

Whole-body single-cell sequencing of the *Platynereis* larva reveals a subdivision into apical versus non-apical tissues

Kaia Achim^{1,#}, Nils Eling^{2,4,#}, Hernando Martinez Vergara¹, Paola Yanina Bertucci¹, Thibaut Brunet¹, Paul Collier⁵, Vladimir Benes⁵, John C Marioni^{2,3,4} and Detlev Arendt¹

¹Developmental Biology Unit, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

²EMBL-European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge CB10 1SD, UK

³Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK

⁴Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK

⁵Genomics Core Facility, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

These authors contributed equally to this work.

Correspondence:

Detlev Arendt: arendt@embl.de

John Marioni: marioni@ebi.ac.uk

Abstract (127 words)

Animal bodies comprise a diverse array of tissues and cells. To characterise cellular identities across an entire body, we have compared the transcriptomes of single cells randomly picked from dissociated whole larvae of the marine annelid *Platynereis dumerilii*¹⁻⁴. We identify five transcriptionally distinct groups of differentiated cells that are spatially coherent, as revealed by spatial mapping⁵. Besides somatic musculature, ciliary bands and midgut, we find a group of cells located at the apical tip of the animal, comprising sensory-peptidergic neurons, and another group composed of non-apical neural and epidermal cells covering the rest of the body. These data establish a basic subdivision of the larval body surface into molecularly defined apical versus non-apical tissues, and support the evolutionary conservation of the apical nervous system as a distinct part of the bilaterian brain⁶.

Introduction

Animal bodies are composed of cells, which are organized into tissues and organs. Recently, a number of studies have applied single cell transcriptomics to assess cellular diversity within tissues, such as the vertebrate pancreas⁷, intestine⁸, or different brain parts such as cortex, basal ganglia or hypothalamus⁹⁻¹¹. This approach allows the molecular characterisation of cell types within a given tissue, as well as an assessment of their heterogeneity. However, the sheer number of cells in conventional vertebrate or insect systems has so far hampered the comparison of cellular identities across tissue boundaries, or even across entire bodies.

Other laboratories have focused on the molecular comparison of different tissues, referred to as tissue transcriptomics. This approach determines and compares the expression profiles of entire tissues via bulk RNAseq^{12,13,14,15}, with the potential to compare across the entire body. Inherent to the approach however, tissue transcriptomics so far refer to the somewhat arbitrary, morphology-based dissection of the body into discernible units. This is relevant because, on the one hand, cell types of a given tissue or organ may be very divergent and on the other hand, similar cell types can populate different tissues. Therefore, while tissue transcriptomics allow in-depth and technically robust analyses and comparisons of larger cellular assemblies, these are somewhat arbitrarily defined and the method lacks the resolution to disentangle and molecularly define cell and tissue identities across entire bodies.

To advance on this, we establish the marine annelid *Platynereis dumerilii*, a molecular model species for development, evolution and neurobiology^{16,17}, as experimental paradigm to explore how cellular identities compare and relate to each other across an entire animal body. We apply single-cell RNAseq to randomly sampled cells from the dissociated whole larvae at 48 hours post fertilization (hpf). The *Platynereis* larval body is especially suited for this task, as it already comprises various cell types and tissues. Besides larval body features (e.g., epidermal cells and ciliary bands) it contains early differentiating cells contributing to the later juvenile worm (musculature, brain and nerve cord)¹⁸. It thus allows us to determine and compare cellular identities at an early stage of the life cycle when cell numbers are still relatively low.

Our whole-body analysis reveals that, at this stage, the larval annelid body comprises five groups of differentiated cells with distinctive expression profiles. In each of these groups,

cells express sets of transcripts that together encode group-specific cellular modules – representing cellular structures and functions that characterize these cells. While some of these groups and their defining modules, such as ciliary bands with motile cilia or larval musculature with striated myofibres, match larval morphology, others shed new light on the basic organization of the annelid body. We find a group of cells located at the apical tip of the animal that appears to be distinct from all other ectodermally-derived, non-apical cells of the body. While the apical cells specifically utilize photosensory and neurosecretory modules, the non-apical tissue instead shows specific expression of various extracellular matrix components – such as ligand-receptor pairs, cellular adhesion molecules, and conserved proteolytic enzymes. Comparative data indicate that this molecular subdivision into apical versus non-apical tissues is an important organisational feature of the body that appears to be conserved in animal evolution.

Results

Single-cell RNA-seq identifies five groups of differentiated cells

To explore cell type diversity on the whole organism level, we dissociated whole larvae of a marine annelid, *P.dumerilii* at 48 hpf and randomly captured cells for single-cell RNA-sequencing (scRNA-seq) (**Fig. 1**). At this stage of development, the larva comprises of relatively few cells (~5000), but many differentiated cell types such as ciliated cells, neurons or myocytes are already present. The collected cells were optically inspected to exclude doublets, multiple cells or cell debris. Sequenced samples were further filtered computationally to remove low complexity transcriptomes, lowly-

expressed genes and transcriptomic doublets (**Extended Data Fig. 1**, Methods). A total of 373 cells and 31300 transcripts passed filtering steps and were used for downstream analysis. To group the cells into distinct clusters we used a sparse clustering strategy, which identified seven groups of cells. We used the *scrn* package to find group specific marker genes and discovered that in pairwise comparisons across all groups, two clusters were consistently highly similar to one other. Therefore, we merge these two closely related groups (**Fig. 1**, **Extended Data Fig. 2**, Methods).

To characterize the remaining six groups further, we identified differentially expressed genes (Methods). The largest group of cells was characterized by the specific expression of genes known to be active in developmental precursors, such as DNA replication (*DNA topoisomerase*, *Rfc3*, *Rfc5*, *Mcm3*, *Mcm4*, *Mcm7*), cell proliferation (*Pcna*, *Md211*, *Apc7*, **Extended Data Fig. 2i**), or chromatin remodeling (*Nucleoplasmin*, *Bptf*). We thus inferred that this group was comprised of undifferentiated cells.

Cells in the other five groups showed significantly lower expression of these markers (**Extended Data Table 1**, FDR < 0.1, Methods) and we thus consider them differentiated cells. For each of these five groups, pairwise differential expression analysis revealed distinct sets of group-specific effector genes (i.e., differentially expressed genes that encode the particular structural and/or functional properties of cells) and group-specific transcription factors, many of which are known to act as terminal selectors in cell type differentiation^{19,20} (**Extended Data Table 2**).

To validate the clustering of differentiated cells into the above groups, and to infer relationships between these groups, we used bootstrap resampling (Methods) to calculate a hierarchical tree (**Extended Data Fig. 3**). This tree confirmed each group's integrity,

albeit with variable support. We then averaged the gene expression across cells in each group and used these values to calculate a tree comprising cell groups only (**Fig. 3**, upper panel). The resulting tree retrieved the group topology of the single cell tree with higher support values.

Spatially coherent, molecularly defined larval tissues

To characterize and localize the five distinct groups of differentiated cells with regard to larval morphology, we used a dual strategy. First, we investigated the spatial distribution in the larval body by mapping constituent cells into a cellular-resolution expression atlas⁵ (**Fig. 2**). To this end, we constructed a cellular resolution atlas of the 48 hpf larva (**Extended Data Fig. 4**), taking advantage of the existing Profiling by Signal Probability mapping (ProSPr) pipeline²¹. Using ProSPr, the majority (95%) of cells could be mapped to distinct regions within the larva. Complementing this, we used wholemount in situ hybridization (WMISH) with probes targeted towards group-specific transcripts to refine the spatial mapping.

Following this strategy, cells from group 1 mapped to the most apical part of the larval ectoderm around the so-called apical organ (**Fig. 2a-b**). Consistent with this, the specific group 1 marker *Phosphodiesterase-9* (*Pde9*) outlined a highly specific expression territory in the apical/dorsal episphere (**Fig. 3a**). With reference to the clear neural character of the cells in this group (see below), we refer to group 1 cells as ‘apical nervous system’. Group 2 cells mapped into the yolk-rich macromeres where the first differentiating cells of the later midgut are located (**Fig. 2c-d**), as confirmed by the highly

specific expression of *Hnf4* in the cellularized portion of the macromeres (**Fig. 3b**). This group thus represents ‘larval midgut’. Cells of group 3 mapped to the differentiating striated myocytes (**Fig. 2e-f**) and, in line with this, *Striated muscle myosin heavy chain* (*St-mhc*) was expressed in all cells belonging to ‘striated musculature’ (**Fig. 3c**). Next, group 4 cells mapped to the ciliary bands (**Fig. 2g-h**), which is composed of multiciliated cells with motile cilia. Expression of *Radial spoke head protein homolog 4* (*Rsph4*) in the multiciliated cells of the ciliary bands of the apical organ, prototroch and paratrochs confirmed this mapping (**Fig. 3d**)¹⁸. Complementing this, group 5 cells covered much of the remaining larval surface, including non-apical territories in the larval trunk (**Fig. 2i-j**). The broad expression of the *Metabotropic glutamate receptor 7* (*Grm7*) in trunk and head clearly illustrates these ‘non-apical surface cells’ (**Fig. 3e**).

