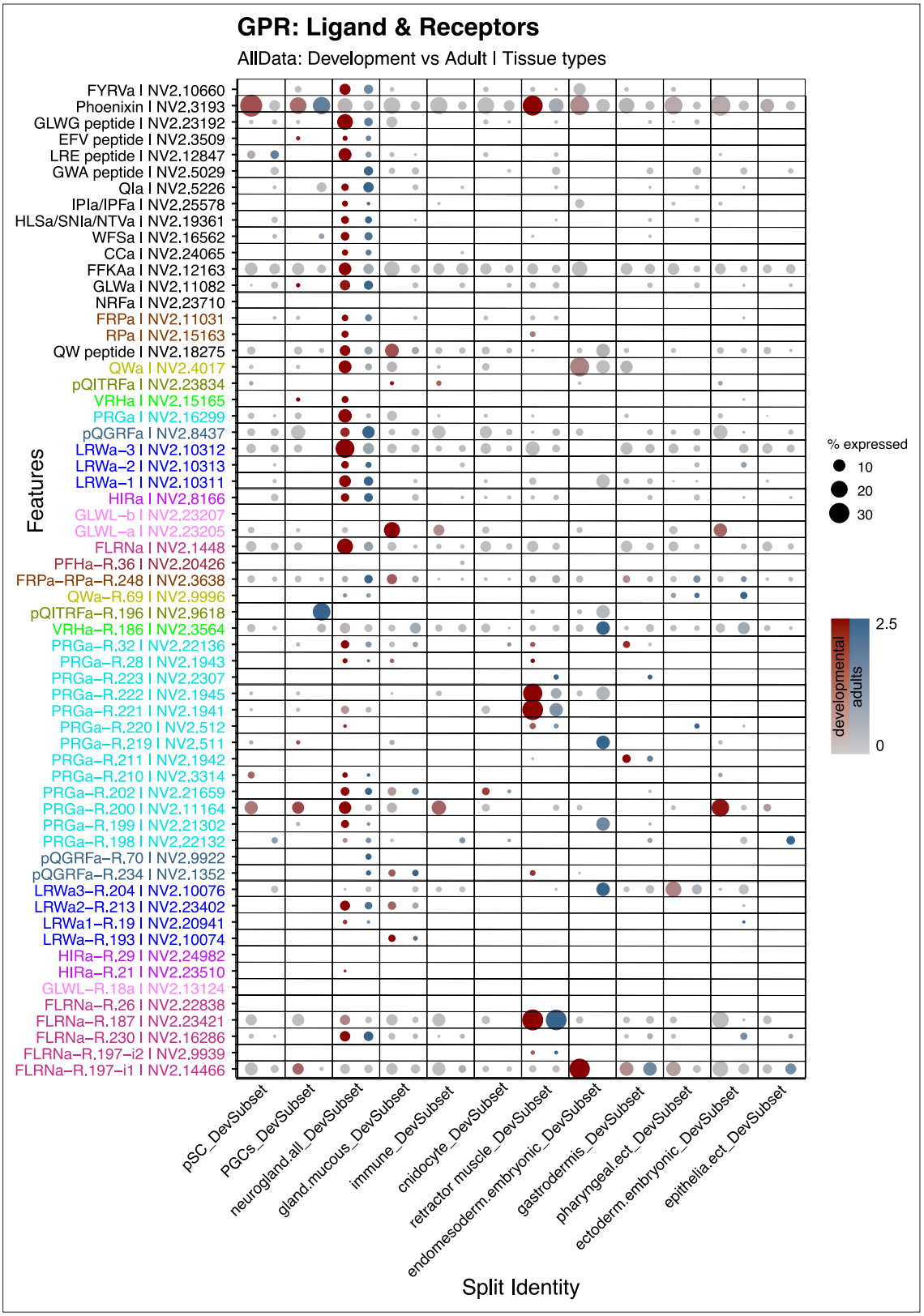


Figure 5—figure supplement 1. Tissue-specific expression of neuropeptide precursors and neuropeptide receptors (G protein-coupled receptors [GPCRs]) in *N. vectensis*. Dotplot for tissue-specific expression of proneuropeptides and GPCRs. Proneuropeptides without a known receptor are also included. Red dots indicate expression in the developmental dataset, blue dots indicate expression in the adult dataset.

