PNAS
www.pnas.org

Supplementary Information for

Prototypical pacemaker neurons interact with the resident microbiota

Alexander Klimovich[a,1], Stefania Giacomello[b,c], Åsa Björklund[d], Louis Faure[e], Marketa Kaucka[e,f,g], Christoph Giez[a], Andrea P. Murillo-Rincon[a], Ann-Sophie Matt[a], Doris Willoweit-Ohl[a], Gabriele Crupi[a], Jaime de Anda[h], Gerard C.L. Wong[h], Mauro D'Amato[i], Igor Adameyko[e,f], Thomas C.G. Bosch[a,1,2]

[1]Correspondence:

Thomas C.G. Bosch: tbosch@zoologie.uni-kiel.de;

Alexander Klimovich: aklimovich@zoologie.uni-kiel.de

[2] Lead contact:

Thomas C.G. Bosch: tbosch@zoologie.uni-kiel.de

This PDF file includes:

Supplementary Methods
Figures S1 to S21
Table S1
Legends for Datasets S1 to S9
Supplementary References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S9

www.pnas.org/cgi/doi/10.1073/pnas.1920469117

Supplementary Methods

FACS isolation of cells

To isolate cells of the interstitial stem cell lineage from the transgenic *Hydra* by FACS (**Fig. S1*A***), the polyps were disintegrated into a cell suspension as previously described (1). Briefly, 100 hydras were treated for 1.5 h in 1 ml dissociation medium (KCl 3.6 mM, $CaCl_2$ 6.0 mM, $MgSO_4$ 1.2 mM, Na citrate 6.0 mM, Na pyruvate 6.0 mM, TES 12.54 mM, glucose 6.0 mM, pH 6.9) supplemented with 50 U/mL Pronase E (Serva, Cat. No. 33635) on an orbital shaker at 200 rpm at 18ºC. The resulting cell suspension was filtered through a 100 µm mesh and centrifuged for 5 min at 1,000 g at 4°C, the supernatant was removed, and the cells were resuspended in 1.0 mL cold dissociation medium. Single-cell suspensions were sorted according to FSC, SSC, eGFP and RFP fluorescence using FacsAria III cell sorting system (BD Biosciences, San Jose, CA, USA) utilizing the 100 µm nozzle and sheath pressure 20 PSI at 4ºC. Gating was performed first using FSC and SSC to eliminate cell debris and doublets, and then the singlets were gated by the intensity of green and red fluorescent signals. First, a small sample was collected in the "bulk-sort" mode, checked using phase contrast and fluorescent microscopy with eGFP and RFP filter sets, and re-analyzed to verify purity of sorted fractions. In all cases, purity of the collected cell populations exceeded 95%. Further, single cells were sorted directly into 382-well plates containing 2.0 µL of lysis buffer supplemented with Triton X-100 and RNase inhibitor, 1.0 µL oligo-dT30VN primer and 1.0 µL dNTP mix per well. Immediately after FACS isolation, sorted cells were centrifuged at 700 g for 1 min, and snap-frozen on dry ice. In total, 1152 individual

cells were harvested: 384 GFP^{-}/RFP^{+} neurons, 384 GFP^{+}/RFP^{-} stem cells, and 384 GFP^{low}/RFP^{low} cells.

Smart-seq2 library preparation and sequencing

To generate cDNA libraries from the isolated cells, previously described Smart-seq2 protocol (2) was implemented with minor modifications. After RNA denaturation at 72°C for 3 min and reverse transcription reaction, 22 cycles of pre-amplification were carried out using IS PCR primers. PCR products were purified on Ampure XP beads and the quality of cDNA libraries was assessed on an Agilent high-sensitivity DNA chip. A successful library had an average size of 1 kb. After tagmentation using Illumina Nextera XT DNA kit and amplification of adapter-ligated fragments using index N5xx and N7xx primers for 12 cycles, the PCR fragments were again purified, their concentration and size distribution were estimated using Qubit and Agilent chips, respectively. The libraries were further diluted to 2 nM, pooled, and paired-end sequenced on Illumina HiSeq2500 instrument. Raw sequences and quality scores for all clusters were extracted using CASAVA software.

scRNA-seq data processing and quality control

Raw data from scRNA-seq were processed using a snakemake single cell RNA-seq pipeline in transcriptome mode. Briefly, data was mapped to the *H. vulgaris* transcriptome (3) using STAR v2.5.2b (4) with stringent gap penalties (*i.e.* outFilterMultimapNmax 100, alignIntronMax 1, alignIntronMin 2, scoreDelOpen -1000, scoreInsOpen -10000). Quality control was performed with the RSeQC package (5) and gene expression was estimated with RSEM (6) with groups of contigs (called clusters) defined as genes as previously described (3). Low quality cells were filtered according to

the following criteria: < 20% uniquely mapping reads, > 26.2% ERCC spike-in mapping, >68.7% rRNA mapping, >13.9% of the reads at the 10% most 3' end of transcripts (using only transcripts with complete ORFs) and < 3,000 genes detected, < 798,442 counts mapped. Any cell that failed two of these criteria was removed from the analysis.

Hierarchical clustering using the whole transcriptome

The input data was TPM normalized in RSEM and subset for the corresponding input cells and genes that underwent dimensionality reduction. A +1 pseudocount was added prior to log2 transformation. Heatmaps were generated in R using the heatmap.2 function based on Euclidean distances and the ward.D2 method. We included genes detected in at least 3 cells and cells that contained at least 5,000 transcripts. This resulted in a total of 928 cells and 166,186 transcripts. Subsequently, we performed normalization using the option "LogNormalize" with scale.factor the mean number of genes per cell. We selected a total of 6,549 variable genes (x.low.cutoff=0.5, x.high.cutoff=50, y.cutoff=0.5) and identified cell clusters using 20 PCA dimensions and a resolution parameter equal to 0.6.

Receptors

We included genes detected in at least 1 cells and cells that contained at least 1 transcript in order to include all receptor genes. This resulted in a total of 926 cells and 112 transcripts (**Supplementary Data S7**). Subsequently, we performed normalization using the option "LogNormalize" with scale.factor the mean number of genes per cell. We used the 112 receptor genes as variable genes (x.low.cutoff=0, x.high.cutoff=7, y.cutoff=-2) and identified cell clusters using 6 PCA dimensions and a resolution parameter equal to 0.6.

Ion channels

We included genes detected in at least 2 cells and cells that contained at least 2 transcripts in order to include all ion channel genes. This resulted in a total of 928 cells and 431 transcripts (**Supplementary Data S8**). Subsequently, we performed normalization using the option "LogNormalize" with scale.factor the mean number of genes per cell. We used the 431 ion channel genes as variable genes (x.low.cutoff=0, x.high.cutoff=50, y.cutoff=-3) and identified cell clusters using 15 PCA dimensions and a resolution parameter equal to 0.6.

Transcription factors

We included genes detected in at least 1 cells and cells that contained at least 1 transcript in order to include all transcription factor genes. This resulted in a total of 928 cells and 364 transcripts (**Supplementary Data S1**). Subsequently, we performed normalization using the option "LogNormalize" with scale.factor the mean number of genes per cell. We used the 364 transcription factors genes as variable genes (x.low.cutoff=0, x.high.cutoff=7, y.cutoff=-3) and identified cell clusters using 5 PCA dimensions and a resolution parameter equal to 0.6.

Reference transcriptome and gene annotation

The reference transcriptome of *H. vulgaris* strain AEP (accession number SRP133389) was assembled from 25 cDNA libraries generated from whole non-transgenic polyps at different conditions sequenced using Illumina HiSeq2500 v4 platform and annotated as previously described (3). To generate the subsets of putative neurotransmitter receptors, ion channels and transcription factors, we used SMART, Pfam domain and PANTHER family predictions generated by InterProScan. To generate

the proliferation signature dataset for cell type annotation, we identified in *Hydra* homologues of 25 highly conserved genes coding for proteins involved in cell division control and collectively known as "proliferation gene expression signature" (7) using BLAST (**Supplementary Data S5**). To get a deeper insight into the top 300 genes differentially expressed in each of 12 clusters, we performed additional annotation steps. First, we performed BLAST analysis of the longest detected ORF predicted within each transcript using BLAST+ tool. Initially, BLAST was performed for all the differentially expressed transcripts from each cluster, using Uniprot database. A more precise BLAST search was then performed on the top 300 DE genes from each cluster, using NCBI nr Refseq database. Further, in order to detect TRGs among the DE genes, the BLAST search was performed first using the whole RefSeq database, and then with Cnidarian taxon (taxid 6073) excluded. A gene was considered as TRG if e-value from BLAST without cnidarian taxon is higher than 10e-10 and e-value from BLAST with cnidarian taxon is lower than 10e-10. Further, to predict the function of the putative gene products, we screened the predicted peptide sequences for the presence of signal peptides using SignalP (8), the distribution of positively and negatively charged amino acids as well as putative strand, coiled or helical domains using GOR IV (9) and inferred putative cellular localization using DeepLoc (10). The sequences were also screened for the presence of putative structural domains using SMART (11). Finally, to test, whether the peptides encoded within TRGs may possess properties of putative AMPs, the sequences were scanned for probability of having membrane disruption activity (σ-score) using a machine learning classifier (12) with a moving window of 20 amino acids. The prediction scores are based on the support vector machine (SVM) distance to margin score, σ.

Peptides with the membrane activity score σ>0.0 were considered as putative AMPs, and peptides with σ >1.0 – as high-confidence AMPs. To refine the prediction of an active peptide encoded within the TRG *cluster62692* precursor, we repeated the membrane activity scanning for this precursor using moving windows 10 to 30 aa. The results of prediction were visualized using heat maps. Supplementary Data S3 provides a list of all the TRGs screened using this machine learning tool with corresponding σ-score values.

Supplementary Results

Expression of immune-related genes

To determine whether neurons in *Hydra* are able to interact with bacteria, we analyzed the seven neuronal subpopulations for expression of putative immune-related genes (**Fig. 3B–D**). We identified *Hydra* orthologues of murine genes coding for components of the Toll-like receptor (TLR/MyD88) pathway, intracellular NACHT- and NB-ARC-domain containing NOD-like receptors, and C-type lectin (CTL) receptors (**Fig. 3B–D**, **Fig. S14**). Analysis of the single cell transcriptome revealed that the neurons in *Hydra* express a remarkably broad repertoire of receptors, signal transducers and effector molecules to interact with bacteria.

In order to ensure that essential immune-related genes were not omitted due to the phylogenetic distance between *Hydra* and mouse, we added three more models into our analysis. We first identified a set of genes involved in immune response in *C. elegans*, *Drosophila* and a number of mollusk species (*Lottia*, *Biomphalaria* and *Crassostrea*) based on an extensive literature survey. We next identified homologues of these genes in *Hydra* (**Supplementary Data S9**). As expected, we uncovered a number of highly conserved genes present in all of the model animals and already uncovered in our survey

using murine sequences, such as genes coding for NFkB, MyD88 and JNK proteins. In addition, we uncovered numerous taxon-specific effector genes that have no homologues in *Hydra,* such as Caenacins and Caenopores from *C. elegans*, Myticins and Mytimacins from clams, and Drosomycin and Diptericin from *Drosophila* (**Supplementary Data S9**). This is consistent with our observations (**Fig. 3**) that non-conserved taxon-restricted genes play a major role in the immune response of *Hydra* as in any other organism. Finally, we identified a small subset of conserved genes that have homologues in *Hydra*, but that were not uncovered using murine homologues. This category includes, for instance, genes coding for scavenger receptors and antiviral response RIG-I proteins (**Supplementary Data S9**). A reciprocal BLAST analysis of these *Hydra* transcripts against the UniProt database, however, retrieved members of other, not immune-related families, indicating a lack of true homology (**Supplementary Data S9**). Moreover, there is no empiric evidence for the role of these putative genes in the immune response of *Hydra*. Hence, we conclude that no essential immune-related genes were omitted in the analysis of *Hydra* homologues of murine immune genes (**Fig. 3B–D**, **Fig. S14**). In sum, our observations provide evidence that neurons in *Hydra* are immunocompetent cell capable of interacting with bacteria.