Notably, while the sorting of differentiated cells into ciliary bands, larval midgut, and striated musculature matched the subdivision of the larval body into well-known, morphologically distinct tissues, the assignment of ectodermally-derived cells into apical and non-apical populations did not correspond to an obvious morphological boundary. Interestingly, however, the topology of the (unrooted) hierarchical tree (**Fig. 3, Extended Data Fig. 3**) supported the general subdivision of the larval body into apical versus non-apical tissues - suggesting a closer relationship of apical nervous system and midgut on one hand, and of the non-apical tissues (non-apical surface, ciliary bands and striated musculature) on the other.

The apical nervous system

To further explore the different molecular nature of the five differentiated cell groups , we investigated their distinctive gene sets. **Fig. 4** shows a collection of the most specific and informative regulatory and effector genes specific for each group.

Cells of the apical group consistently expressed the neuropeptide-cleaving *Prohormone convertase 2 (Phc2)*²² and many neuropeptides, such as the broadly expressed *GNXQNpeptide* or *GGNamide*, two lophotrochozoan neuropeptides²³, indicating neurosecretory release from these cells. We also noted that cells in the apical group highly expressed non-calcium-binding members of the synaptotagmin family implicated in the generation and fusion of large dense core vesicles for neurosecretion, such as *Synaptotagmin17/B/K (Syt17)*, *Syt α* and *Syt4* (**Fig. 4a-b, Fig. 5, Extended Data Fig. 5**)²⁴⁻²⁶. This is in line with previous reports of neurosecretory cells populating the dorsomedian larval brain²², and with ultrastructural observations, which show that cells in the dorsomedian *Platynereis* brain are rich in dense core vesicles while synapses are often sparse or absent²⁷.

In addition, the bistable photopigment *Peropsin*²⁸ and components of cGMP-based signal transduction such as *Pde9* and the CNG channel components *Cngb*, *Cnga- α* were broadly expressed in the cells of the apical group, indicative of light reception via vertebrate-type phototransduction²⁹ (**Fig. 4a**). Specific cells within this group co-expressed effector genes and transcription factors, such as *c-opsin* or *c-opsin2* and *Neuropeptide Y (Npy)* (representing brain ciliary photoreceptors²⁹), or *Vasotocin-neurophysin*, *Proenkephalin* and *c-opsin* (representing photosensory-vasotocinergetic cells²²). Adding to this, some apical cells expressed the neural effector gene *Neuroglobin*, encoding a monomeric globin evolutionarily older than hemoglobin or myoglobin that reversibly binds oxygen

and may represent an oxygen reservoir for highly metabolic neurons (**Fig. 4a**)³⁰. Other cells expressed *Insulin-like peptide*, *Pigment dispersing factor (Pdf)* or *Allatostatin-C*, and subsets of cells expressed the transcription factors *Foxq2*, *Bsx* and *T-brain (Tbr1)*. In sum, we conclude that the apical ectodermal cells form part of the photosensory-neurosecretory dorsomedian brain^{4,22} and correspond to the ‘apical nervous system’ (ANS) as defined previously^{4,6,27,31}.

Midgut cells and striated musculature

Cells allocated to the early differentiating midgut specifically expressed the midgut-specific transcription factor *Hnf4* as well as *Chymotrypsin-A*, a digestion enzyme (**Fig. 4b**). In addition, this group was enriched in the expression of signal peptides such as *CCWamide*, *PYpeptide*, *WLDpeptide*, indicating that peptidergic neurosecretion is a characteristic feature of this group, similar to the apical nervous system (but using different neuropeptides).

The striated musculature cells were characterized by the relatively higher expression of *St-mhc*, *Mrlc*, *Troponin-I* and *Troponin-T*, which together are known to assemble striated myocyte contractile fibres³² (**Fig. 4c**), and by the heterogeneous expression of the myocyte-specific terminal selectors *MyoD*, *Hand*, *Paraxis*, *Mox* and *Pitx*³³. We used WMISH and co-expression analysis using ProSPR to validate the expression of *MyoD* in a subset of muscles (**Extended Data Fig. 5**). Notably, we detected a low level of the striated myocyte markers in other cell groups (**Fig. 4c**). Given the exceptionally strong expression of these genes in the myocytes, and their up to 100-fold lower expression in the other groups, it is yet unclear whether this finding represents biological or technical noise.

Non-apical surface cells and ciliary bands

The cells of ectodermally-derived, non-apical surface and the ciliary band cells expressed different, partially overlapping gene sets indicative of mixed epithelial and neural characteristics (**Fig. 4d**). First, we identified a set of genes broadly co-expressed by non-apical surface ectoderm, ciliary bands and - consistently but at lower level – in striated musculature. These genes included *LaminaA/C*, a nuclear fibrous protein forming the nuclear lamina on the interior of the nuclear envelope³⁴ (**Fig. 4d, Extended Data Fig. 5**), and the related *Neurofilament-70 (Nf70)*, a neural intermediate filament found in Lophotrochozoans (**Fig. 4d**). Interestingly, Lamins and cytoplasmic intermediate filaments evolved via duplication from an ancient, Lamin-like gene (with the intermediate filament precursor losing the nuclear localization sequence and the CaaX motif characteristic for Lamins, and giving rise to intermediate filaments and keratins several times independently³⁵). In addition, this group comprised the conserved *Extended Synaptotagmin 2 (Esyt2)*, with an apparent function in membrane lipid composition dynamics, extracellular signal transduction and cell adhesion³⁶; and finally, *P4ha2* encodes a component of prolyl 4-hydroxylase, a key enzyme in collagen synthesis (**Fig. 4d**).

Another gene set, shared between non-apical surface cells and ciliary bands, was enriched for extracellular matrix activities. For example, orthologs of *Anoctamin*, a transepithelial chloride channel³⁷, *Protocadherin-15 (Pcdh15)*, an atypical cadherin mediating structural integrity of ciliated sensory receptors³⁸ (**Extended Data Fig. 5**), and *Semaphorin 2 (Sema2)*, a secreted extracellular guidance molecule involved in axon pathfinding³⁹ were

co-expressed in these groups (**Fig. 4d**). Most noticeably, however, several genes encoded matrix-modifying proteases such as a serine protease *Plasminogen* related to vertebrate Plasmin⁴⁰ and Hepatocyte Growth Factor (HGF)⁴¹ (**Extended Data Table 2**); metalloproteinases such as *Matrix Metalloproteinase (Mmp9)*⁴² together with its specific substrate *Cd109*⁴³; and the cysteine proteinase *Cathepsin L*⁴⁴ (**Fig. 6b**). Regarding transcription factors, these groups shared the epidermis-specific transcription factor *Grainyhead*, known for its conserved role in maintaining the epidermal barrier⁴⁵, and the *ETS-homologous factor* that likewise specifies epithelial cell type identities⁴⁶ (**Fig. 4d**).

Likewise, the genes most specific for non-apical surface cells abundantly encoded proteins known to specifically interact in extracellular matrix and basal lamina (**Fig. 4d, Fig. 6c**) – many of which with reported roles in nervous system development and plasticity. For example, we identified orthologs of *Neurotactin*, which encodes a cell surface glycoprotein of the serine esterase superfamily⁴⁷, and its specific binding partner *Amalgam*⁴⁸; both involved in axonal pathfinding in *Drosophila*^{47,49}. We also found two orthologs of *Neurotrypsin* (**Fig. 6a**), a matrix-modifying serine protease, together with the extracellular proteoglycan *Agrin*⁵⁰, indicating that the Neurotrypsin-Agrin system is active in these cells⁵¹ (**Fig. 4d**). The *Hspg2* gene, encoding a conserved proteoglycan related to Agrin, was also present. Besides extracellular matrix components, the non-apical surface cells also expressed *Wnt* ligands such as *Wnt4*⁵², and the bHLH HES transcription factors *Hes1*, *Hes2*, *Hes6* and *Hes11*. Homeodomain transcription factors with conserved roles in neuronal specification such as *Uncx4*, *Tbx20* or *Phox2*^{2,53,54} were expressed in subsets of cells in this group (**Fig. 4d**), indicating neuronal identities.

Finally, we also identified a group of genes most specific for the ciliary bands: *Rsph*, *Ift*, *Dynein* and *Kinesin* effector genes, which are constituent parts of motile cilia (**Fig. 4e**, **Extended Data Fig. 5**, **Extended Data Table 2**)^{55,56}. Notably, all of these genes were also expressed in other groups, albeit in a patchy manner and at lower level. While most of the ciliated cells mapped to the larval trunk, two of the ciliary band cells mapped to the head and co-expressed the transcription factors *Rx* and *Tailless*, which specify head regions and cell types in brain and eye in a broad spectrum of bilaterians⁵⁷. This suggest that the ciliary band cells may split into distinct subgroups corresponding to the ciliary cells of the head vs trunk.

Distributed nature of neural cells

Interestingly, scRNAseq indicated that pan-neuronal differentiation genes such as *Synaptotagmin1*, *Rab3*, *Munc13*, *Rims*, *Rimsbp* and *Complexin*⁵⁸ were expressed across all groups of differentiated cells (**Fig. 7b**). This is consistent with observations that neuronal features are not restricted to the nervous system⁵⁸. For *Platynereis*, the ciliated cells of the larval prototroch are reported to show action potentials⁵⁹ and therefore likely to also express neuronal markers; and the *Platynereis* differentiated midgut has recently been show to show autonomous contraction waves at juvenile stages⁶⁰.