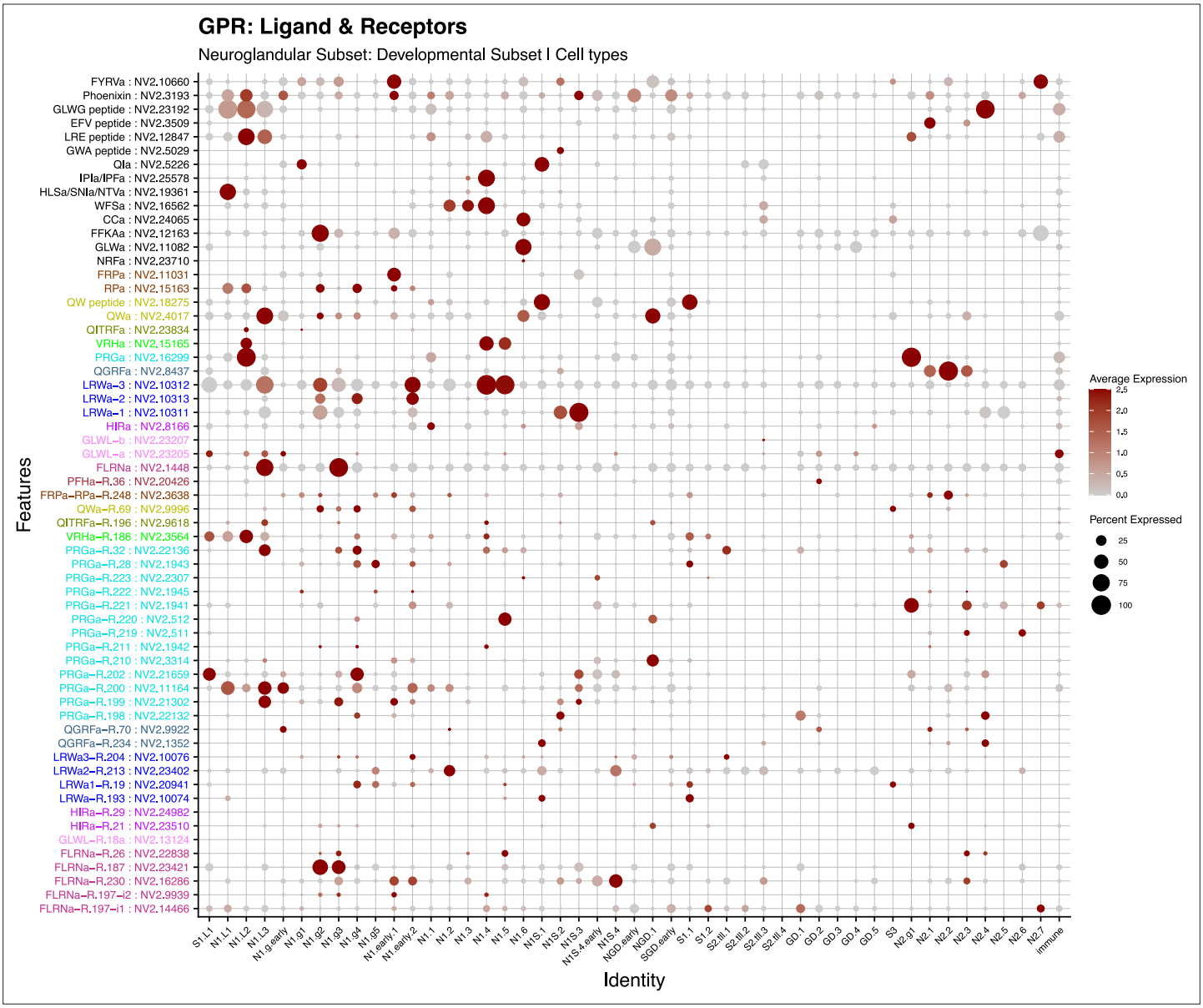


Figure 5—figure supplement 2. Expression of neuropeptide precursors and G protein-coupled receptors (GPCRs) in neuroglandular cell types in the developmental dataset. Dot size indicates percentage of cells expressing the corresponding gene, color intensity indicates average expression.

