Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of *Platynereis dumerilii*

Hernando Martínez Vergara[a], Paola Yanina Bertucci[a], Peter Hantz[a,b], Maria Antonietta Tosches[a,c], Kaia Achim[a], Pavel Vopalensky[a,d], and Detlev Arendt[a,1]

[a]Developmental Biology Unit, European Molecular Biology Laboratory, 69012 Heidelberg, Germany; [b]Laboratory of Physics of Complex Matter, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; [c]Max Planck Institute for Brain Research, 60438 Frankfurt am Main, Germany; and [d]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

The comparative study of cell types is a powerful approach toward deciphering animal evolution. To avoid selection biases, however, comparisons ideally involve all cell types present in a multicellular organism. Here, we use image registration and a newly developed "Profiling by Signal Probability Mapping" algorithm to generate a cellular resolution 3D expression atlas for an entire animal. We investigate three-segmented young worms of the marine annelid *Platynereis dumerilii*, with a rich diversity of differentiated cells present in relatively low number. Starting from whole-mount expression images for close to 100 neural specification and differentiation genes, our atlas identifies and molecularly characterizes 605 bilateral pairs of neurons at specific locations in the ventral nerve cord. Among these pairs, we identify sets of neurons expressing similar combinations of transcription factors, located at spatially coherent anterior-posterior, dorsal-ventral, and medial-lateral coordinates that we interpret as cell types. Comparison with motor and interneuron types in the vertebrate neural tube indicates conserved combinations, for example, of cell types cospecified by Gata1/2/3 and Tal transcription factors. These include V2b interneurons and the central spinal fluid-contacting Kolmer-Agduhr cells in the vertebrates, and several neuron types in the intermediate ventral ganglionic mass in the annelid. We propose that Kolmer-Agduhr cell-like mechanosensory neurons formed part of the mucociliary sole in protostome-deuterostome ancestors and diversified independently into several neuron types in annelid and vertebrate descendants.

gene-expression atlas | evo-devo | cell-type evolution | Kolmer-Agduhr cells | ProSPr

In recent years, the molecular comparison of cell types has grown into an exciting new discipline. One major aim is to understand the diversification of cell types in animal evolution and to reconstruct their ancient repertoire in remote ancestors (1, 2). Cell types represent evolutionary units, defined as "sets of cells that change in evolution together, partially independent of other cells, and evolutionarily more closely related to each other than to other cells" (2). One important focus of interest is the evolution of neurons and nervous systems, with several studies comparing sensory neurons (3–6), interneurons (7), or motor neurons (8, 9). The starting point for these comparisons has been the observation that molecular regions in the developing nervous system are conserved from medial to lateral, between vertebrate and fly (10), and vertebrate and annelids (9) (Fig. 1). If these mediolateral regions already existed in urbilaterian ancestors, it is likely that they gave rise to region-specific neural cell types in these ancestors; however, given that conserved developmental regions can result in entirely different morphologies (11, 12), it is clear that any further homology statement requires an exhaustive analysis of the cell types that differentiate from the compared regions.

In vertebrates, the alar plate of the neural tube contains several sets of dorsal interneurons, and the basal plate ventral interneurons and motor neurons that generate the locomotor

patterns controlling movement (Fig. 1*A*) (13). The distinct identities of these neuron types are established by combinations of transcription factors (14, 15), which form so-called "core regulatory complexes" (CoRCs) that activate the neural differentiation genes specific for a given cell type (2, 16). For example, a complex of the LIM homeodomain proteins Lhx3 and Islet1, the EGH homeodomain factor Hb9, and the cofactor NLI specifies somatic motor neurons (17); another complex of the Gata2 factor, together with the Tal1/Scl bHLH factor, the LIM-only protein Lmo4, and NLI specifies the inhibitory descending V2b interneurons (18) (Fig. 1*A*); and V2a interneurons are controlled by a CoRC comprising Lhx3 and NLI. Any large-scale comparison of cell types across phyla should thus focus on the neuron-type–specific, combinatorial expression of transcription factors. Indeed, combinations of Lhx3, Islet, and Hb9 have been found to specify somatic motor neurons also in fly and nematode (8). On these grounds, it was proposed that somatic motor neurons specified by Hb9, Islet, and Lhx3 existed in bilaterian ancestors (8, 9).

However, it was also noted that in all groups, additional motor neurons exist, specified by different, partially overlapping combinations of transcription factors (8, 14), blurring the picture of cell-type interrelationships between phyla. Obviously, clade-specific cell-type diversification events have taken place in the divergent bilaterian lineages, and cell types may have been lost in specific lineages (e.g., the "chordate bottleneck" hypothesis; see ref. 8). It follows that, to infer cell-type interrelationships between remote animal groups, it will be misleading to restrict comparisons to few candidate neurons: cell types that are even more closely related may exist elsewhere in the body, which—once included—might produce a different picture of cell-type interrelationships within or across species. Instead, cell-type comparisons should provide molecular information for as many cell types as possible in as many species as possible.