Upon examining the spatial expression pattern of pan-neuronal genes using WMISH we noted that their expression was highest in the apical nervous system and in the ventral portion of the non-apical, ectodermally-derived tissue (**Fig. 7a**), consistent with the known location and extent of the neuroectoderm at this stage². With regard to the apical

nervous system, a large fraction of its constituent cells indeed mapped in to the apical domain of *Syt1* expression (**Extended Data Fig. 6a**). Substantiating this, we validated the co-expression of the pan-neuronal marker genes and apical cell group markers by WMISH (**Fig. 7a**).

With regard to the non-apical surface cells, spatial mapping showed that some of those cells also mapped into the *Syt1* expression domain on the ventral body side (**Extended Data Fig. 6b**), supporting their neuronal character. ProSPr analysis likewise confirmed an overlap of non-apical surface markers with the pan-neuronal marker *Rab3* (**Fig. 7a**). Adding to this, we validated the expression of the non-apical group marker *Neurotrypsin* in the ventral neuroectoderm by WMISH (**Fig. 6a**). Together, these data indicate that, at least at this larval stage, ‘neurons’ did not fall into one homogenous group (in contrast to striated myocytes or cells with motile cilia).

Discussion

For this study we analyzed the transcriptomes of 373 cells randomly collected from dissociated, entire larvae of the marine annelid *Platynereis dumerilii*. Among these we identified 123 differentiated cells, chosen for further analysis and comparison. Previous studies had captured cells from specific body parts and tissues only (such as the mouse forebrain^{9,10}); our data is unique as it allows single-cell level comparison of differentiated cells across the entire body. We were able to sort all differentiated cells into five major groups, based on similarities and differences in gene expression profiles. Spatial mapping into the 48 hpf ProSPr cellular-resolution expression atlas, in combination with

wholemound in situ hybridisation of marker genes, identified these groups as apical nervous system, larval midgut, ciliary bands, non-apical surface cells, and striated musculature.

What is the nature of these groups? Interestingly, within each of the five groups, there existed considerable heterogeneity in the expression of transcription factors and effector genes (as evident from **Fig. 4**), suggesting the presence of several cell types within each group. For example, cells of the apical nervous system differ with regard to opsin expression, neuropeptide content, and signaling components – indicating that each, or almost each of the cells might represent a separate type. Therefore, at the current level of analysis the five groups of differentiated cells rather appear to represent some kind of molecularly defined larval tissues – uniting cells with similar molecular properties shared beyond the cell type level. We believe that the number of sequenced cells would have to be increased by at least one order of magnitude to resolve individual cell types within these groups.

Comparing the cell groups further, we noted an unexpected divide between ectodermally-derived cells in the most apical part of the larva (apical nervous system), *versus* the non-apical surface cells and ciliary bands. Cells of the apical nervous system shared the expression of two major sets of modules, enabling the synthesis and secretion of neuropeptides, and mediating the detection of ambient light for the circadian control of body physiology. Most of the specifically expressed effector genes detected for these modules, encoding for example prohormone convertase, specific neuropeptides, components of the phototransducing cascade, and ciliary opsins, have previously been implicated in apical nervous system function^{4,22}.

In contrast, the genes characteristic for the non-apical surface cells revealed mixed epithelial and neural properties. One broadly expressed set of genes – found in non-apical surface, ciliary band and striated muscle cells – encoded proteins relating to basic properties of the surface epidermis – such as stabilizing intermediate filaments (related to vertebrate keratins), and an enzyme mediating collagen synthesis into the extracellular matrix. Another set of genes – shared between non-apical surface and ciliary band cells, but absent from myocytes – included matrix-modifying proteinases, many of which implicated in tissue remodeling, axonogenesis and synapse formation in insects and vertebrates⁴⁰. And yet another set – identified as most specific for the non-apical surface cells proper – included genes encoding neural properties, such as *Neurotactin/Amalgam*, *Neurotrimin*, and *Semaphorin* involved in axon guidance and *Neurotrypsin/Agrin* with reported roles in synaptogenesis^{51,50}. This would indicate that the non-apical surface cells utilize specific molecular toolkits to modify extracellular matrix, guide outgrowing axons, and establish synapses – that are not employed in the apical nervous system.

The specific and broad expression of the nicotinic acetylcholine receptor alpha-9 in non-apical surface cells is also noteworthy. In vertebrates, this unconventional receptor plays restricted roles in the mechanosensory hair cells, where it is involved in the reception of efferent signals⁶¹, and in epidermal keratinocytes, where it triggers epithelialization via local, non-neuronal acetylcholine cytotransmission⁶².

What is the significance of the fundamental divide between apical and non-apical cells in the *Platynereis* larval surface ectoderm? While some of the genes that are differentially expressed between apical and non-apical cell groups relate to functions carried out by fully differentiated cells (e.g., neurosecretion, phototransduction, cholinergic

cytotransmission), others indicate earlier stages of differentiation (e.g., proteolytic cleavage for extracellular matrix modification, and axonal guidance). Thus, and given that our analysis only covers one single stage of larval development, we cannot rule out that some of the differences between apical and non-apical cells are due to differences in developmental timing; more extensive studies including additional stages will be required to solve this issue. However, preliminary WMISH of ‘apical’ and ‘non-apical’ marker genes already indicates that expression differences between apical and non-apical tissues are also observed at earlier developmental and later life-cycle stages (**Extended Data Fig. 7**). Developmental timing is therefore unlikely to fully account for the apical/non-apical differences.

Another possible explanation for the difference between apical and non-apical cells is evolutionary divergence. This notion assumes that an ancient subdivision of the animal body into functionally and structurally different parts occurred early in animal evolution. These segregating parts would have started to differ in epithelial properties, extracellular matrix remodeling capacities and cell-cell communication strategy. To validate and substantiate this notion, similar datasets from both, closely and remotely related species will be required; they should show similar overall molecular tissue subdivisions if these were evolutionarily conserved.

As a starting point, and in line with the second option, the restricted activity of apical nervous system marker genes in the vertebrate forebrain indicates evolutionary conservation of the apical nervous system and of its specific properties⁶. For example, in vertebrates the neurosecretory-specific *Syt4* is specifically active in the neuroendocrine hypothalamus, where it regulates release of the neuropeptide Oxytocin⁶³. *Syt17*, initially

isolated from the vasopressin-secreting supraoptic and paraventricular nuclei, also shows strong expression in the hypothalamus⁶⁴, as does *Neuroglobin*⁶⁵⁻⁶⁷. The T-box transcription factor *T-brain* is specifically expressed in the developing para- and periventricular and supraoptic hypothalamic nuclei⁶⁸; and finally, expression of *ciliary opsins* and other genes of the ciliary photosensory cascade in the vertebrate retina and pineal indicates evolutionary conservation of apical nervous system components²⁹. These similarities add to previously reported conserved features shared between vertebrate hypothalamus and annelid apical nervous system, involving specifying transcription factors⁶⁹, a neurosecretory centre releasing conserved neuropeptides and hormones^{22,27}, and non-visual ciliary photoreceptors²⁹ active in melatonin synthesis and the control of circadian behavior⁴. Our transcriptomics data thus support the evolutionary conservation of the apical nervous system as a distinct part of the bilaterian brain, specialized on the perception of ambient light for the control of body physiology via the release of neuropeptides and hormones. They are furthermore consistent with the recent 'chimeric brain' model, which stipulates that the apical nervous system evolved as an independent center of neuronal condensation in the course of bilaterian brain evolution⁶.

METHODS

Cell capture and sequencing

Platynereis larvae were cultured, the cells were dissociated, and single-cell cDNA synthesis was performed using the Fluidigm C1 Single-Cell Auto Prep System, and the sequencing libraries were prepared as previously described⁵. We used Fluidigm IFCs for mRNA-seq with capture sites for 5-10 μm and 10-17 μm to collect the cells, allowing the

capture of a full range of cell sizes in 48hf stage of *Platynereis*. In total, 9 IFC-s, each representing an independent collection of cells derived from a different batch of animals, were processed for this study. The information about the samples distribution on chips, the chip sizes, cell number per sample and the pooling of samples on sequencing lanes described in Extended Data Table 3. 100 bp paired-end sequences were generated using the Illumina HiSeq2000 platform. For one library (CN62), 75 bp paired-end sequences were generated using the Illumina NextSeq500 platform. Libraries were sequenced to an average depth of 2.1 million reads.

Data availability

The raw sequence data for the cells analysed in this study are available from ArrayExpress, accession numbers E-MTAB-2865 (see **Extended Data Table 5** for matching the cell ID-s between the E-MTAB-2865 and this study) and E-MTAB-XXXX (the accession number will be updated as the dataset is released). The R code used for analysis is available at: <https://github.com/MarioniLab/Platynereis2017>

Read alignment

FastQC⁷⁰ was used for the quality trimming of the raw sequencing reads using default settings. For read mapping, the quality-filtered sequencing reads were mapped, using bowtie2, against a *Platynereis* reference transcriptome combined from the two previously published assemblies^{5,23}. Briefly, the transcriptome assemblies were concatenated and contigs showing more than 94% identities were removed using CD-HIT⁷¹, leaving 44977 transcripts. In thus generated reference, each gene should be represented by one

transcript, reducing the problem of multiple mapping of the reads. Expression counts for each transcript were obtained using HTSeq⁷² while only considering uniquely mapped reads.