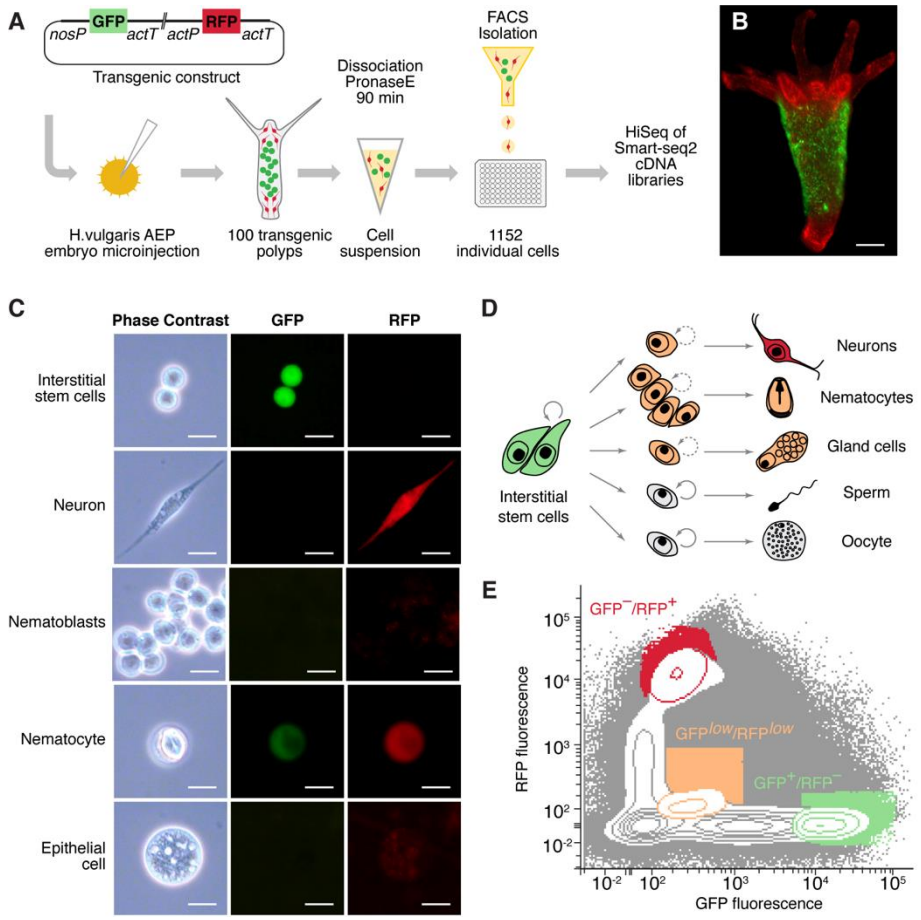


Fig. S1. Differential fluorescent labeling and FACS-mediated isolation of interstitial cell lineage. (**A**) The workflow of the single-cell molecular profiling of *Hydra* interstitial cell lineage. (**B**) A transgenic polyp expressing GFP in the interstitial stem cells in the middle body compartment and RFP in the neurons concentrated in the tentacles, hypostome, and the foot region. (**C**) Phase contrast and fluorescent microscopy of disintegrated cells from transgenic *Hydra*. Interstitial stem cells are strongly GFP-positive but show now RFP fluorescence ($GFP^{+/}RFP^{-}$). Mature neurons are strongly RFP-positive, while demonstrating no GFP fluorescence (GFP^{-}/RFP^{+}). Nematoblasts and nematocytes showed weak GFP and weak RFP fluorescence (GFP^{low}/RFP^{low}). Epithelial cells demonstrate only weak unspecific red fluorescence that is microscopically visible

but is not detected by FACS. Scale bar: 10 µm. (**D**) Combination of two fluorescent proteins allowed differential labeling of all somatic cell types within the interstitial cell lineage. Interstitial stem cells continuously proliferate and differentiate into neurons, gland cells, nematocytes, and gametes. (**E**) Three cell populations were isolated from dissociated transgenic *Hydra* polyps using FACS based on their relative GFP and RFP fluorescence: GFP^{+}/RFP^{-}, GFP^{-}/RFP^{+}, and GFP^{low}/RFP^{low}.

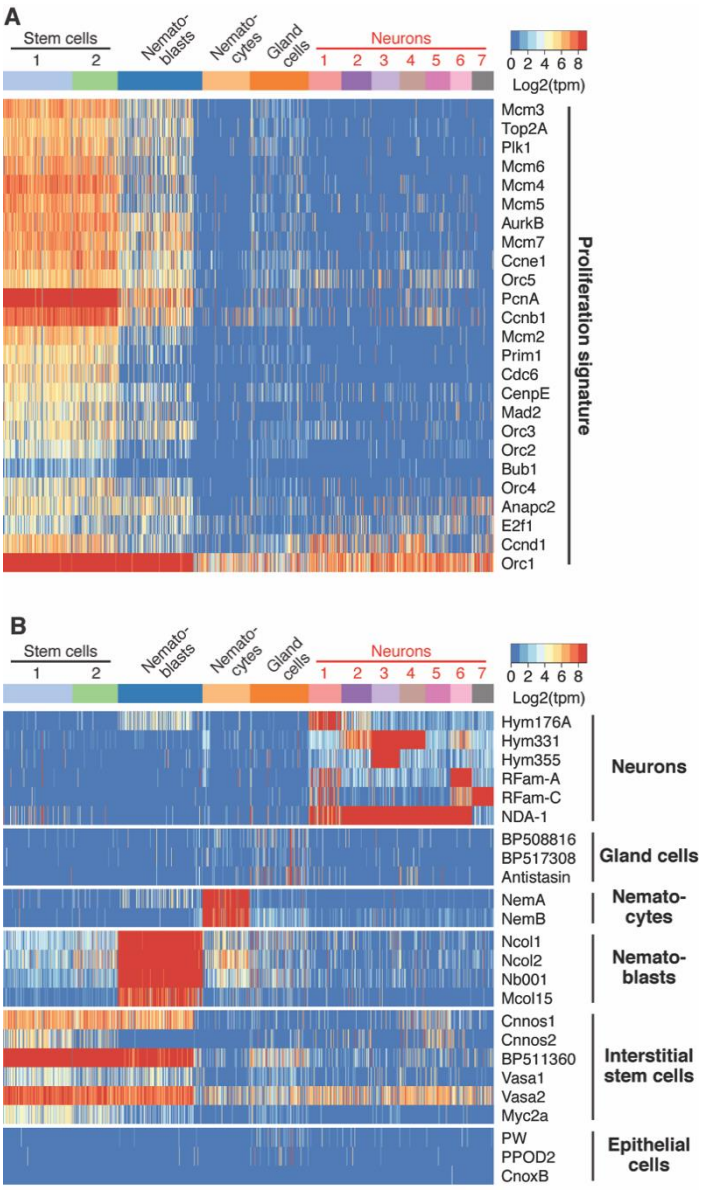


Fig. S2. Annotation of cell populations using the expression profiles of the proliferation signature and cell-type specific genes. (A) Heatmap illustrating expression of genes comprising the proliferation signature (see Data S5). These genes are strongly expressed in proliferating stem cells and nematoblasts and are not expressed in the differentiated non-dividing neurons and nematocytes. **(B)** Heatmap illustrates expression of specific marker genes used to annotate the 12 clusters (see Data S6).

Markers specific for the interstitial stem cells and their progeny (neurons, nematoblasts, nematocytes, gland cells) demonstrate strong and distinct expression pattern. Marker genes characteristic for the epithelial cells are not expressed in the dataset, providing evidence for specific profiling of the interstitial lineage without contamination from the ectodermal and endodermal cells.

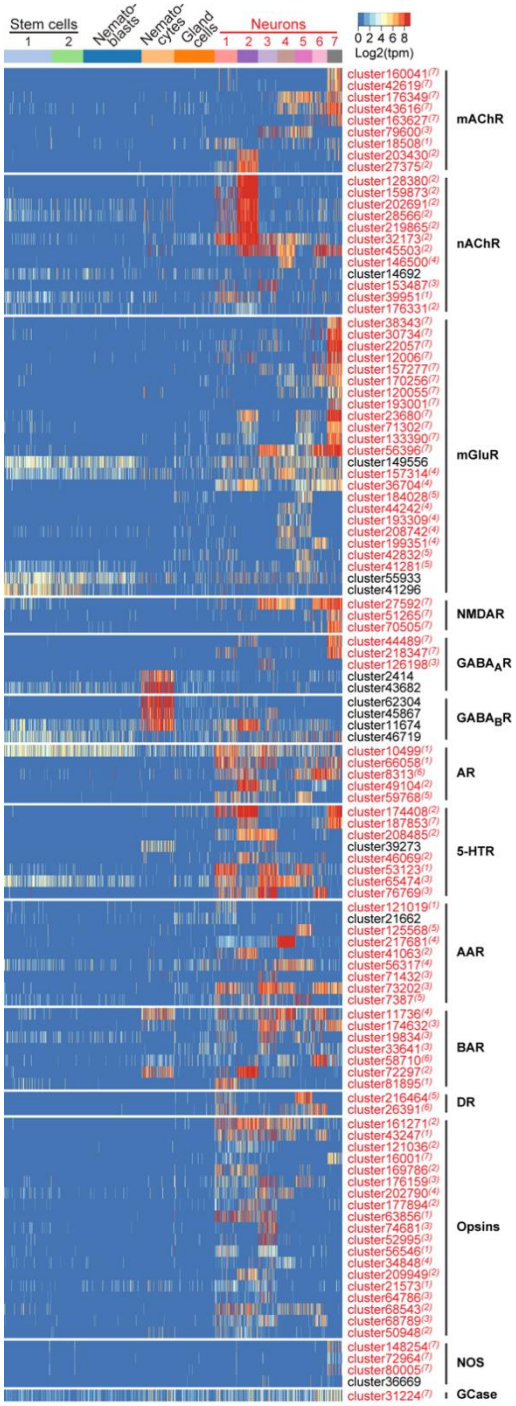

Stem cells
1
2
Nemato-blasts
Nemato-cytes
Gland Cells
Neurons
1 2 3 4 5 6 7
0 2 4 6 8
Log2(tpm)
cluster160041(7)
cluster42619(7)
cluster176349(7)
cluster43616(7)
cluster163627(7)
cluster79600(3)
cluster18508(1)
cluster203430(2)
cluster27375(2)
mAChR
cluster128380(2)
cluster159873(2)
cluster202691(2)
cluster28566(2)
cluster219865(2)
cluster32173(2)
cluster45503(2)
cluster146500(4)
cluster14692
cluster15348(3)
cluster39951(1)
cluster176331(2)
nAChR
cluster38343(7)
cluster30734(7)
cluster22057(7)
cluster12006(7)
cluster157277(7)
cluster170256(7)
cluster120055(7)
cluster193001(7)
cluster23680(7)
cluster71302(7)
cluster133390(7)
cluster56396(7)
cluster149556
cluster157314(4)
cluster36704(4)
cluster184028(5)
cluster44242(4)
cluster193309(4)
cluster208742(4)
cluster199351(4)
cluster42832(5)
cluster41281(5)
cluster55933
cluster41296
mGluR
cluster27592(7)
cluster51265(7)
cluster70505(7)
NMDAR
cluster44489(7)
cluster218347(7)
cluster126198(3)
cluster2414
cluster43682
GABA$_A$R
cluster62304
cluster45867
cluster11674
cluster46719
GABA$_B$R
cluster10499(1)
cluster66058(1)
cluster8313(6)
cluster49104(2)
cluster59768(5)
AR
cluster174408(2)
cluster187853(7)
cluster208485(2)
cluster39273
cluster46069(2)
cluster53123(1)
cluster65474(3)
cluster76769(3)
5-HTR
cluster121019(1)
cluster21662
cluster125568(5)
cluster217681(4)
cluster41063(2)
cluster56317(4)
cluster71432(3)
cluster73202(3)
cluster7387(5)
AAR
cluster11736(4)
cluster174632(3)
cluster19834(3)
cluster33641(3)
cluster58710(6)
cluster72297(2)
cluster81895(1)
BAR
cluster216464(5)
cluster26391(6)
DR
cluster161271(2)
cluster43247(1)
cluster121036(2)
cluster16001(7)
cluster169786(2)
cluster176159(3)
cluster202790(4)
cluster177894(2)
cluster63856(1)
cluster74681(3)
cluster52995(3)
cluster5646(1)
cluster34848(4)
cluster209949(2)
cluster21573(1)
cluster64786(3)
cluster68543(2)
cluster68789(3)
cluster50948(2)
Opsins
cluster148254(7)
cluster72964(7)
cluster80005(7)
cluster36669
NOS
cluster31224(7)
GCase

Fig. S3. Each neuronal population is characterized by a specific set of transcripts coding for receptors for neuromediators. Heatmap illustrating expression levels of genes coding for putative neurotransmitter receptors (see Data S7) in the interstitial stem cell lineage: muscarinic (mAChR) and nicotinic (nAChR) acetylcholine receptors, metabotropic (mGluR) and N-methyl-D-aspartate (NMDAR) glutamate receptors, A-type ($GABA_AR$) and B-type ($GABA_BR$) γ-aminobutyric acid, adenosine (AR) and serotonin (5-HTR) receptors, alpha- (AAR) and beta- (BAR) adrenergic receptors, dopamine receptor (DR). Multiple transcripts coding for putative light-sensitive receptors of opsin family are present in the neurons of *Hydra*. Genes coding for nitric oxide synthases (NOS) are expressed exclusively in the neuronal population N7. Transcripts coding for the guanylate cyclase enzymes (GCase) that may serve as a receptor for NO secondary messenger are also enriched in the subpopulation N7. Transcripts specifically upregulated in the neurons are labelled red, superscript numbers indicate the nerve cell cluster (N1-N7) where the transcripts are significantly (padj<0.05) enriched.