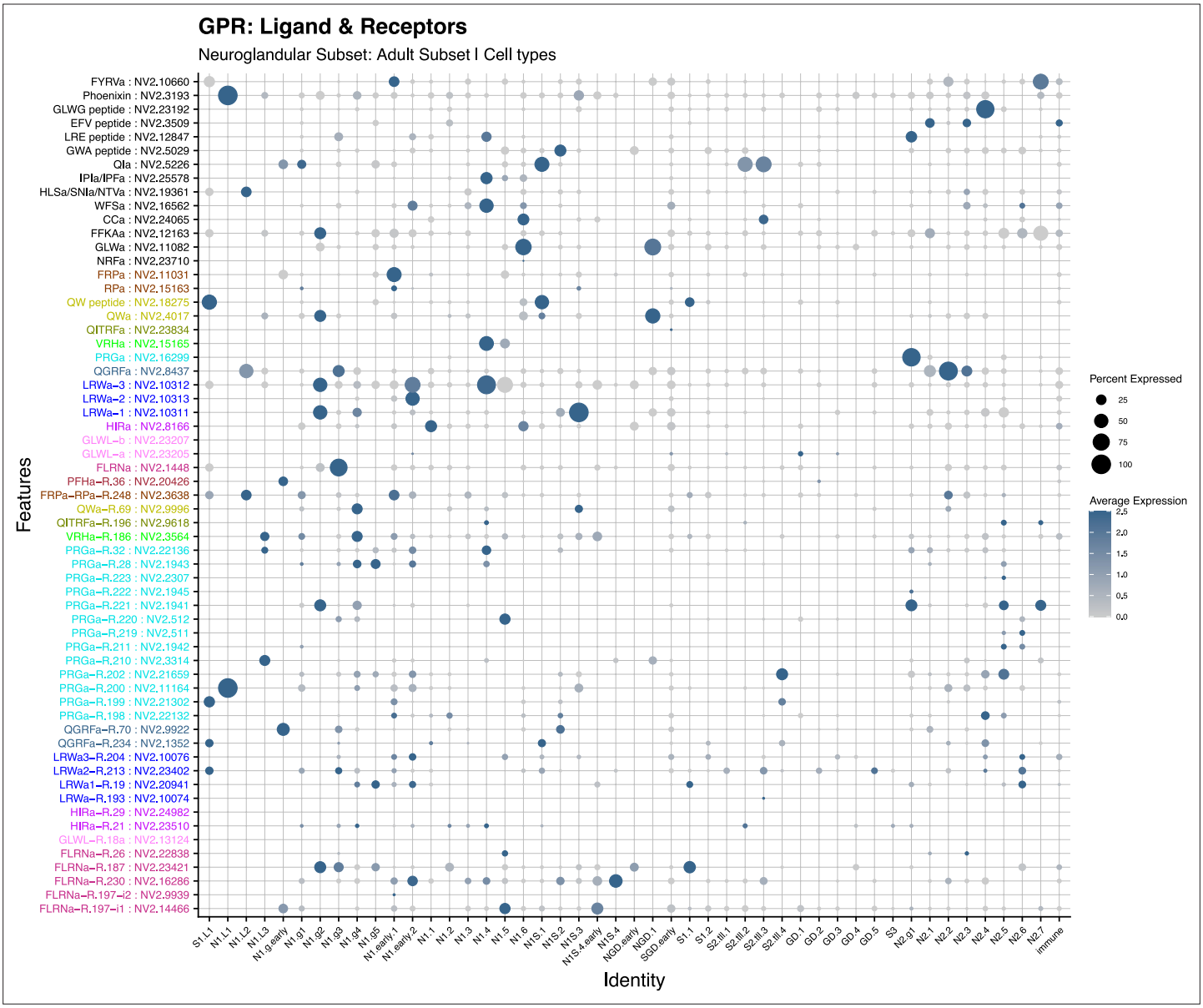


Figure 5—figure supplement 3. Expression of neuropeptide precursors and G protein-coupled receptors (GPCRs) in neuroglandular cell types in the adult dataset. Dot size indicates percentage of cells expressing the corresponding gene, color intensity indicates average expression.