Gene annotation

To annotate genes of the two transcriptome assemblies^{5,23}, reciprocal BLAST comparison of genes sequences against the Uniref90 protein database (Arendt assembly) or against Swissprot (Jékely assembly) was performed. For each transcript, the BLAST hit with highest E-value was selected for annotation.

Quality control

Low-quality cells were removed from the dataset based on the following filtering criteria. Visual inspection of capture sites on the IFC revealed empty wells, wells with multiple cells or debris captured. Only wells with positive cell capture were further processed for sequencing. For downstream analysis, libraries marked as single cells were chosen. Cells with less than 1000 reads mapping to ERCC spike-ins, less than 100,000 mapped transcriptomic reads or less than 60% of mapped reads being allocated to the transcriptome were removed. Additionally, cells that express unusually small number of genes (< 2000) were removed. Prior to normalization, genes with more than 1,000,000 reads or less than 10 reads across all cells were removed.

Normalization

The BASiCS package⁷³ was used to normalize read counts by incorporating ERCC spike-ins for technical noise estimation. Specific ERCC spike-ins were removed if not detected in the dataset. Posterior estimates for model parameters were computed by Markov chain Monte Carlo (MCMC) simulation with 40,000 iterations.

The *BASiCS_DenoisedCounts* function was utilized to compute normalized read counts. Running the MCMC independently on the different batches confirmed similar small levels of technical noise in the data (**Extended Data Fig. 1h**),

Spatial mapping

Spatial mapping of the sequenced single-cell transcriptomes onto the *Platynereis* reference atlas was performed as described previously⁵. The *Platynereis* reference atlas is provided as **Extended Data Table 4** and the spatial mapping results for all the sequenced cells are provided in **Extended Data Table 5**. We defined functional regions within the embryo based on spatial expression patterns of known marker genes. The region in which *Phc2* is expressed comprises the apical nervous system. The muscle region of the animal is defined by the expression of *St-mhc*. *Foxj* expression defines the ciliated cells and *Hnf4* expression the midgut region. The trunk ectoderm can be described by *Uncx4* expression. To assign mapped cells to particular regions, we identified the voxel cluster with highest mapping confidence for each cell. For visualizing the mapping of individual cells, the centroids of all voxel clusters were plotted. To compute the overlap of mapped cells to spatial gene expression patterns, we focused on the cluster with highest mapping confidence for each cell. If more than 50% of these voxels also show a particular gene expression (e.g. *Syt1*), we consider the cell as falling into this gene region.

Detecting highly variable genes

The *BASiCS_DetectHVG* function in BASiCS allows the detection of highly variable genes by incorporating spike-ins to estimate expected technical variability. Testing was done after MCMC simulation with a variance threshold of 0.98 and an evidence threshold of 0.7.

Clustering

To detect clusters of cells based on the expression of highly variable genes, we used a sparse K-means clustering approach taken from the *sparcl* package in R^{74,75}. We clustered the data using $K=2, \dots, 10$ and chose $K=7$ since only small subpopulations emerged at $K=8, \dots, 10$. An elbow plot showing the averaged within-cluster sum of squares is non-informative for the choice of K . $AWSS = (\sum_k \sum_i \sum_j (x_{ijk} - \mu_{ik})^2) / m$, where x_{ijk} is the \log_{10} -transformed transcript count of gene i in cell j and cluster k , μ_{ik} the \log_{10} -transformed mean expression of gene i in cluster k and m the number of clusters. The parameters were tuned with $2 < \text{wbounds} < 100$ using 15 permutations. Genes with weights > 0.03 define the cluster-characteristic features of the data.

Removing cell doublets

Possible cell doublets were removed after clustering and marker gene detection. The percentage of group specific marker gene expression ($\#$ marker genes expressed/ $\#$ all

genes) was calculated for each cell. Cells with unusually high expression of markers non-specific for their group were removed from the dataset.

Tree building

To investigate the transcriptional similarity between cell types, we hierarchically clustered cell types before and after averaging gene expression across all cells within each group. The averaging strategy increases stability of the hierarchical clustering algorithm by averaging out dropout events in single cell data. Distances were calculated based on the Spearman dissimilarity: $\sqrt{\frac{1}{2}(1 - r_S)}$ ⁷⁶. To compute bootstrapped p-values, we used the *pvclust* package in R⁷⁷. Hierarchical clustering was performed on the log₁₀-transformed expression counts before and after averaging within each group using the spearman dissimilarity as distance, complete linkage clustering and 1000 bootstrap iterations to evaluate cluster stability.

Differential expression

To identify cell group specific enriched marker genes, we used two approaches. First, we used *scde* package⁷⁸ to compare each of the identified cell group against all the remaining samples in order to identify novel group specific genes from the unannotated portion of the *Platynereis* transcriptome. These genes were then annotated, added to the reference, and the clustering was repeated. Second, differential expression analysis was performed using the *findMarkers* function from the *scrn* package⁶². This approach uses *limma*⁷⁹ on the log₂-transformed, normalized counts. Group specific marker genes were defined as

differentially expressed genes with a FDR < 0.1 and $\log_2\text{FC} > 0$ for all pairwise comparisons. Initially, differential expression was tested between all 7 detected groups. Group 6 and group 7 consistently grouped together when hierarchical clustering (default options from the *hclust* package) was performed on the \log_2 fold changes in expression between one group and all others (**Extended Data Fig. 2**). We therefore merged group 6 and group 7 and tested differential expression between the now largest group and all other groups (**Extended Data Table 1**). In the next step, we tested differential expression pairwise between all differentiated cell groups (**Extended Data Table 2**).

***In situ* hybridization**

For *in situ* hybridization (ISH), we collected 48 hpf or 72 hpf *Platynereis* larvae. Animals were fixed overnight in 4% PFA/1.5xPBS and ISH was performed as described previously⁸⁰.

Cloning of *Platynereis* genes and ISH probe synthesis

For the cloning of *P. dumerilii* cDNA sequences, wild-type *Platynereis* RNA was isolated by Trizol/phenol/chloroform extraction method, reverse transcribed using SuperScriptIII reverse transcriptase (Life Technologies, cat. # 18080044), and amplified by PCR using TaKaRa ExTaq DNA polymerase (Clontech, cat. # RR001A) and respective gene specific primers. The resulting PCR fragments were cloned into pCRII-TOPO vector (Life Technologies, cat # K4610-20). The cloned gene sequences and the gene specific primer sequences are available upon request.

For the synthesis of ISH probes, cDNA plasmids were linearized and antisense RNA probes were transcribed using SP6 or T7 RNA polymerase (Roche, cat. #11487671001 and 10881775001, respectively) and DIG RNA-labeling mix (Roche, cat. #11277073910) or Fluorescein RNA labeling mix (Roche, cat. # 11685619910).

Microscopy and image processing

For imaging of WMISH samples for ProSP α , samples were mounted in 97% 2,2'-thiodiethanol (Sigma, cat. # 166782) and imaged on a Leica TCS SP8 confocal microscope, using a combination of fluorescence and reflection microscopy⁸¹. The colocalization analyses and image post-processing was performed using Fiji⁸² software. The colorimetric WMISH samples were imaged on Zeiss AXIO Imager M1 microscope. The figure panels were compiled using Adobe Illustrator and Adobe Photoshop software.

REFERENCES

- 1 Tomer, R., Denes, A., Tessmar-Raible, K. & Arendt, D. Cellular resolution expression profiling reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* **142**, 800-809 (2010).
- 2 Denes, A. S. *et al.* Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell* **129**, 277-288, doi:10.1016/j.cell.2007.02.040 (2007).
- 3 Randel, N. *et al.* Neuronal connectome of a sensory-motor circuit for visual navigation. *Elife* **3**, doi:10.7554/eLife.02730 (2014).
- 4 Tosches, M. A., Bucher, D., Vopalensky, P. & Arendt, D. Melatonin signaling controls circadian swimming behavior in marine zooplankton. *Cell* **159**, 46-57, doi:10.1016/j.cell.2014.07.042 (2014).
- 5 Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* **33**, 503-509, doi:10.1038/nbt.3209 (2015).
- 6 Tosches, M. A. & Arendt, D. The bilaterian forebrain: an evolutionary chimaera. *Curr Opin Neurobiol* **23**, 1080-1089, doi:10.1016/j.conb.2013.09.005 (2013).
- 7 Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst* **3**, 385-394 e383, doi:10.1016/j.cels.2016.09.002 (2016).
- 8 Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251-255, doi:10.1038/nature14966 (2015).
- 9 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).
- 10 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-346, doi:10.1038/nn.4216 (2016).
- 11 Romanov, R. A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci* **20**, 176-188, doi:10.1038/nn.4462 (2017).
- 12 Ramskold, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol* **5**, e1000598, doi:10.1371/journal.pcbi.1000598 (2009).
- 13 Breschi, A. *et al.* Gene-specific patterns of expression variation across organs and species. *Genome Biol* **17**, 151, doi:10.1186/s13059-016-1008-y (2016).
- 14 Henrichsen, C. N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* **41**, 424-429, doi:10.1038/ng.345 (2009).
- 15 Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-348, doi:10.1038/nature10532 (2011).
- 16 Williams, E. A. & Jekely, G. Towards a systems-level understanding of development in the marine annelid *Platynereis dumerilii*. *Curr Opin Genet Dev* **39**, 175-181, doi:10.1016/j.gde.2016.07.005 (2016).
- 17 Fischer, A. & Dorresteyn, A. The polychaete *Platynereis dumerilii* (Annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *Bioessays* **26**, 314-325, doi:10.1002/bies.10409 (2004).