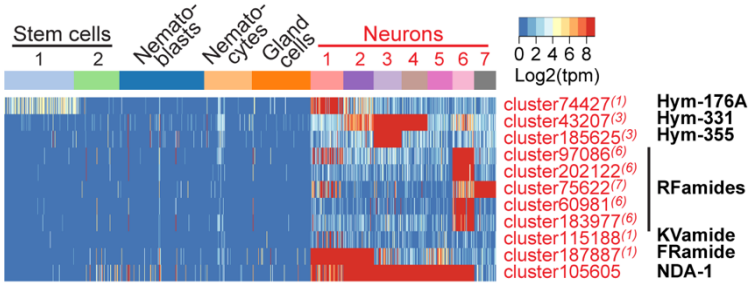


Fig. S4. Each neuronal population is characterized by a specific set of transcripts coding for neuropeptides. Heatmap illustrating expression of genes coding for characterized (13) peptides Hym-176, Hym-331, and Hym-355, five paralogues of RFamide family, a KVamide, FRamide, and recently characterized neuron-derived antimicrobial peptide NDA-1 (14). Transcripts specifically upregulated in the neurons are labelled red, superscript numbers indicate the nerve cell population (N1-N7) where the transcripts are significantly (padj<0.05) enriched.

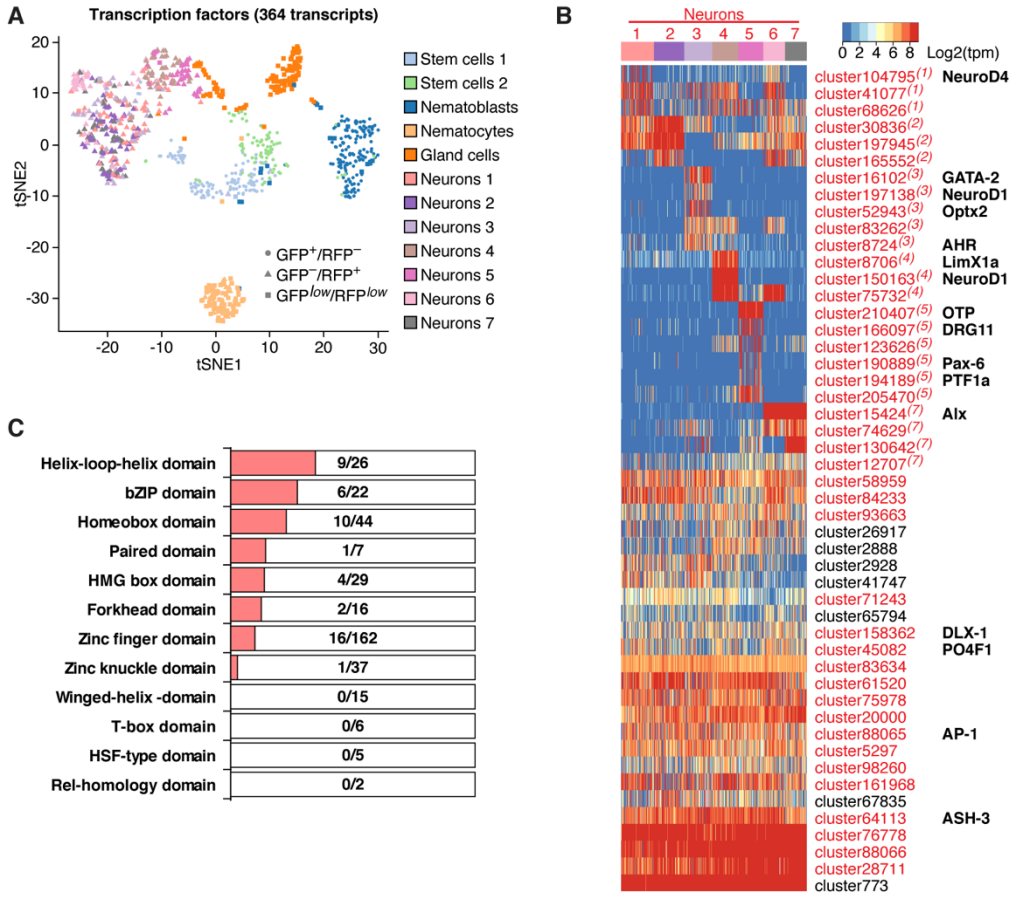
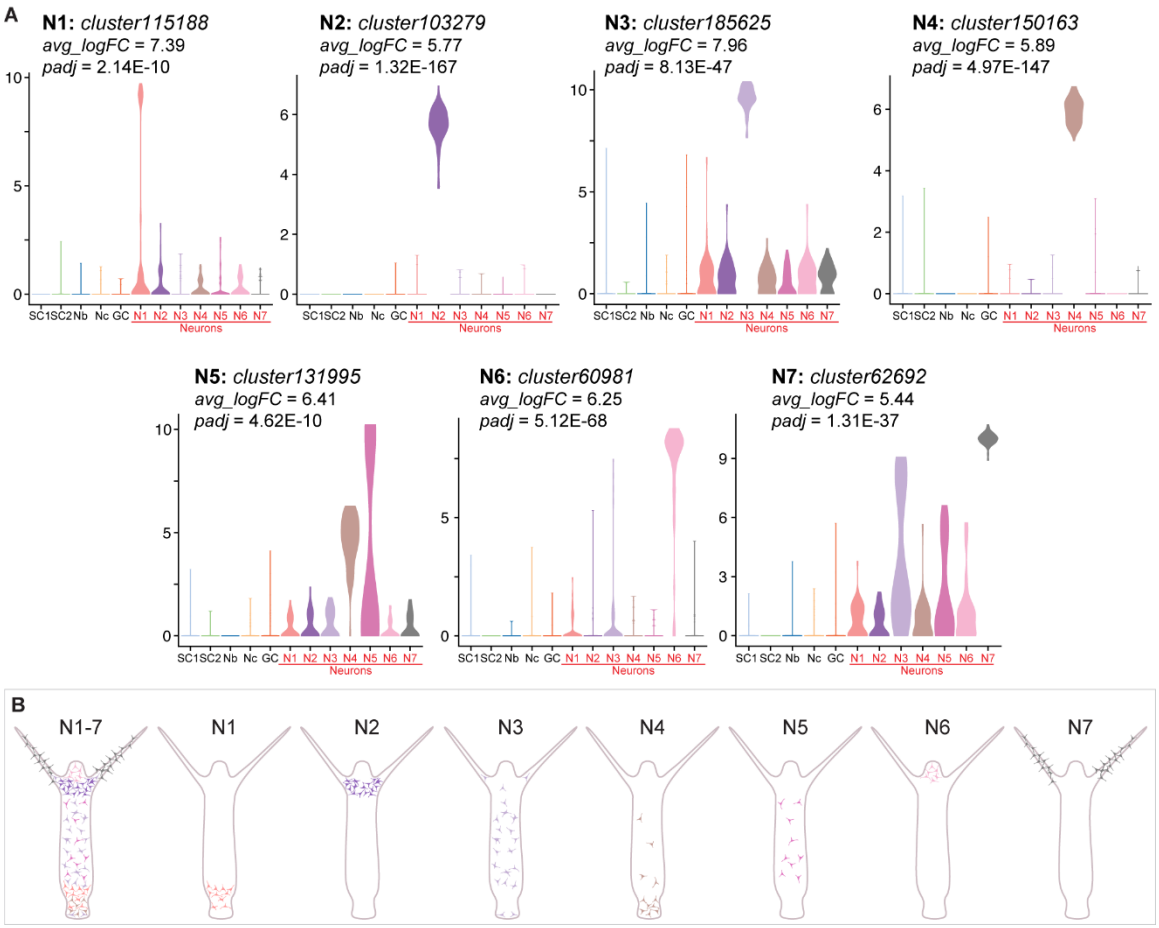


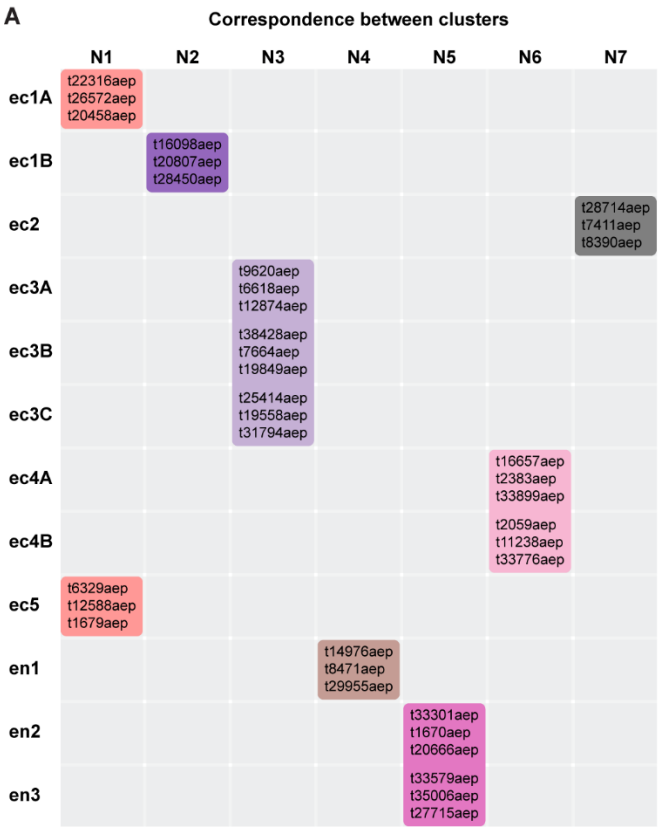
Fig. S5. Seven neuronal subpopulations express a common set of transcription factors. (**A**) t-SNE map based on expression analysis of 364 transcripts coding for transcriptional factors (TFs, see Data S1). All seven populations of neurons express common set of TFs and are not segregated on the t-SNE plot, in contrast to the stem cells, nematoblasts, nematocytes and the gland cells. (**B**) Heatmap illustrating expression levels of transcripts coding for TFs and highly abundant in the neurons. Transcripts specifically upregulated in the neurons are labelled red, superscript numbers indicate the nerve cell cluster (N1-N7) where the transcripts are significantly (padj<0.05) enriched. In addition to common neuron-specific TFs, each neuronal population is characterized by a combinatorial expression of few genes encoding other TF, likely acting as selector genes.

Notably, multiple homologues of neurogenic TFs found in Bilateria (right column) are present in distinct neuronal populations, suggesting their conserved role in neurogenesis in the animal kingdom. (**C**) The neuron-specific TF signature is composed of mainly of Zn-finger, homeodomain and helix-loop-helix DNA-binding proteins. On the contrary, members of other families, such as Winged-helix, T-box, and Rel-homology domain TFs, are absent from the neurons.



Fog. S6. Expression of marker genes strongly enriched specific nerve cells subpopulations (N1-N7) used to map the neuronal populations in *Hydra*. (**A**) Violin plots demonstrating expression levels of seven marker genes used to map distinct neuronal populations by *in situ* hybridization (**Fig. 1*J***) in all cells within the interstitial cell lineage. Y-axis represents expression level as LogRPKM. Average upregulation fold change (avg_logFC) and adjusted p-value (padj) are indicated for each gene. (**B**) The nerve net of *Hydra* is organized in seven spatially restricted cell populations. Localization of the populations is inferred from the *in situ* hybridization results (**Fig. 1*J***).