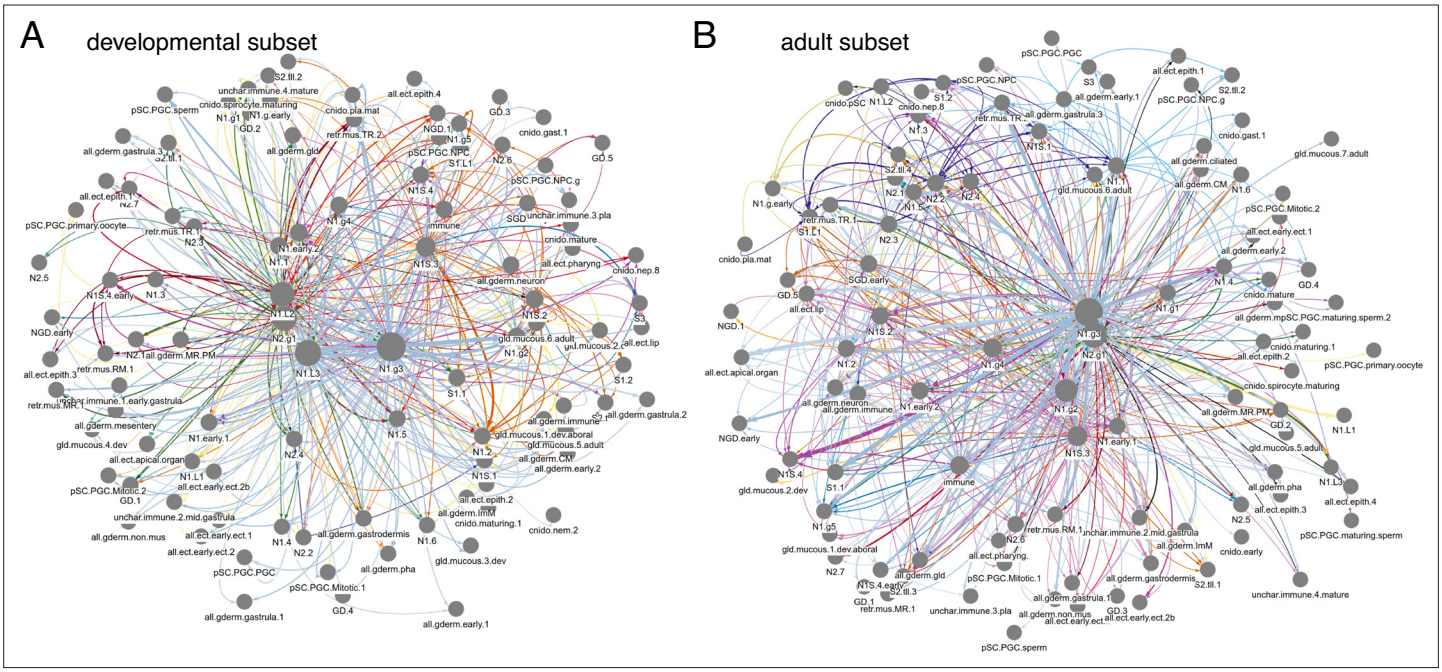


Figure 6. Multilayer peptidergic connectomes in *Nematostella*. Peptidergic networks in the (**A**) developmental and (**B**) adult subset. Nodes represent cell types, connections represent potential peptidergic signaling from neuropeptide-expressing cells to cells expressing one or more of the receptors for that neuropeptide. Colors represent different peptide-receptor signal channels (the different layers in the multilayer connectome).