- 18 Fischer, A. H., Henrich, T. & Arendt, D. The normal development of *Platynereis dumerilii* (Nereididae, Annelida). *Front Zool* **7**, 31, doi:10.1186/1742-9994-7-31 (2010).
- 19 Hobert, O. Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* **27**, 681-696, doi:10.1146/annurev-cellbio-092910-154226 (2011).
- 20 Arendt, D. *et al.* The origin and evolution of cell types. *Nat Rev Genet* **17**, 744-757, doi:10.1038/nrg.2016.127 (2016).
- 21 Vergara, H. M. *et al.* A whole-organism cellular gene expression atlas reveals conserved cell type clusters in the ventral nerve cord of *Platynereis dumerilii*. *Proc Natl Acad Sci U S A* **114**, 5878-5885, doi:10.1073/pnas.1610602114 (2016).
- 22 Tessmar-Raible, K. *et al.* Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* **129**, 1389-1400, doi:10.1016/j.cell.2007.04.041 (2007).
- 23 Conzelmann, M. *et al.* The neuropeptide complement of the marine annelid *Platynereis dumerilii*. *BMC Genomics* **14**, 906, doi:10.1186/1471-2164-14-906 (2013).
- 24 Moghadam, P. K. & Jackson, M. B. The functional significance of synaptotagmin diversity in neuroendocrine secretion. *Front Endocrinol (Lausanne)* **4**, 124, doi:10.3389/fendo.2013.00124 (2013).
- 25 Gustavsson, N., Wu, B. & Han, W. Calcium sensing in exocytosis. *Adv Exp Med Biol* **740**, 731-757, doi:10.1007/978-94-007-2888-2_32 (2012).
- 26 Park, D., Li, P., Dani, A. & Taghert, P. H. Peptidergic cell-specific synaptotagmins in *Drosophila*: localization to dense-core granules and regulation by the bHLH protein DIMMED. *J Neurosci* **34**, 13195-13207, doi:10.1523/JNEUROSCI.2075-14.2014 (2014).
- 27 Williams, E. A. *et al.* Synaptic and peptidergic connectome of a neurosecretory centre in the annelid brain. *bioRxiv* (2017).
- 28 Nagata, T., Koyanagi, M., Tsukamoto, H. & Terakita, A. Identification and characterization of a protostome homologue of peropsin from a jumping spider. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **196**, 51-59, doi:10.1007/s00359-009-0493-9 (2010).
- 29 Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteyn, A. W. & Wittbrodt, J. Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain. *Science* **306**, 869-871 (2004).
- 30 Burmester, T., Weich, B., Reinhardt, S. & Hankeln, T. A vertebrate globin expressed in the brain. *Nature* **407**, 520-523, doi:10.1038/35035093 (2000).
- 31 Arendt, D., Tosches, M. A. & Marlow, H. From nerve net to nerve ring, nerve cord and brain--evolution of the nervous system. *Nat Rev Neurosci* **17**, 61-72, doi:10.1038/nrn.2015.15 (2016).
- 32 Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol Rev* **91**, 1447-1531, doi:10.1152/physrev.00031.2010 (2011).
- 33 Dobi, K. C., Schulman, V. K. & Baylies, M. K. Specification of the somatic musculature in *Drosophila*. *Wiley Interdiscip Rev Dev Biol* **4**, 357-375, doi:10.1002/wdev.182 (2015).

- 34 Kim, Y., McDole, K. & Zheng, Y. The function of lamins in the context of tissue building and maintenance. *Nucleus* **3**, 256-262, doi:10.4161/nucl.20392 (2012).
- 35 Peter, A. & Stick, R. Evolutionary aspects in intermediate filament proteins. *Curr Opin Cell Biol* **32**, 48-55, doi:10.1016/j.ceb.2014.12.009 (2015).
- 36 Herdman, C. & Moss, T. Extended-Synaptotagmins (E-Syts); the extended story. *Pharmacol Res* **107**, 48-56, doi:10.1016/j.phrs.2016.01.034 (2016).
- 37 Pedemonte, N. & Galiotta, L. J. Structure and function of TMEM16 proteins (anoctamins). *Physiol Rev* **94**, 419-459, doi:10.1152/physrev.00039.2011 (2014).
- 38 Seiler, C. *et al.* Duplicated genes with split functions: independent roles of protocadherin15 orthologues in zebrafish hearing and vision. *Development* **132**, 615-623, doi:10.1242/dev.01591 (2005).
- 39 Bates, K. E. & Whittington, P. M. Semaphorin 2a secreted by oenocytes signals through plexin B and plexin A to guide sensory axons in the *Drosophila* embryo. *Dev Biol* **302**, 522-535, doi:10.1016/j.ydbio.2006.10.015 (2007).
- 40 Sonderegger, P. & Matsumoto-Miyai, K. Activity-controlled proteolytic cleavage at the synapse. *Trends Neurosci* **37**, 413-423, doi:10.1016/j.tins.2014.05.007 (2014).
- 41 Tyndall, S. J. & Walikonis, R. S. The receptor tyrosine kinase Met and its ligand hepatocyte growth factor are clustered at excitatory synapses and can enhance clustering of synaptic proteins. *Cell Cycle* **5**, 1560-1568, doi:10.4161/cc.5.14.2918 (2006).
- 42 Michaluk, P. *et al.* Influence of matrix metalloproteinase MMP-9 on dendritic spine morphology. *J Cell Sci* **124**, 3369-3380, doi:10.1242/jcs.090852 (2011).
- 43 Vadon-Le Goff, S., Hulmes, D. J. & Moali, C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. *Matrix Biol* **44-46**, 14-23, doi:10.1016/j.matbio.2015.02.006 (2015).
- 44 Felbor, U. *et al.* Neuronal loss and brain atrophy in mice lacking cathepsins B and L. *Proc Natl Acad Sci U S A* **99**, 7883-7888, doi:10.1073/pnas.112632299 (2002).
- 45 Boglev, Y. *et al.* The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Dev Biol* **349**, 512-522, doi:10.1016/j.ydbio.2010.11.011 (2011).
- 46 Feldman, R. J., Sementchenko, V. I. & Watson, D. K. The epithelial-specific Ets factors occupy a unique position in defining epithelial proliferation, differentiation and carcinogenesis. *Anticancer Res* **23**, 2125-2131 (2003).
- 47 Speicher, S. *et al.* Neurotactin functions in concert with other identified CAMs in growth cone guidance in *Drosophila*. *Neuron* **20**, 221-233 (1998).
- 48 Zeev-Ben-Mordehai, T. *et al.* Amalgam, an axon guidance *Drosophila* adhesion protein belonging to the immunoglobulin superfamily: over-expression, purification and biophysical characterization. *Protein Expr Purif* **63**, 147-157, doi:10.1016/j.pep.2008.09.019 (2009).

- 49 Fremion, F. *et al.* Amalgam is a ligand for the transmembrane receptor neurotactin and is required for neurotactin-mediated cell adhesion and axon fasciculation in *Drosophila*. *EMBO J* **19**, 4463-4472, doi:10.1093/emboj/19.17.4463 (2000).
- 50 Stephan, A. *et al.* Neurotrypsin cleaves agrin locally at the synapse. *FASEB J* **22**, 1861-1873, doi:10.1096/fj.07-100008 (2008).
- 51 Matsumoto-Miyai, K. *et al.* Coincident pre- and postsynaptic activation induces dendritic filopodia via neurotrypsin-dependent agrin cleavage. *Cell* **136**, 1161-1171, doi:10.1016/j.cell.2009.02.034 (2009).
- 52 Pruitt, M. M., Letcher, E. J., Chou, H. C., Bastin, B. R. & Schneider, S. Q. Expression of the wnt gene complement in a spiral-cleaving embryo and trochophore larva. *Int J Dev Biol* **58**, 563-573, doi:10.1387/ijdb.140084ss (2014).
- 53 Nomaksteinsky, M. *et al.* Ancient origin of somatic and visceral neurons. *BMC biology* **11**, 53, doi:10.1186/1741-7007-11-53 (2013).
- 54 Rabe, T. I. *et al.* The transcription factor Uncx4.1 acts in a short window of midbrain dopaminergic neuron differentiation. *Neural Dev* **7**, 39, doi:10.1186/1749-8104-7-39 (2012).
- 55 Thomas, J. *et al.* Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. *Biol Cell* **102**, 499-513, doi:10.1042/BC20100035 (2010).
- 56 Brooks, E. R. & Wallingford, J. B. Multiciliated cells. *Curr Biol* **24**, R973-982, doi:10.1016/j.cub.2014.08.047 (2014).
- 57 Joly, J. S., Recher, G., Brombin, A., Ngo, K. & Hartenstein, V. A Conserved Developmental Mechanism Builds Complex Visual Systems in Insects and Vertebrates. *Curr Biol* **26**, R1001-R1009, doi:10.1016/j.cub.2016.08.017 (2016).
- 58 Stefanakis, N., Carrera, I. & Hobert, O. Regulatory Logic of Pan-Neuronal Gene Expression in *C. elegans*. *Neuron* **87**, 733-750, doi:10.1016/j.neuron.2015.07.031 (2015).
- 59 Jekely, G. *et al.* Mechanism of phototaxis in marine zooplankton. *Nature* **456**, 395-399, doi:10.1038/nature07590 (2008).
- 60 Brunet, T. *et al.* The evolutionary origin of bilaterian smooth and striated myocytes. *Elife* **5**, doi:10.7554/eLife.19607 (2016).
- 61 Murthy, V., Taranda, J., Elgoyhen, A. B. & Vetter, D. E. Activity of nAChRs containing alpha9 subunits modulates synapse stabilization via bidirectional signaling programs. *Dev Neurobiol* **69**, 931-949, doi:10.1002/dneu.20753 (2009).
- 62 Chernyavsky, A. I., Arredondo, J., Vetter, D. E. & Grando, S. A. Central role of alpha9 acetylcholine receptor in coordinating keratinocyte adhesion and motility at the initiation of epithelialization. *Exp Cell Res* **313**, 3542-3555, doi:10.1016/j.yexcr.2007.07.011 (2007).
- 63 Zhang, G. *et al.* Neuropeptide exocytosis involving synaptotagmin-4 and oxytocin in hypothalamic programming of body weight and energy balance. *Neuron* **69**, 523-535, doi:10.1016/j.neuron.2010.12.036 (2011).