A



B

Transcript ID, Siebert et al.	Transcript ID, our study	Cluster, our study	logFC	Padj
t22316aep	cluster60609	N1	5,13	3,33E-49
t26572aep	cluster33557	N1	3,16	1,30E-32
t20458aep	cluster200791	N1	3,08	1,05E-10
t16098aep	cluster103279	N2	5,77	1,32E-167
t20807aep	cluster66153	N2	4,43	1,40E-84
t28450aep	cluster119726	N2	5,53	2,52E-88
t28714aep	cluster56664	N7	2,72	2,53E-86
t7411aep	cluster94531	N7	8,77	5,72E-38
t8390aep	cluster62692	N7	5,44	1,31E-37
t9620aep	cluster199921	N3	3,10	1,48E-42
t6618aep	cluster124097	N3	1,05	6,07E-38
t12874aep	cluster185625	N3	7,96	8,13E-47
t38428aep	cluster52943	N3	5,33	9,33E-58
t7664aep	cluster160659	N3	7,20	1,79E-46
t19849aep	cluster155448	N3	4,31	2,19E-26
t25414aep	cluster65043	N3	3,54	1,85E-34
t19558aep	cluster137620	N3	4,77	8,58E-87
t31794aep	cluster106527	N3	5,75	1,42E-56
t16657aep	cluster183977	N6	8,18	1,68E-39
t2383aep	cluster58169	N6	5,72	2,69E-109
t33899aep	cluster60981	N6	6,25	5,12E-68
t2059aep	cluster202122	N6	8,34	6,69E-50
t11238aep	cluster127881	N6	1,54	5,79E-03
t33776aep	cluster61788	N6	2,79	4,78E-22
t6329aep	cluster115188	N1	7,39	2,14E-10
t12588aep	cluster193988	N1	6,77	4,91E-71
t1679aep	cluster74427	N1	7,60	1,12E-35
t14976aep	cluster150163	N4	5,89	4,97E-147
t8471aep	cluster133757	N4	8,11	1,00E-56
t29955aep	cluster160146	N4	5,94	2,31E-80
t33301aep	cluster206126	N5	7,03	3,66E-33
t1670aep	cluster138649	N5	7,43	1,42E-39
t20666aep	cluster153425	N5	7,09	2,85E-47
t33579aep	cluster53954	N5	7,27	6,79E-24
t35006aep	cluster44629	N5	6,35	8,80E-102
t27715aep	cluster108398	N5	5,02	1,49E-16

Fig. S7. Correspondence between 7 neuronal clusters identified in our study (N1-7) and 12 neuronal subtypes reported by Siebert and co-authors (15)**.** Expression analysis of marker genes specifically expressed in each of 12 neuronal subtypes reported by Siebert and co-authors (see Fig. S41 in ref. (15)) uncovers a clear correspondence between the clusters. Marker genes characteristic for 12 neuronal subtypes from study of Siebert and co-authors have corresponding transcripts in our dataset that are significantly enriched in specific neuronal subpopulations (on B).

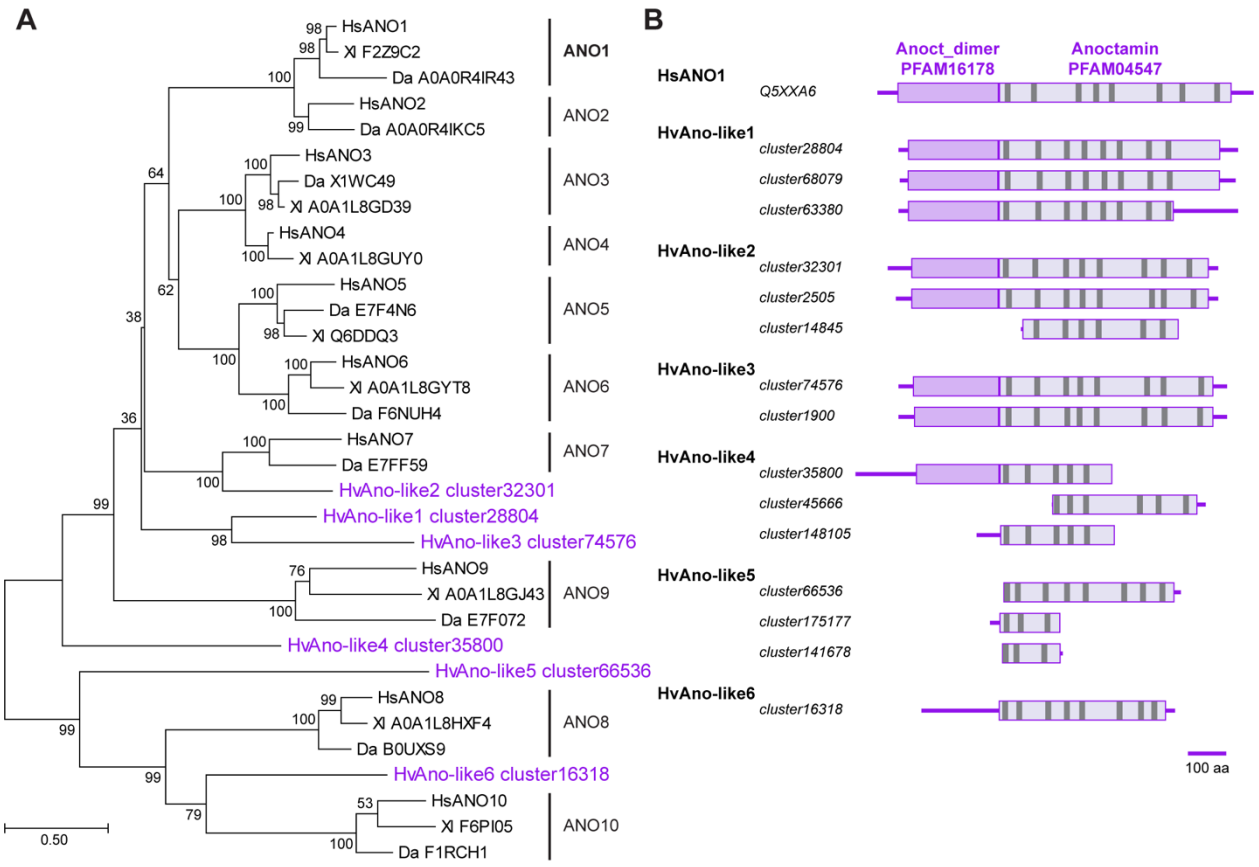


Fig. S8. Pacemaker-specific Ca-activated chlorine channels of Anoctamin family are highly conserved in *Hydra*. (**A**) Detailed phylogenetic analysis of Anoctamin ion channels from human (*Homo sapiens*, Hs), frog (*Xenopus laevis*, Xl), zebrafish (*Danio rerio*, Da), and hydra (*Hydra vulgaris*, Hv). *Hydra* possesses six genes coding for anoctamin-like ion channels – HvAno-like1–6. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel model (16). The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Only fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 549 positions in the final dataset. Evolutionary analyses

were conducted in MEGA7 (17). (**B**) Domain structure analysis using SMART and TMHMM algorithms uncovers remarkably high conservation in structure between the human ANO1 protein and six anoctamin-like channels from *Hydra*. Multiple transcripts corresponding to the same gene were frequently identified in the reference transcriptome of *Hydra*, suggesting either existence of alternative splice-variants or poor assembly of the full-length transcripts, most likely due to very low expression level of the genes (see Fig. S11). Localization of specific Anoctamin dimer (PFAM16178) and Anoctamin (PFAM04547) domains and transmembrane domains is illustrated.

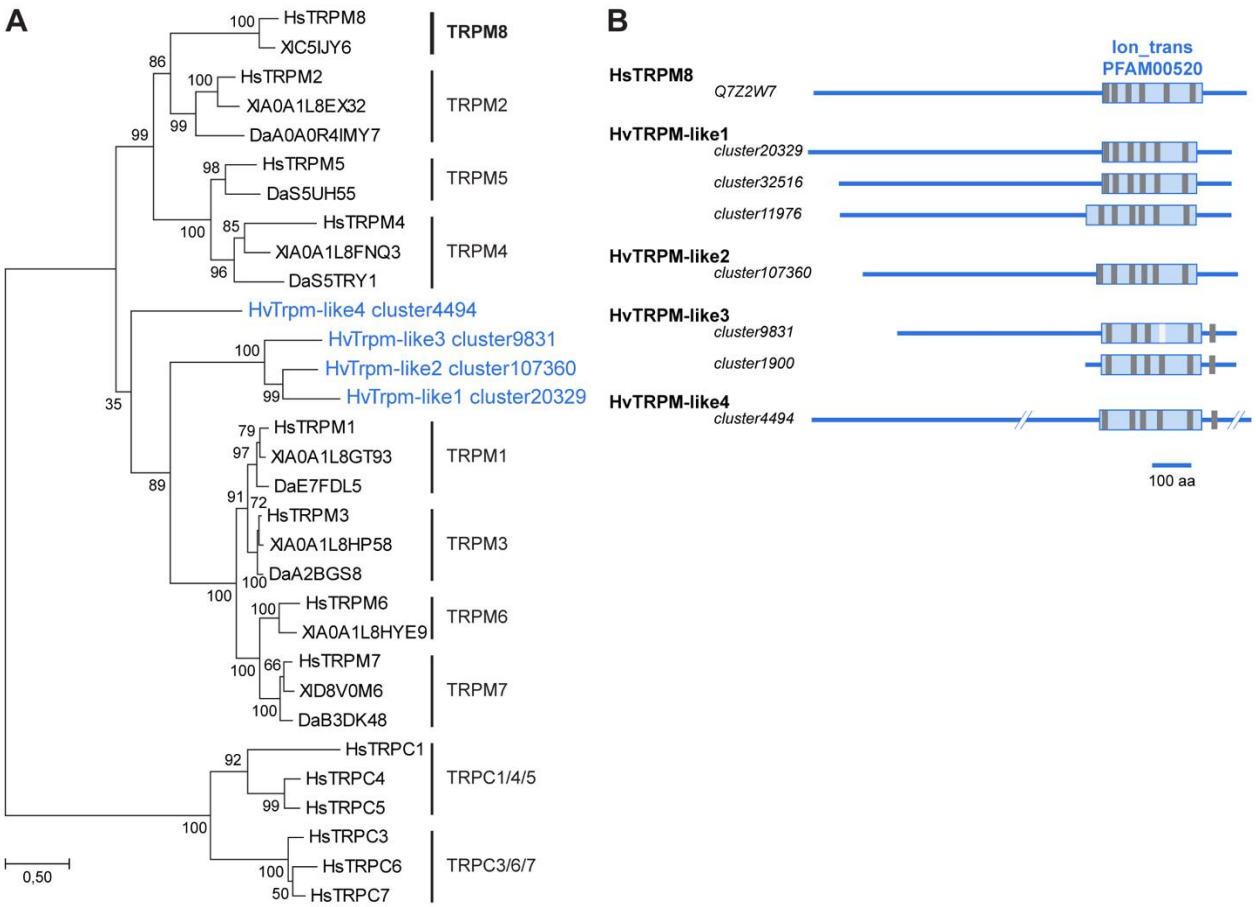


Fig. S9. Pacemaker-specific cation TRPM-like ion channels are highly conserved in *Hydra*. (A) Detailed phylogenetic analysis of TRPM-like ion channels from human (*Homo sapiens*, Hs), frog (*Xenopus laevis*, Xl), zebrafish (*Danio rerio*, Da), and hydra (*Hydra vulgaris*, Hv). Human TRPC-like channels are used as an outgroup. *Hydra* has four genes coding for TRPM-like ion channels – HvTRPM-like1–4. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel model (16). The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Only fewer than 10% alignment gaps, missing data, and ambiguous bases were

allowed at any position. There were a total of 679 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (17). (**B**) Domain structure analysis using SMART and TMHMM algorithms uncovers high conservation in structure between the human TRPM8 protein and four TRPM-like channels in *Hydra*. Multiple transcripts corresponding to the same gene were frequently identified in the reference transcriptome of *Hydra*, suggesting either existence of alternative splice-variants or poor assembly of the full-length transcripts, most likely due to the very low expression level of the genes (see Fig. S11). Localization of specific Ion_trans (PFAM00520) domains and transmembrane domains (grey) is illustrated.