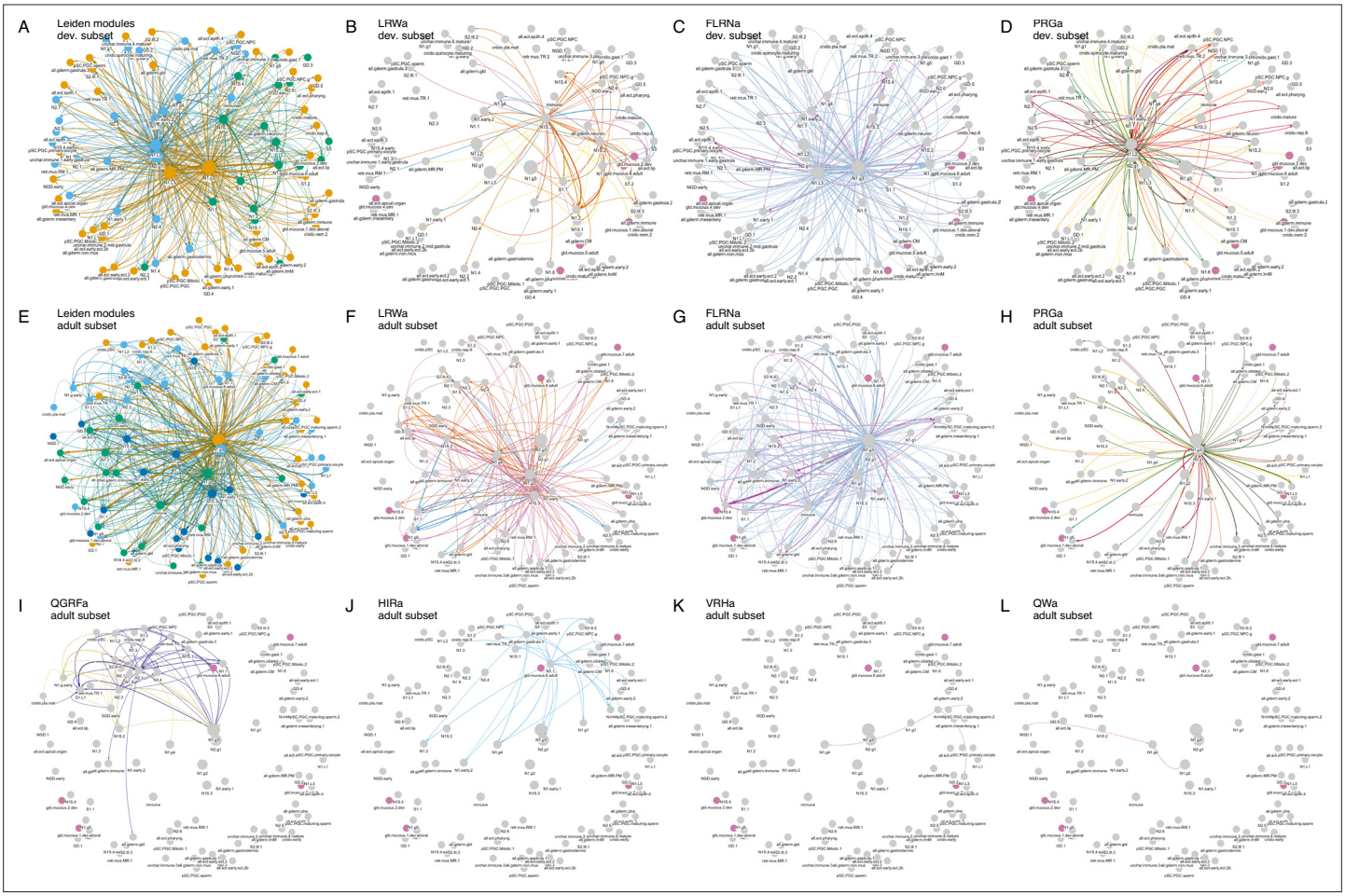


Figure 6—figure supplement 1. Multilayer peptidergic connectomes in *Nematostella*. (**A**) Network of all peptide-receptor pairs for the developmental subset, colored by network Leiden module. (**B–D**) Networks of LRWa (**B**), FLRNa (**C**), and PRGa (**D**) for the developmental subset. (**E**) Network of all peptide-receptor pairs for the adult subset, colored by network Leiden module. (**F–L**) Networks of LRWa (**F**), FLRNa (**G**), PRGa (**H**), QGRFa (**I**), HIRa (**J**), VRHa (**K**), and QWa (**L**) for the adult subset. Nodes represent cell types, connections represent potential peptidergic signaling from neuropeptide-expressing cells to cells expressing one or more of the receptors for that neuropeptide. For peptides with more receptors, different colors represent different peptide-receptor signal channels (the different layers in the multilayer connectome).