- 64 Lee, M. Y., Choi, S. H., Shin, S. L., Chin, H. & Kwon, O. J. Distribution of B/K protein in rat brain. *Cell Tissue Res* **303**, 47-56 (2001).
- 65 Fabrizius, A. *et al.* Critical re-evaluation of neuroglobin expression reveals conserved patterns among mammals. *Neuroscience*, doi:10.1016/j.neuroscience.2016.07.042 (2016).
- 66 Hundahl, C. A. *et al.* Circadian behaviour in neuroglobin deficient mice. *PLoS One* **7**, e34462, doi:10.1371/journal.pone.0034462 (2012).
- 67 Hundahl, C. A. *et al.* Neuroglobin in the rat brain: localization. *Neuroendocrinology* **88**, 173-182, doi:10.1159/000129698 (2008).
- 68 Dominguez, L., Gonzalez, A. & Moreno, N. Patterns of hypothalamic regionalization in amphibians and reptiles: common traits revealed by a genoarchitectonic approach. *Front Neuroanat* **9**, 3, doi:10.3389/fnana.2015.00003 (2015).
- 69 Marlow, H. *et al.* Larval body patterning and apical organs are conserved in animal evolution. *BMC biology* **12**, 7, doi:10.1186/1741-7007-12-7 (2014).
- 70 Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
- 71 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).
- 72 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
- 73 Vallejos, C. A., Marioni, J. C. & Richardson, S. BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLoS Comput Biol* **11**, e1004333, doi:10.1371/journal.pcbi.1004333 (2015).
- 74 Witten, D. M. & Tibshirani, R. A framework for feature selection in clustering. *J Am Stat Assoc* **105**, 713-726, doi:10.1198/jasa.2010.tm09415 (2010).
- 75 R Core Team. *R: A language and environment for statistical computing.*, <<http://www.R-project.org/>> (2014).
- 76 Stijn van Dongen, A. J. E. Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv:1208.3145* (2012).
- 77 Ryota Suzuki, H. S. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542 (2006).
- 78 Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat Methods* **11**, 740-742, doi:10.1038/nmeth.2967 (2014).
- 79 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 80 Tessmar-Raible, K., Steinmetz, P. R., Snyman, H., Hassel, M. & Arendt, D. Fluorescent two-color whole mount in situ hybridization in *Platynereis dumerilii* (Polychaeta, Annelida), an emerging marine molecular model for evolution and development. *Biotechniques* **39**, 460, 462, 464 (2005).

- 81 Jekely, G. & Arendt, D. Cellular resolution expression profiling using confocal detection of NBT/BCIP precipitate by reflection microscopy. *Biotechniques* **42**, 751-755 (2007).
- 82 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
- 83 Fukuda, M. Molecular cloning and characterization of human, rat, and mouse synaptotagmin XV. *Biochem Biophys Res Commun* **306**, 64-71 (2003).
- 84 Adolfsen, B., Saraswati, S., Yoshihara, M. & Littleton, J. T. Synaptotagmins are trafficked to distinct subcellular domains including the postsynaptic compartment. *J Cell Biol* **166**, 249-260, doi:10.1083/jcb.200312054 (2004).

Acknowledgements

We thank Catalina Vallejos for comments on the implementation of BASiCS and all members of the Marioni and Arendt labs for feedback on the project and manuscript.

Author contributions

KA and NE planned and conducted experiments, interpreted the data, designed the study and wrote the manuscript. H MV, PB and PC conducted experiments. TB interpreted data. VB, JCM and DA funded the study. JCM and DA designed the study and wrote the manuscript.

Author information

The authors declare no competing interest.

Figures and Figure legends

Figure 1

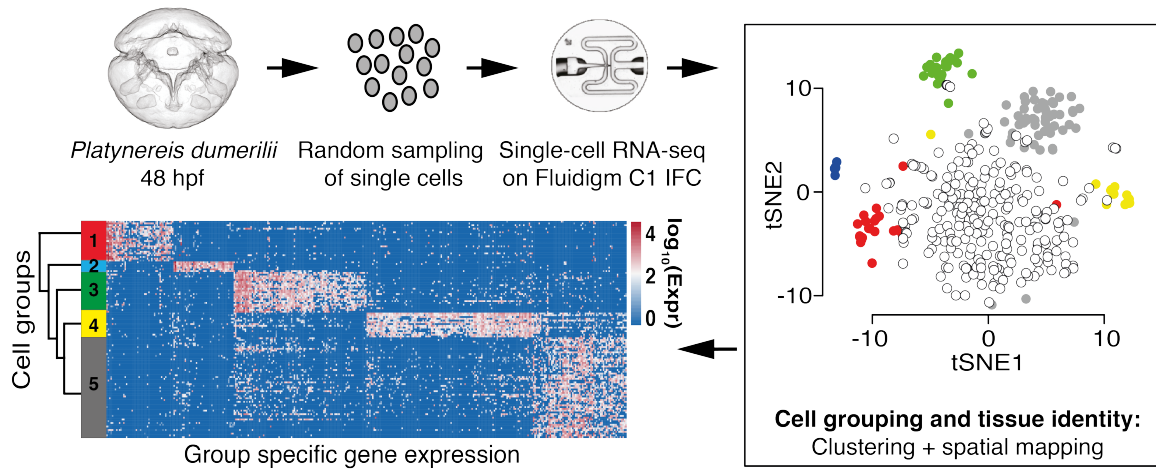


Figure 1. Single cell transcriptomics of *Platynereis* 48 hpf larvae.

Cells of the 48 hpf larvae were dissociated and randomly selected for single cell RNA-sequencing using the Fluidigm C1 Single-cell AutoPrep system. Combining sparse clustering with spatial positioning of single cells allows the identification of robust cell groups within the data. The clustering approach enables identification of genes that characterise each cell type. Finally, we used hierarchical clustering to investigate the similarity between the identified cell clusters.

Figure 2

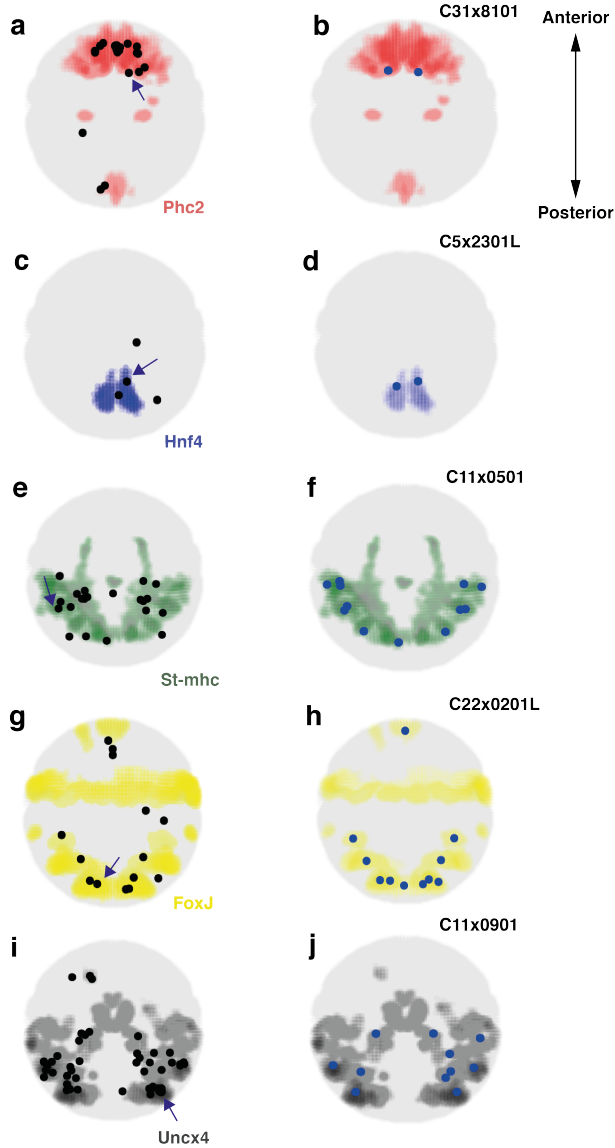


Figure 2. Spatial mapping of single cells to tissues characterized by specific marker gene expression.

a-j, Left panels: Centroids of the voxel clusters to which cells were mapped with highest confidence are shown as black dots. Tissue specific marker gene expression is visualized as colored regions within the embryo. Left panels: Mapping results for all cells belonging to the respective group. Right panels: For each cell group, we show an example of the

mapping result for one individual cell (indicated by an arrow on the left panel). **a**, Cells of the apical ectoderm group (n = 23 cells) map to the *Phc2* expressing region in the embryo. **b**, Mapping result for the cell C31x8101. **c**, Midgut cells (n = 4 cells) map to the *Hnf4* expressing region. **d**, Mapping result for the cell C5x2301L. **e**, Striated musculature cells (n = 23 cells) map to the *St-mhc* expressing region. **f**, Mapping result for the cell C11x0501. **g**, Cells of the ciliary bands (n = 14 cells) map to the *Foxj* expressing region. **h**, Mapping result for the cell C22x0201L. **i**, Cells of the non-apical ectodermal group (n = 55 cells) map to the *Uncx4* expression domain. **j**, Mapping result for the cell C11x0901.