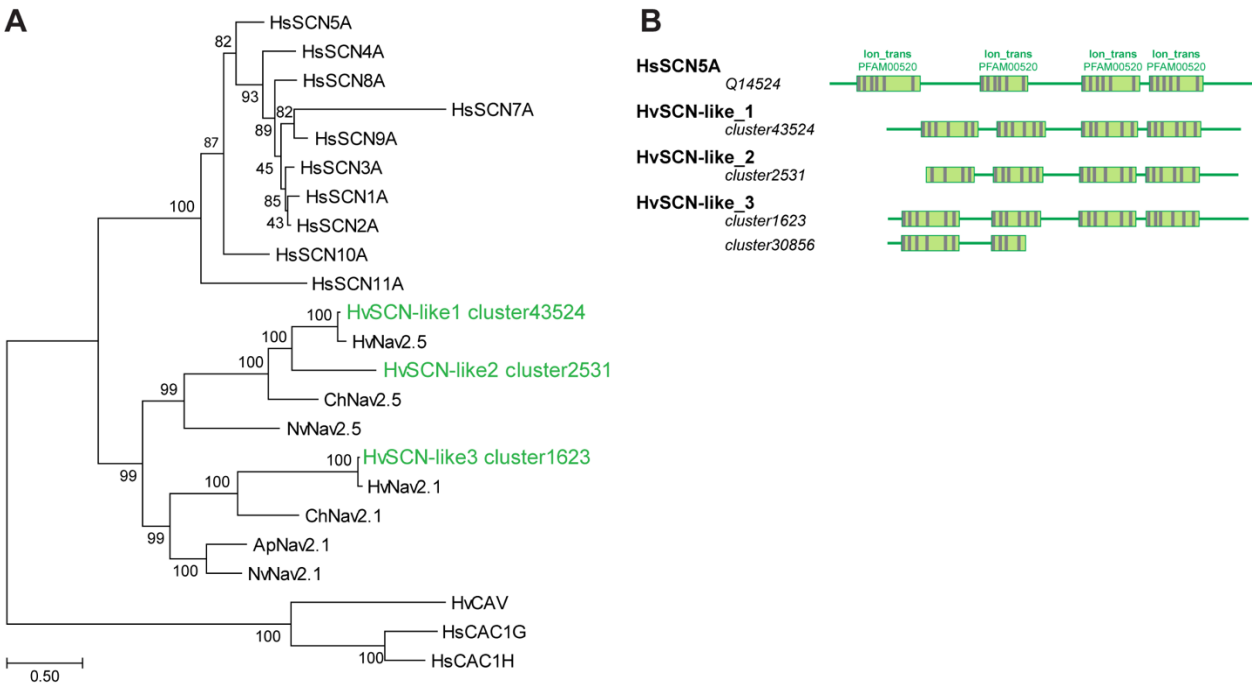


Fig. S10. Pacemaker-specific voltage-gated sodium SCN-like ion channels are highly conserved in *Hydra*. (**A**) Detailed phylogenetic analysis of SCN-like ion channels from human (*Homo sapiens*, Hs), and hydra (*Hydra vulgaris*, Hv), as well as previously characterized (18) Nav channels from sea anemones (*Nematostella vectensis*, Nv, and *Aiptasia pallida*, Ap) and hydroid jellyfish (*Clytia hemisphaerica*, Ch). Human CAV and CAC channels are used as an outgroup. *Hydra* has three genes coding for SCN-like ion channels – HvSCN-like1–3. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel model (16). The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Only fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1429 positions in the final dataset. Evolutionary analyses were conducted in MEGA7

(17). (**B**) Domain structure analysis using SMART and TMHMM algorithms uncovers high conservation in structure between the human SCN5A protein and three SCN-like channels in *Hydra*. Two transcripts corresponding to the same gene coding for HvSCN-like3 were identified in the reference transcriptome of *Hydra*, suggesting either existence of alternative splice-variants or poor assembly of the full-length transcripts, most likely due to the very low expression level of the genes (see Fig. S11). Localization of specific Ion_trans (PFAM00520) domains and transmembrane domains (grey) is illustrated.

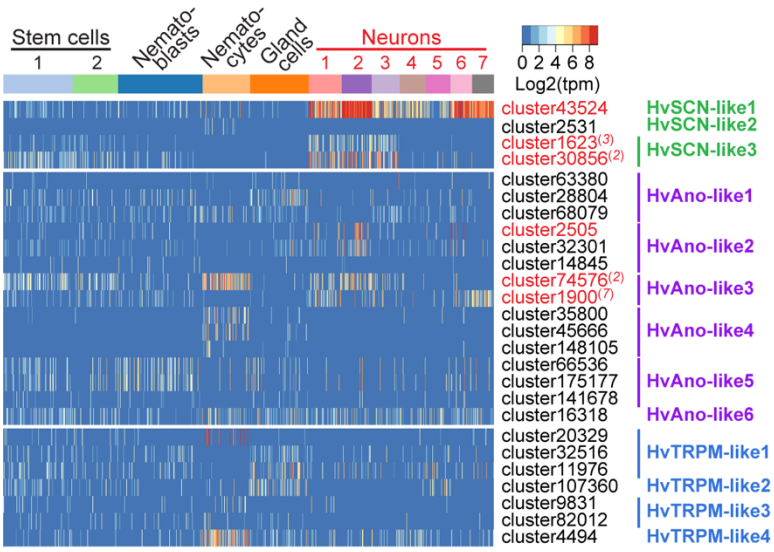


Fig. S11. Pacemaker-specific ion channels are expressed in *Hydra* neurons. Heatmap illustrating expression levels of transcripts coding for putative SCN-, ANO1-, and TRPM-like ion channels in the interstitial stem cell lineage of *Hydra*. Six transcripts (marked red) are significantly upregulated in the neurons, superscript numbers indicate the nerve cell population (N1-N7) where the transcripts are significantly (padj<0.05) enriched.

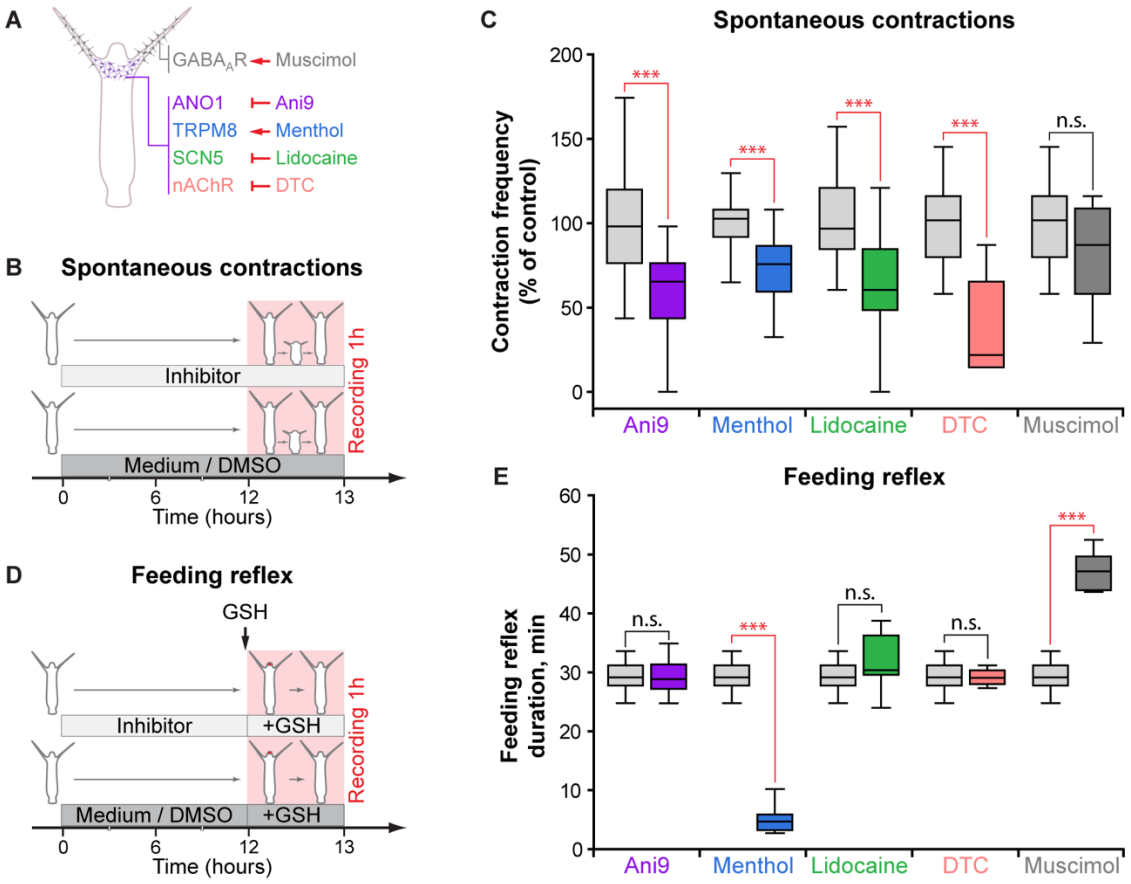


Fig. S12. Pharmacological substances modulate activity of ion channels and receptors and differentially affect behaviors generated by two distinct neuronal populations in *Hydra*. (**A**) Four pharmacological agents (Ani9, Menthol, Lidocaine, and DTC) target ion channels and receptors expressed on the neuronal population N2 in the basis of tentacles. Muscimol is an agonist of $GABA_A$ receptors expressed exclusively in the neuronal population in the tentacles. (**B**) Effects of chemicals onto rhythmic spontaneous contractions of *Hydra* were assayed in the following experimental setup: polyps were incubated in the inhibitors for 12 hours prior to a 1 hour recording of contractile behavior. Polyps incubated in 0.16% DMSO-supplemented (for Ani9) or pure (all other chemicals) Hydra-medium served as control. (**C**) Contraction frequency is significantly reduced in the presence of all chemicals targeting the channels expressed on

the pacemaker population N2, but not affected in the presence of muscimol, that interferes with the population N7. (**D**) Effects of pharmacological substances onto feeding reflex of *Hydra* were assayed in the following experimental setup: polyps were incubated in the inhibitors for 12 hours, the feeding reflex was induced by reduced glutathione, and the duration of feeding response (*i.e.* time between mouth opening and closure) was recorded. Polyps incubated in 0.16% DMSO-supplemented (for Ani9) or pure (all other chemicals) Hydra-medium served as control. (**E**) Three chemicals (Ani9, Lidocaine, and DTC) show no effect onto duration of the feeding reflex, while muscimol prolongs the feeding response. Complementary patterns of pharmacological effects indicate that two behaviors, the rhythmic contractions and feeding reflex, are controlled by two independent neuronal populations. Only menthol had effects onto both behaviors, reduced contraction frequency and shortened the feeding response. This may be explained by the low specificity of this chemical. In contrast to other four chemicals, that are known to be highly specific and target only narrow class of ion channels or receptors, menthol may interfere with the activity of very broad range of molecules, including ion channels, neurotransmitter receptors and enzymes (19). Collectively, these data corroborate that the feeding reflex in *Hydra* is controlled by the GABAergic neuronal population N7, and the spontaneous contractile behavior – by the cholinergic neuronal population N2. n=10–49 animals (contraction frequency), n=11–14 animals (feeding response duration), * - $p<0.05$; ** - $p<0.005$; *** - $p<0.0005$; n.s. – $p>0.05$.

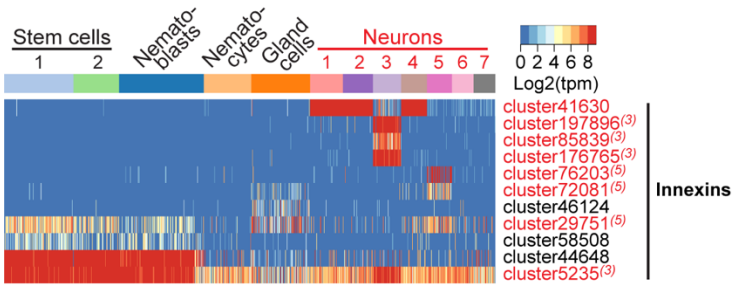


Fig. S13. Neurons in *Hydra* are electrically coupled by gap junctions made of innexin proteins. Heatmap illustrating expression levels of transcripts coding for putative innexin channels in the interstitial stem cell lineage of *Hydra*. Multiple innexin genes are expressed in the neuronal populations N1–7. All cells within the population N2 that contains the pacemakers express the *innexin* transcript *cluster41630*, that is also expressed in the populations N1 and N4 localized in the foot region of a polyp (**Fig. 1*J***). This gene codes for Innexin-2 protein that has been previously reported as essential for body contractions of *Hydra* (20).