Figure 3

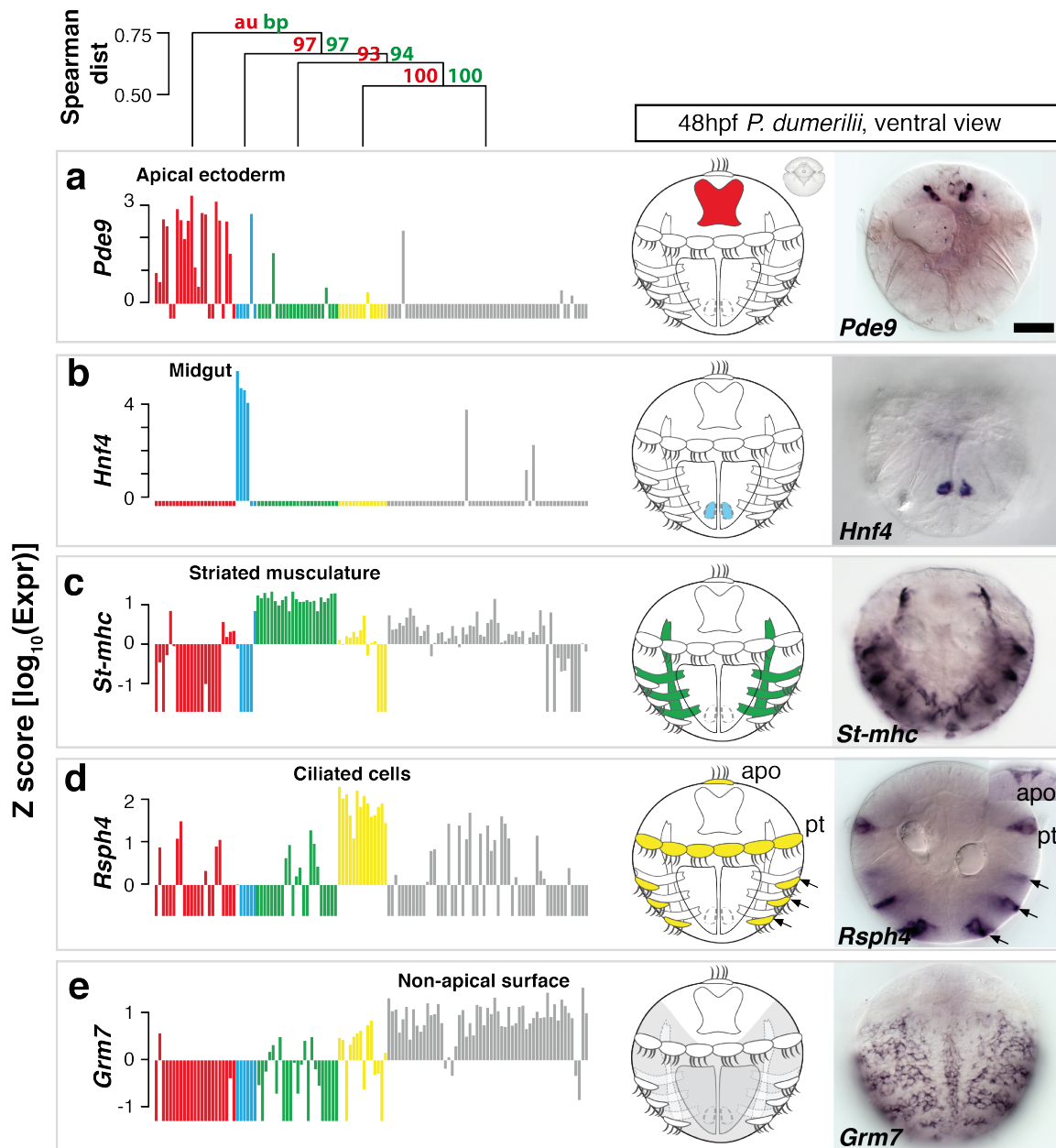


Figure 3. Identification and validation of tissue-specific genes.

On top, Distance tree showing the hierarchical relationships between the differentiated cell groups. **a-e**, For each identified cell group, the expression of a group-specific marker gene is shown in a bar plot, the respective tissue is shown schematically in the ventral view of *Platynereis* larva, and visualized by WMISH with respective probes: **a**, *Pde9*

expression in the apical ectoderm (red); **b**, *Hnf4* expression in the midgut (cyan); **c**, *Stmhc* expression in striated muscle (green); **d**, *Rsph4* expression in ciliated cells (yellow); **e**, *Grm7* expression characterizes the non-apical surface cells (gray). Note that *Pde9*, *Hnf4*, *Rsph4* and *Grm7* are novel markers for the respective cell groups. Each ISH pattern was replicated in at least 6 animals. Scale bar, 50 μm . Apo, apical organ; pt, prototroch, Z factor = $\frac{x_{ij} - \mu_i}{\sigma_i}$, x_{ij} : Expression of gene i in cell j, μ_i : Mean expression of gene i, σ_i :

Standard deviation of gene i.

Figure 4

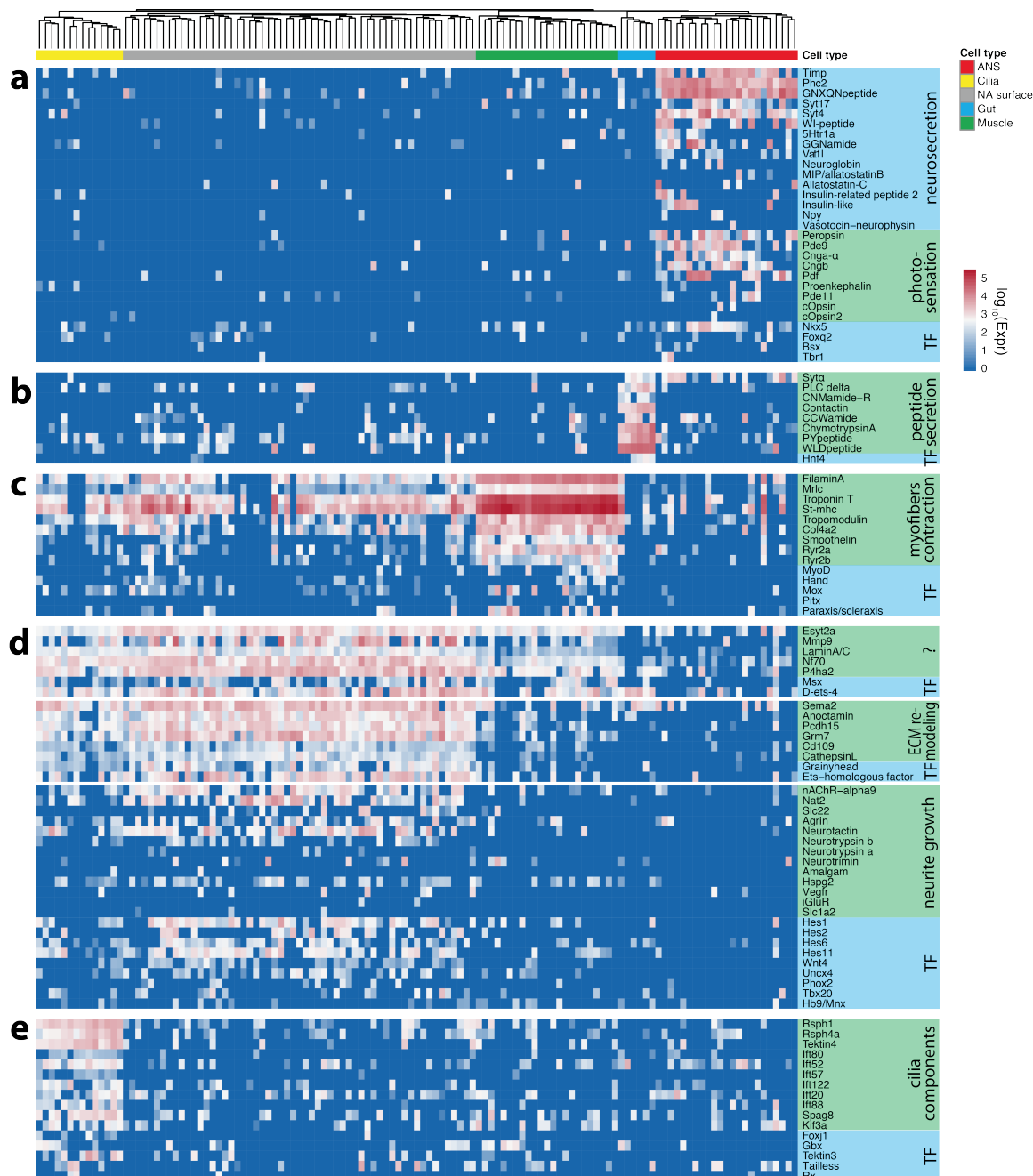


Figure 4. Tissue specific marker genes reflect cellular functions. For each of the differentiated tissues, we show a heatmap of tissue specific gene expression: **a**, expression profile of ANS-specific genes; **b**, expression profile of midgut genes; **c**, expression profile of ciliated cell-specific genes; **d**, expression profile of non-apical

surface-specific genes; and **e**, expression profile of myocyte-specific genes. Functionally related groups of genes are highlighted.