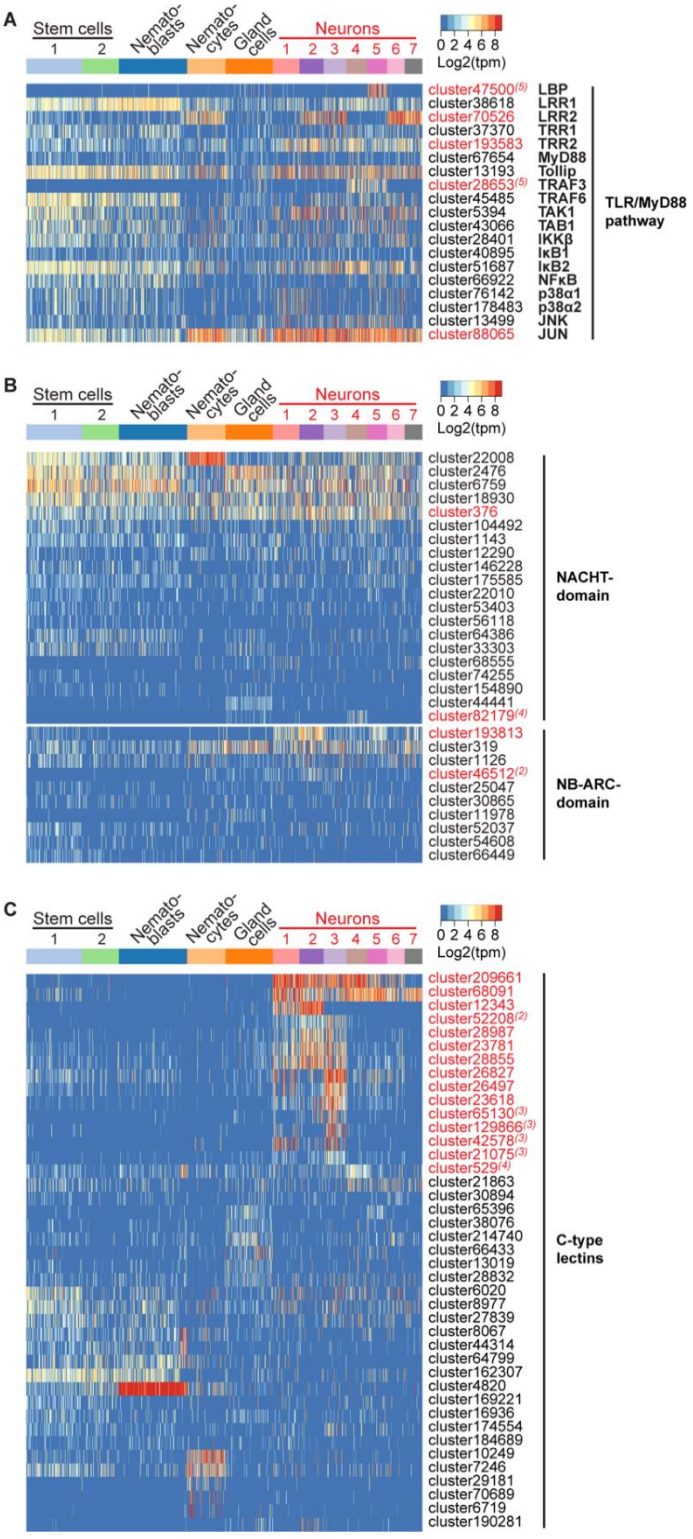


Fig. S14. Neurons in *Hydra* are immunocompetent cells. (**A–C**): Heatmaps illustrating expression levels of genes coding for putative components of immune-related

TLR/MyD88 pathway (on A), NACHT- and NB-ARC-domain containing NOD-like receptors (on B) and C-type lectin receptors (on C) in the interstitial stem cell lineage. Transcripts specifically upregulated in the neurons are labelled red, superscript numbers indicate the nerve cell population (N1-N7) where the transcripts are significantly (padj<0.05) enriched.

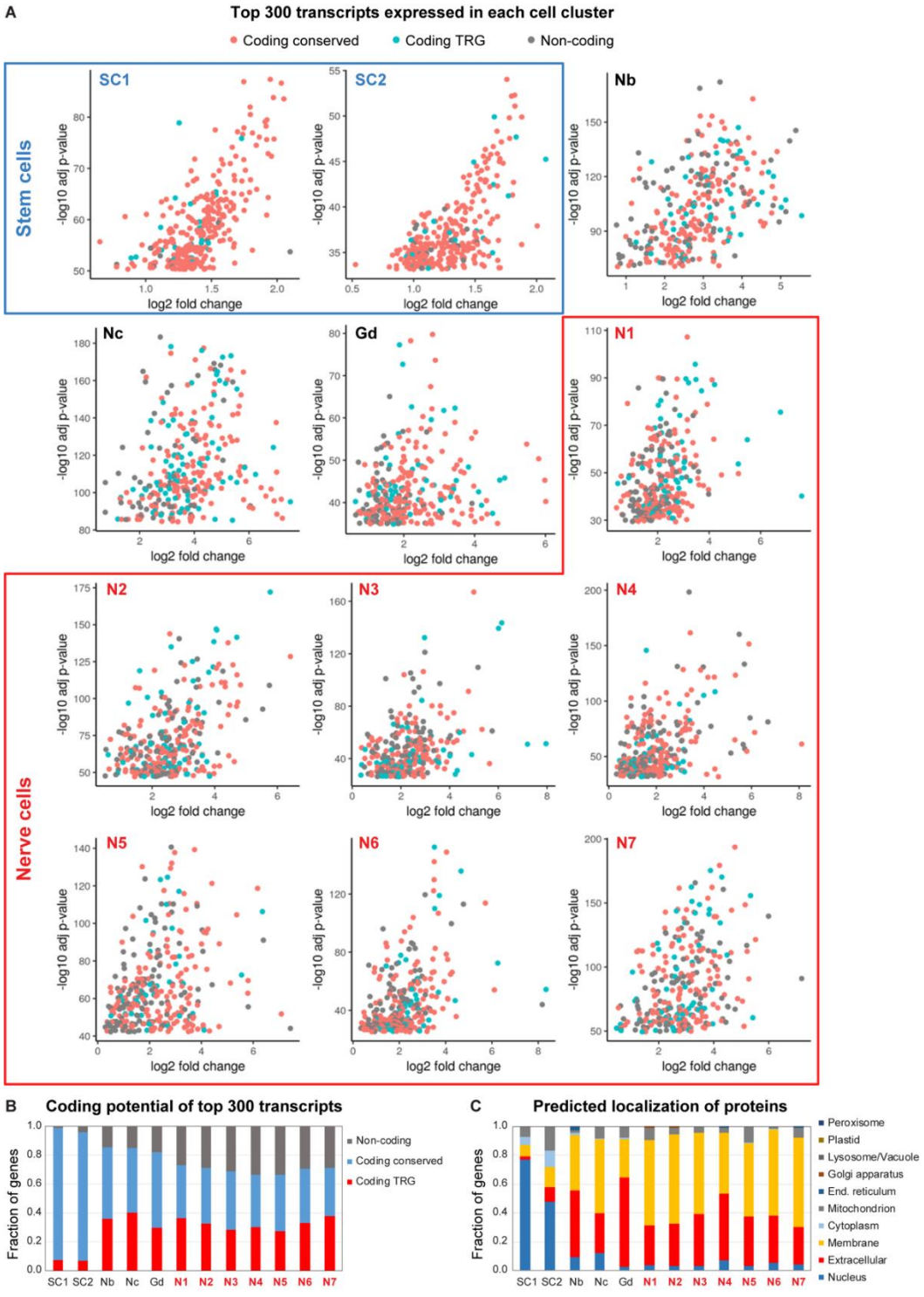


Fig. S15. Taxon-restricted genes (TRGs) determine the identity of each neuronal population in *Hydra* and code for predominantly secreted peptides. (**A**) Scatter plots illustrating expression levels of top 300 genes differentially significantly up-regulated in

each of 12 cell clusters. Highly-conserved protein-coding genes (torquise) dominate the signature of the stem cells. The identity of the neurons is determined by taxon-restricted protein-coding transcripts (red) and non-coding RNAs. (**B**) Relative abundance of conserved and taxon-restricted protein-coding transcripts and transcripts with no detectable open reading frame among top 300 genes differentially expressed in each of 12 clusters. (**C**) Annotation of protein-coding transcripts among top 300 genes differentially expressed in each of 12 clusters. Transcripts coding for proteins localized in the nucleus (likely, transcription factors and RNA-binding proteins; blue) dominate the signature of the interstitial stem cells. On the contrary, the contribution of transcripts encoding nuclear proteins to the identity of the neurons is minimal, consistent with our clustering based on expression of transcription factors (see Fig. S5). Instead, nearly 90% of peptides coded by transcripts specifically expressed in each neuronal population are annotated as extracellular or membrane-associated proteins (red and yellow). This suggests that the neuronal phenotype is characterized by a complex set of secreted factors (secretome), comprising mainly novel, taxon-restricted products. Some of them may mediate the communication of the nervous system with the microbiota.

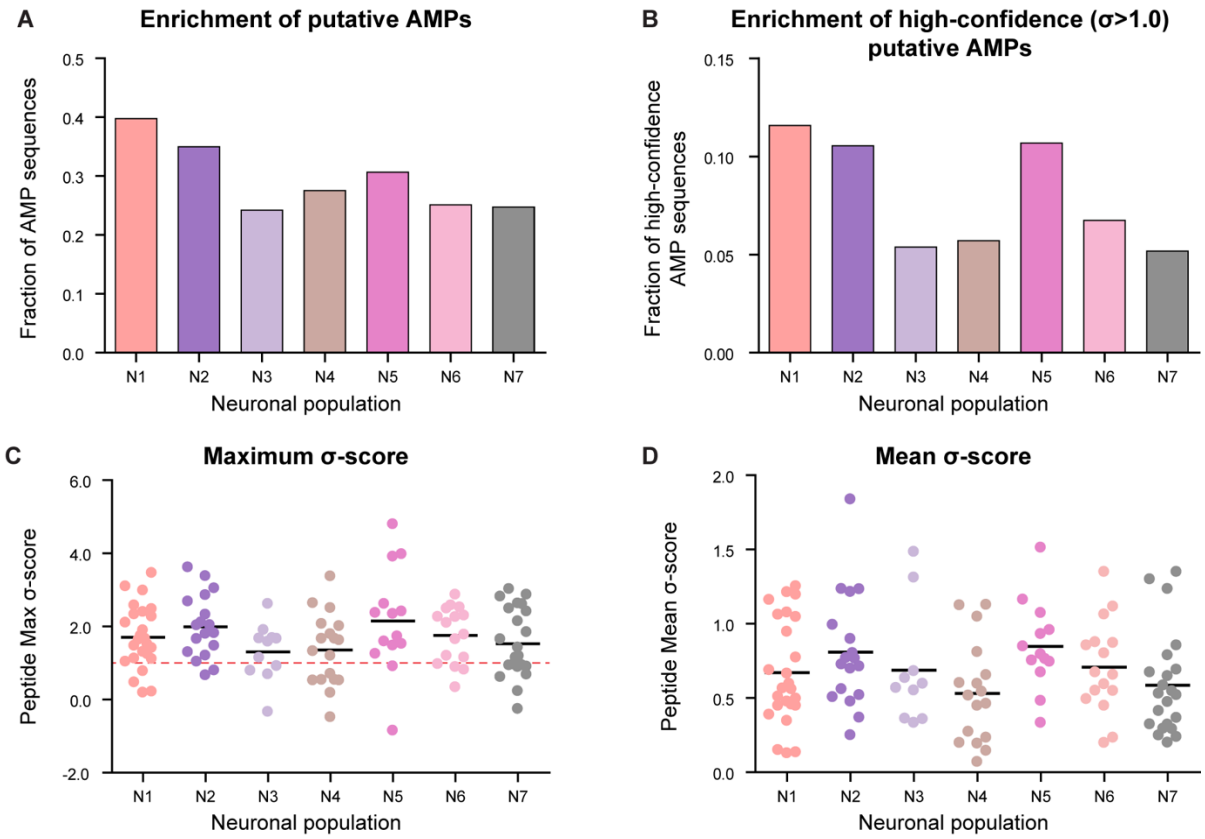


Fig. S16. Secreted peptides encoded in neuron-specific TRGs contain putative antimicrobial peptides. (A) A substantial fraction of secreted peptides encoded by neuron-specific TRGs contains small peptides that may function as antimicrobial peptides, as revealed by a moving window (20 aa) small-peptide scan. (**B**) Neuronal populations N1, N2 and N5 are particularly enriched in putative AMPs with high confidence (σ>1.0) membrane destabilizing activity. (**C**) Distribution of maximum σ-score values for each of secreted peptides encoded by neuron-specific TRGs in seven neuronal subpopulations. Raw values are provided in Data S3. Majority of the TRG-encoded peptides contain at least one subsequence of 20 aa with a positive σ-score. The neuronal population N2 that contains the pacemakers is particularly rich in secreted peptides with high-confidence (σ>1.0, red dotted line) antimicrobial activity. (**D**)

Distribution of mean σ-score values for each of secreted peptides encoded by neuron-specific TRGs in seven neuronal subpopulations illustrates a high likelihood of containing an AMP for the peptides. Raw values are provided in Data S3.