Figure 5

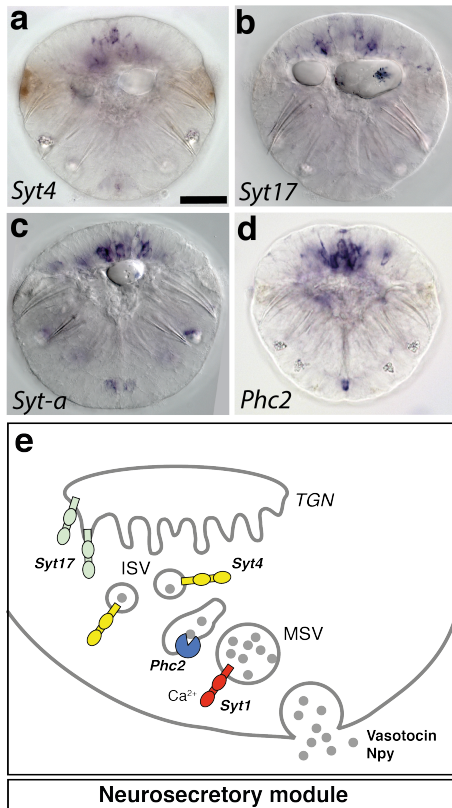


Figure 5. Neurosecretory functional module is specific to ANS cells. a-d, WISH analysis of *Syt4*, *Syt17*, *Syt-a* and *Phc2* expression in *Platynereis* larvae at 48hpf. Ventral view. Scale bar, 50 μ m e, Schematic of the neurosecretion cellular module. Syt17 contains an N-terminal cysteine cluster mediating membrane association with the trans-Golgi network (TGN)⁸³. Syt4 is involved in the maturation of secretory vesicles⁶³. Syt-a functions in vesicle trafficking of specific subclasses of neuropeptides and/or neuromodulators⁸⁴. TGN, trans-golgi network; ISV, immature secretory vesicle, MSV, mature secretory vesicle.

Figure 6

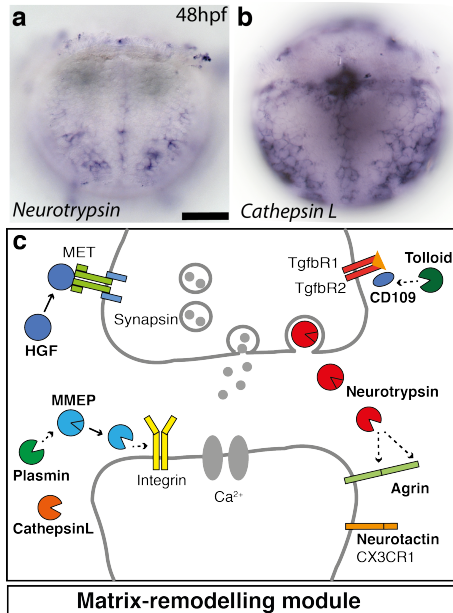


Figure 6. Extracellular matrix remodeling module is specific to non-apical surface cells. **a-b**, WMISH analysis of *Neurotrypsin* and *Cathepsin L* expression in *Platynereis* larvae at 48hpf. Ventral view focused at the surface. Scale bar, 50 μ m **c**, Schematic of the synapse formation cellular module. Neurotrypsin cleaves agrin locally at the synaptic cleft, triggering the formation of new synapses⁴⁰. HGF and its receptor MET enhance clustering of synaptic proteins at excitatory synapses⁴¹. Plasmin cleaves selected synaptic target proteins such as NMDA receptor and matrix metalloproteinases⁴⁰. Several members of the family of matrix-metalloproteinases such as MMP-9 have been in synapse formation and remodeling⁴². *Tolloid-like (Tll)* and its substrate *CD109* have been implicated in the control of TGFb signaling and extracellular matrix synthesis⁴³. *Cathepsin L* is likewise implicated in extracellular matrix remodeling and neuronal survival⁴⁴. The genes specifically expressed in non-apical ectodermal cells are marked in bold.

Figure 7

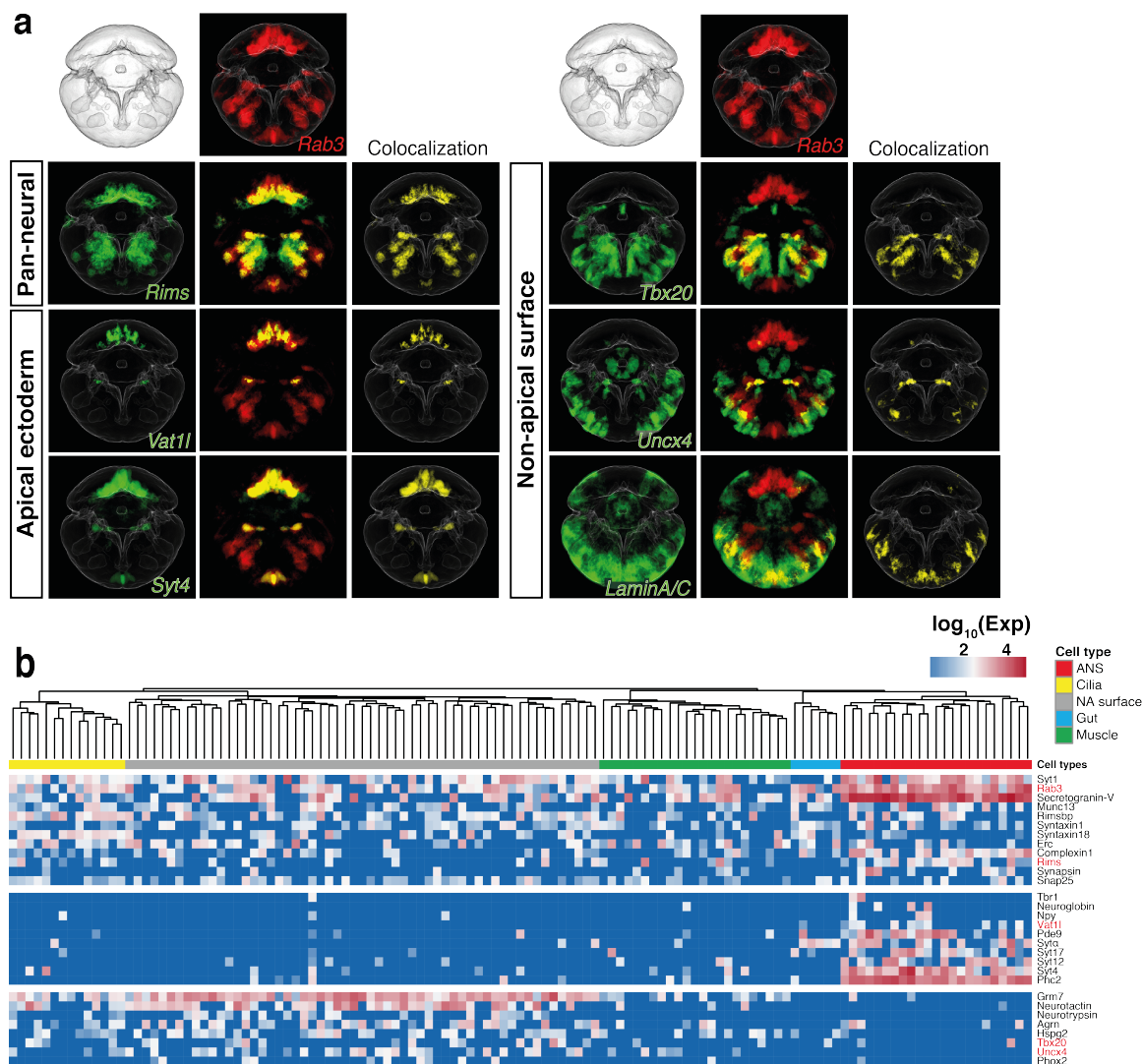


Figure 7. Shared and unique characteristics of the apical and non-apical nervous system. a, Averaged gene expression patterns and *in silico* analysis of co-expression with pan-neuronal marker gene *Rab3*, for selected marker genes. **b,** Heatmap showing the expression profile of pan-neuronal markers, as well as neuronal genes specific to apical ectoderm and non-apical surface cells. Red labels, the genes for which ISH is shown in (a).