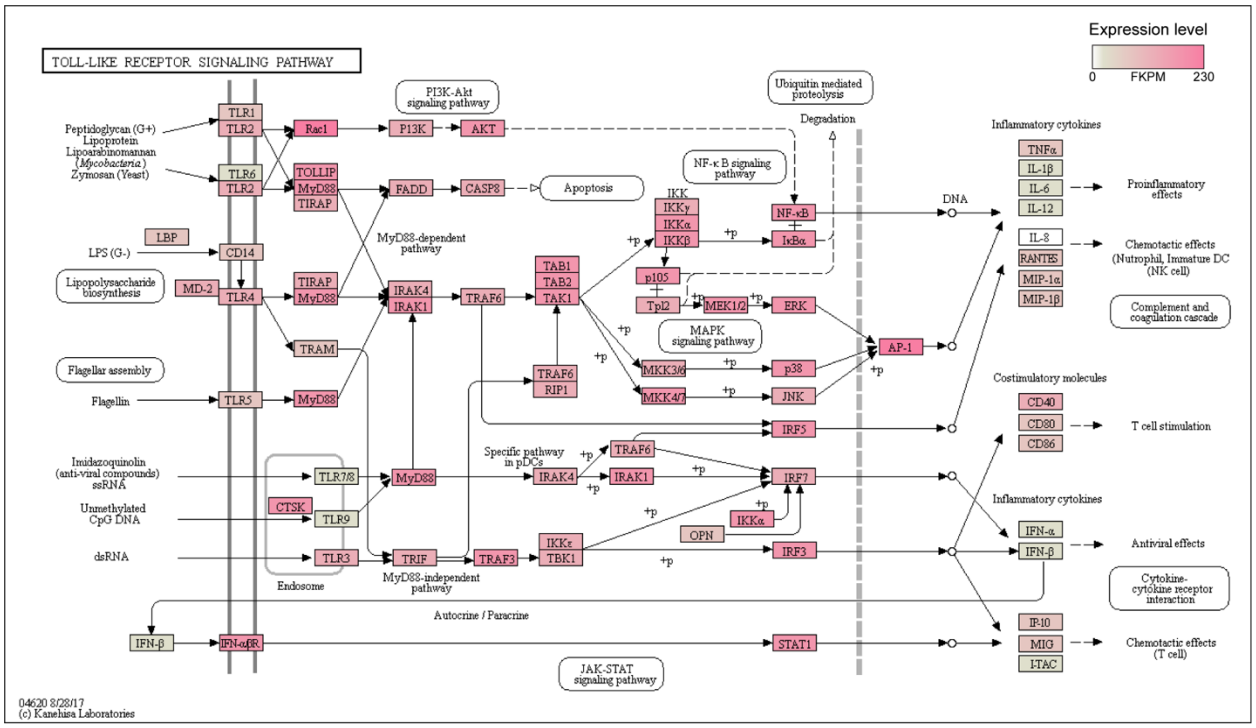


Fig. S17. Murine pacemaker cells, the interstitial cells of Cajal (ICC), express immune-related TLR/MyD88 pathway components. KEGG-pathway mapping. Genes highly abundant in the transcriptome of ICC cells (21) are marked red, poorly expressed genes in gray, transcripts missing in the dataset are labeled white. Original data are presented in Data S4.

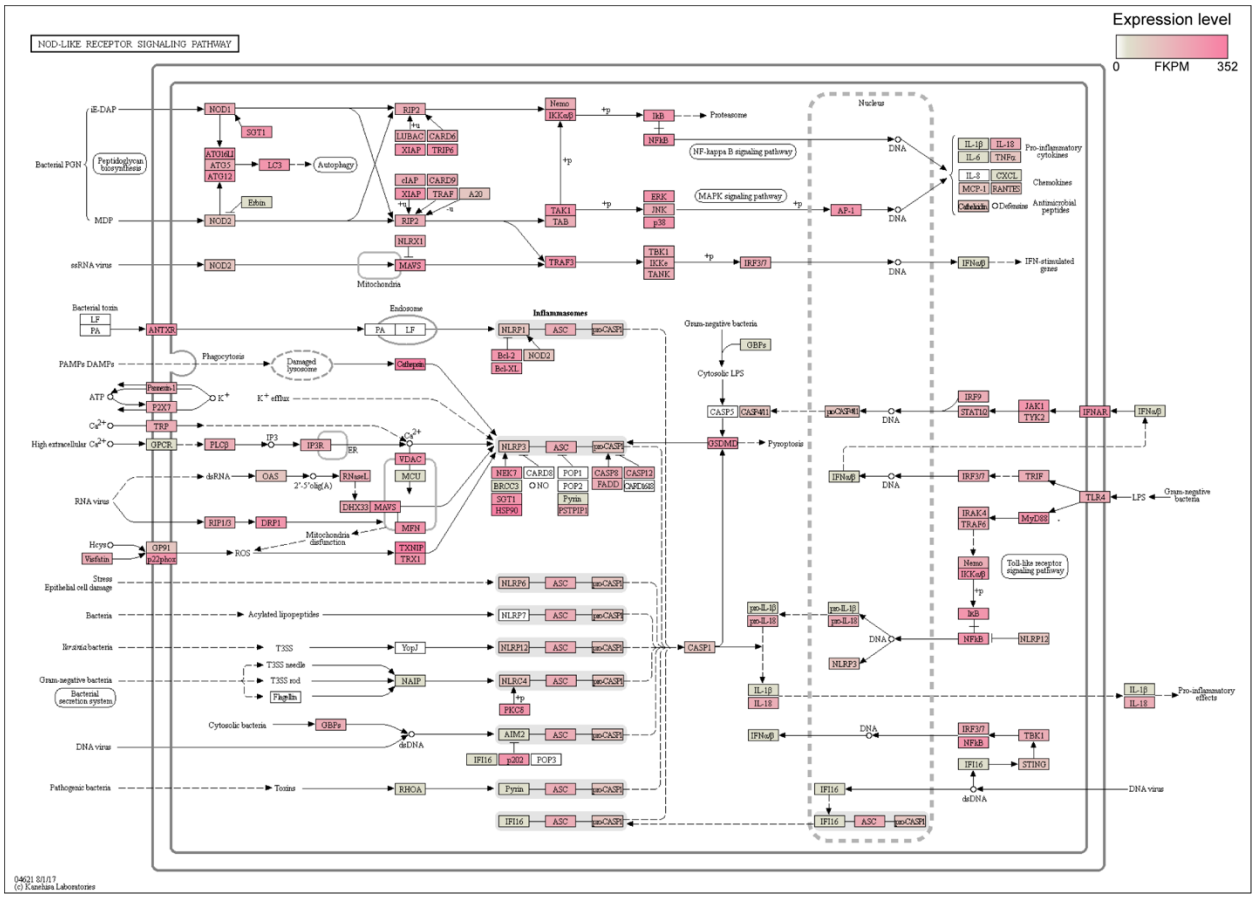
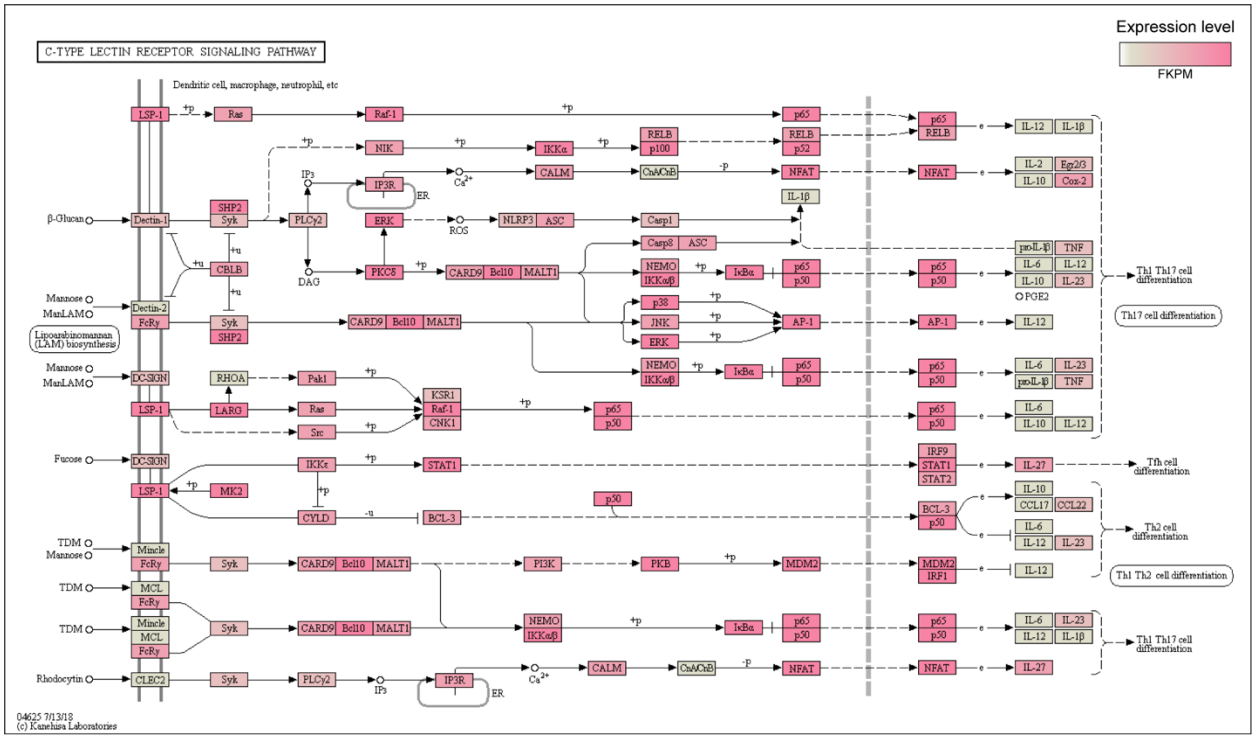


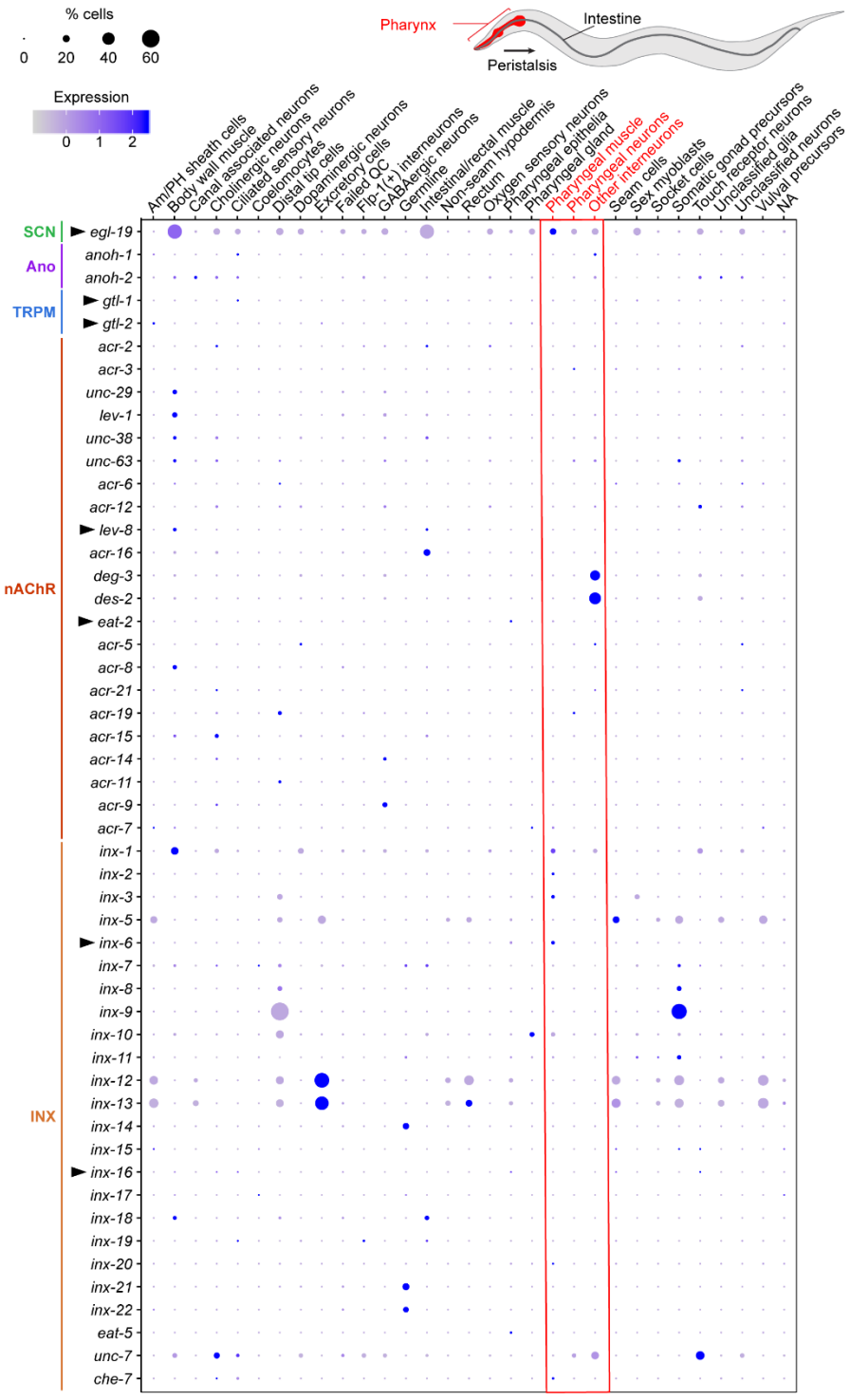
Fig. S18. Murine pacemaker cells, the interstitial cells of Cajal (ICC), express immune-related NOD-like receptor pathways. KEGG-pathway mapping. Genes highly abundant in the transcriptome of ICC cells (21) are marked red, poorly expressed genes in gray, transcripts missing in the dataset are labeled white. Original data are presented in Data S4.

Fig. S19



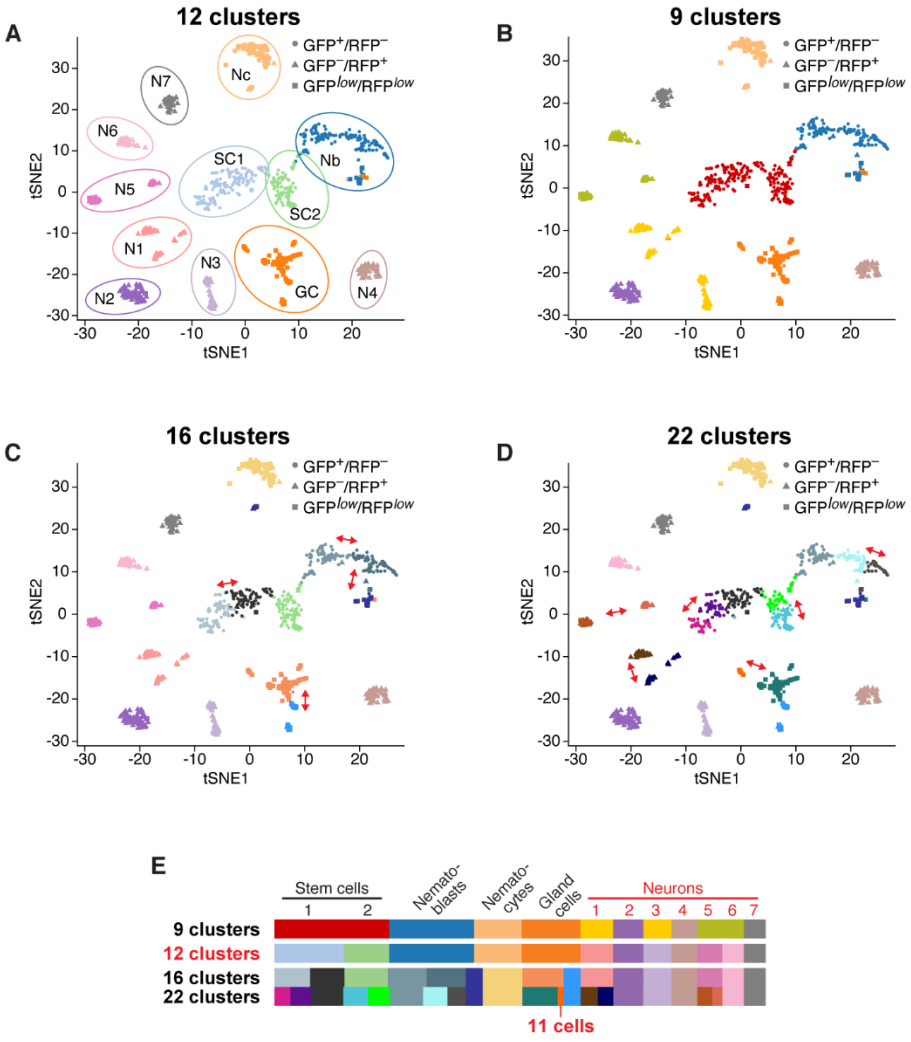
Murine pacemaker cells, the interstitial cells of Cajal (ICC), express immune-related C-type lectin pathway components. KEGG-pathway mapping. Genes highly abundant in the transcriptome of ICC cells (21) are marked red, poorly expressed genes in gray, transcripts missing in the dataset are labeled white. Original data are presented in Data S4.

Fig. S20


% cells
0
20
40
60
Expression
0
1
2
Pharynx
Intestine
Peristalsis
Am/PH sheath cells
Body wall muscle
Canal associated neurons
Cholinergic neurons
Ciliated sensory neurons
Coelomocytes
Distal tip cells
Dopaminergic neurons
Excretory cells
Failed QC
Flp-11(+) Interneurons
GABAergic neurons
Germline
Intestinal/rectal muscle
Non-seam hypodermis
Rectum
Oxygen sensory neurons
Pharyngeal epithelia
Pharyngeal gland
Pharyngeal muscle
Pharyngeal neurons
Other interneurons
Seam cells
Sex myoblasts
Socket cells
Somatic gonad precursors
Touch receptor neurons
Unclassified glia
Unclassified neurons
Vulval precursors
NA
SCN
egl-19
Ano
anoh-1
anoh-2
TRPM
gtl-1
gtl-2
nAChR
acr-2
acr-3
unc-29
lev-1
unc-38
unc-63
acr-6
acr-12
lev-8
acr-16
deg-3
des-2
eat-2
acr-5
acr-8
acr-21
acr-19
acr-15
acr-14
acr-11
acr-9
acr-7
INX
inx-1
inx-2
inx-3
inx-5
inx-6
inx-7
inx-8
inx-9
inx-10
inx-11
inx-12
inx-13
inx-14
inx-15
inx-16
inx-17
inx-18
inx-19
inx-20
inx-21
inx-22
eat-5
unc-7
che-7

Pacemaker-specific genes are expressed in the neuro-muscular pacemaker system in *C. elegans* pharynx. The plot illustrates expression of *C. elegans* homologues of the genes that comprise the pacemaker signature in both *Hydra* and mouse (coding for SCN, ANO and TRPM ion channels, nAChR receptors and innexin gap junction proteins) in 27 cell types of *C. elegans* identified by Cao and co-authors (22). The pharynx of *C. elegans* represents a neuro-muscular unit with pacemaker activity that controls regular peristaltic movements of the intestine (scheme above). Pharyngeal neurons, pharyngeal muscles and other neurons (outlined) express multiple genes of the pacemaker signature. Genetic analysis identified that mutants for at least seven *C. elegans* pacemaker-relevant genes (*egl-19*, *gtl-1* and *-2*, *lev-8*, *eat-2*, *inx-6* and *-16*) are characterized by disrupted pharyngeal pumping and constipation, providing evidence that these genes are functionally indispensable for the pacemaker function of the pharynx.

Fig. S21



Different hierarchical clustering strategies used in the initial phase of the project. In our study, we used the clustering scheme that partitioned 928 sequenced cells into 12 clusters, including 7 populations of neurons (on A), since it is most consistent with our previous observations and the literature data. It results in clusters composed of at least 40 cells each and was functionally validated in our study. Independently on the settings, the neuronal clusters are the most robust. For instance, cluster N2, N4 and N7 remain not affected in all clustering schemes. Reducing the number of clusters, for instance to 9 (on

B and E) clearly compromises the accuracy of population identification and contradicts the existing observations. Increasing the clustering depth to 16 (on C and E) or 22 clusters (on D and E) results in partitioning the stem-cell and nematoblast populations in smaller cell groups most likely corresponding to different cell cycle and differentiation states. Only if the SEURAT algorithm is forced to generate 22 clusters, the neuronal clusters start splitting. However, by this moment some clusters of non-neuronal cells are as small as 11 cells (on E), and the smallest neuronal cluster becomes 19 cells. We consider this cluster size as critically small and do not attempt to obtain even more clusters. Overall, the total amount of cells sequenced (only 928) does not allow us to achieve higher clustering depth.

Table S1

Oligonucleotide primers used to amplify gene fragments in qRT-PCR.

Target gene	**Forward primer 5' -> 3'**	**Reverse primer 5' -> 3'**
actin	gaatcagctggtatccatgaaac	aacattgtcgtaccacctgatag
elongation factor 1	gcagtactggtgagtttgaag	cttcgctgtatggtggttcag
cluster43524	*tggtggtgatgcgctgag*	*tcgatcgactttgtttgaccc*
cluster63380	*tgcttgcacgggttggac*	*gtagtttcaggtagacacagg*
cluster28804	*ctgagcagcacgatcgtg*	*aggcgacattgctggctc*
cluster2505	*gctccattagcgctacgtc*	*aagtagtcaggaggactcatag*
cluster32301	*gaccaagggattcaagaagag*	*ccatctcgaaagtaagtccc*
cluster74576	*tcaaactgatcgctcaagaattg*	*ccatcatgaacaggataacaatc*
cluster11976	*aaacgaagagacaagcttagttg*	*tctgttggtgtctcttgtgg*
cluster128380	*aagttcgcccagcgatgc*	*ctgaccaacccatacattactg*
cluster159873	*ggagaagacaccaatgacgtc*	*gatcgaggcaaagtattgggag*
cluster32173	*gacactgtgcctctacttgtc*	*aacactacgcacccaacctg*
cluster39951	*tgggaagacgttgttcgctc*	*gtctttggtggaaacgcgaag*
cluster41630	*cacggcaaaatagcggaagg*	*acgtgtgcgaaatcccttgc*
cluster30836	*ccgtaccattgtctgggaga*	*tgcgaatgacgaacgtgtgc*
cluster197945	*caactataacggtaacttgagcg*	*tttcatttggcagtggtccaac*
cluster165552	*tacggaagcccacagtacg*	*acaaaaaactgttttccgcacgg*

Data S1 (separate file)

Dataset of 364 transcripts coding for putative transcription factors.

Data S2 (separate file)

Dataset of top 300 transcripts expressed in each of 12 populations within the interstitial cell lineage. Cluster numbers, best hit annotation using UniProt and RefSeq databases, predicted peptide length and domains, signal peptide and cellular localization.

Data S3 (separate file)

Dataset of putative secreted antimicrobial peptides encoded by neuron-specific TRGs. Predicted membrane destabilizing activity estimated using machine learning tool with a moving window of 20 amino acids.

Data S4 (separate file)

Expression of genes coding for components of immune-related TLR/MyD88 pathways, NOD-like and C-type lectin receptors in the murine intestinal pacemaker cells, interstitial cells of Cajal (ICC).

Data S5 (separate file)

Dataset of 25 transcripts comprising the proliferation signature used for annotation of cell clusters.

Data S6 (separate file)

Dataset of 24 transcripts of cell-type specific marker genes used for annotation of cell clusters.

Data S7 (separate file)

Dataset of 112 transcripts coding for putative neurotransmitter receptors.

Data S8 (separate file)

Dataset of 431 transcripts coding for putative ion channels.

Data S9 (separate file)

Homologues of immune-related genes from *C. elegans*, *Drosophila* and molluscs in Hydra.

SI References

1. 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).
2. S. Picelli, *et al*., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096 (2013).
3. B. M. Mortzfeld, *et al*., Temperature and insulin signaling regulate body size in Hydra by the Wnt and TGF-beta pathways. *Nat. Commun*. **10**, 3257 (2019).
4. A. Dobin, *et al*., STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**, 15–21 (2013).
5. L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
6. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
7. A. L. Shaffer, *et al*., Signatures of the immune response. *Immunity* **15**, 375–385 (2001).
8. T. N. Petersen, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785 (2011).
9. J. Garnier, J.-F. Gibrat, B. Robson, "GOR method for predicting protein secondary structure from amino acid sequence" in *Methods in Enzymology*, (Elsevier, 1996), pp. 540–553.
10. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep

learning. *Bioinformatics* **33**, 3387–3395 (2017).

11. I. Letunic, P. Bork, 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. **46**, D493–D496 (2017).

12. E. Y. Lee, B. M. Fulan, G. C. L. Wong, A. L. Ferguson, Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proc. Natl. Acad. Sci*., **113**, 13588–13593 (2016).

13. T. Fujisawa, Hydra peptide project 1993-2007. *Dev. Growth Differ*. **50**, S257–268 (2008).

14. R. Augustin, et al., A secreted antibacterial neuropeptide shapes the microbiome of Hydra. *Nat. Commun.* **8**, 698 (2017).

15. S. Siebert, *et al.*, Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science.* **365**, eaav9314 (2019).

16. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

17. S. Kumar, G. Stecher, K. Tamura, S. Q. Le, O. Gascuel, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

18. Y. Moran, M. G. Barzilai, B. J. Liebeskind, H. H. Zakon, Evolution of voltage-gated ion channels at the emergence of Metazoa. *J. Exp. Biol.* **218**, 515–525 (2015).

19. M. Oz, E. G. El Nebrisi, K.-H. S. Yang, F. C. Howarth, L. T. Al Kury, Cellular and Molecular Targets of Menthol Actions. *Front. Pharmacol.* **8**, 472 (2017).

20. Y. Takaku, *et al.*, Innexin gap junctions in nerve cells coordinate spontaneous

contractile behavior in Hydra polyps. *Sci. Rep.* **4** (2014).

21. M. Y. Lee, *et al.*, Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures. *PLoS One* **12**, e0176031 (2017).

22. J. Cao, *et al*., Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).