Here, we follow a systemic approach to molecularly characterize and compare all differentiated cell types in an entire animal. Our focus is on the marine annelid *Platynereis dumerilii*,
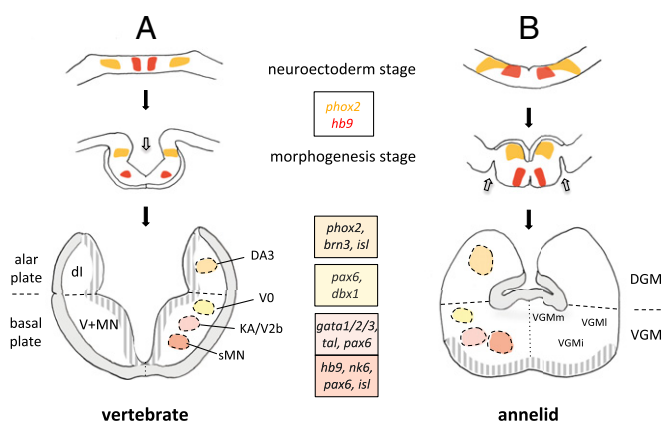
Fig. 1. Comparison of mediolateral patterning, morphogenesis, and molecular neuroanatomy in vertebrate and annelid. Schemes illustrate similarities and differences between vertebrate and annelid developing trunk nervous systems. Note that, because of different infolding strategies, the basal neuropil is outside of the tube in vertebrates, but inside the cord in annelids. Hatched gray, proliferative layer; Solid gray, neuropil. (*A*) Generalized development of the vertebrate neural tube. *hb9* expression after Dichmann and Harland (46) and *phox2* expression after Talikka et al. (47). White arrow in the *Middle* panel indicates invagination of the neural tube centered on the neural midline. Molecular anatomy of neural tube simplified after refs. 14 and 15. (*B*) Generalized development of the annelid nerve cord. *hb9* and *phox2* expression after Denes et al. (9) and this study. White arrows in the *Middle* panel indicate the bilateral, lateral infolding of the DGM. Molecular neuroanatomy according to this study. DGM, dorsal ganglionic mass; dI, dorsal interneurons; KA, Kolmer-Agduhr; MN, motorneurons; sMN, somatic motorneurons; V, ventral interneurons; VGM, ventral ganglionic mass.

which, in a number of recent studies (19–21) has proven to retain ancestral neural characters not found in fly or nematode, and is thus especially suited for long-range evolutionary comparisons. We find that, in *Platynereis*, the ventral nerve cord (VNC) is almost fully differentiated after 6 days post fertilization (dpf) in three-segmented young worms, which are composed of around 12,000 cells in total, making this stage an ideal system for in toto analysis (22). We present a technique, Profiling by Signal Probability mapping (ProSPr), which allows us to generate an atlas for around 100 genes and map gene expression onto a close-to-complete cellular model of the whole animal. We then focus on the VNC neurons to determine distinct molecular identities for 605 bilateral pairs of cells. Similarity-based clustering reveals neuron types expressing specific combinations of transcription factors. Besides several types of segmentally iterated, putative motor neuron types defined by coexpression of *nk6*, *hb9*, and *islet*, we also identify spatially coherent clusters of neurons that might be evolutionary related to different classes of vertebrate interneurons (Fig. 1). We propose that one of these groups, coexpressing *gata1/2/3* and *tal* transcription factors in a highly specific manner, may relate to the similarly defined sensory Kolmer-Agduhr neurons (23) and V2b ventral interneurons (14) in the vertebrates.

Results

In Toto Molecular Profiling of *Platynereis* Nectochaete Larvae. To select the most suitable life-cycle stage with a close-to-complete complement of cell types on the one hand, but as few cells as possible on the other, we determined the earliest stage when most neurons are differentiated in *Platynereis*. We treated developing larvae with 24-h pulses of EdU (that labels mitotic cells) from 2 to 6 dpf and found a significant decrease of cellular proliferation in the developing VNC over time, with just a few cells dividing during 5–6 dpf (Fig. S1*A*). In line with this finding, the vast majority of VNC cells appeared to express the neuronal markers *elaV* and *synaptotagmin* (Fig. S1*B*) at these later stages, indicative of

neuronal differentiation. Interestingly, this stage undergoes pelago-benthic transition (24), when animals start feeding and crawling on the substrate, performing periodic alternations of movements of their lateral body appendages, called parapodia. For this reason, the *Platynereis* VNC at 6 dpf represents a good system to study the molecular profile of neuronal cell types in a prototype protostome central nervous system. It should be stressed, however, that the 6-dpf VNC likely also contains quiescent progenitors and glial cells interspersed between neurons; in the absence of reliable markers, however, we are so far unable to identify these.

In a first attempt to produce an expression atlas at 6 dpf, we used Profiling by Image Registration (PrImR), a technique successfully used to generate cellular resolution atlases at earlier developmental stages (25, 26). This technique overlays and aligns different expression patterns onto a common, highly stereotypic spatial framework, such as the axonal scaffold, or DAPI-stained nuclei. However, we found that, at this more variable later stage with complex body features, the variability in position of a single, uniquely identifiable cell is larger than the average cell radius, in conflict with cellular resolution (Fig. S1 *B–D*).

To gain cellular resolution, we reasoned that sampling of an increased number of individuals should provide enough spatial information to produce a 3D map of probabilities for the location of the expression signal that could subsequently be used to calculate its average spatial location (Fig. 2*A*). We refer to such representation as a signal probability map (SPM). To test whether an SPM would recover the positional information of a given cell, we compared the region of highest expression probability with the arithmetic mean position (centroid) of this cell in different individuals. For this, we manually segmented a uniquely identifiable cell in several registered individuals, determined the centroid for all cells, and calculated the center of mass of the population (Fig. S1). We then compared the center of mass to the peak probability region within the SPM and determined that the distance between them remained below one average cell radius when using more than 12 samples for generating the SPM (Fig. 2*B*), and that increasing the imaging resolution over 660 nm per pixel did not result in a significantly better performance (Fig. 2*C*).

Taking these values into account, we generated a high-resolution reference (Fig. S2 and *Methods*) using DAPI scans of 153 individuals, which we used to register all individual larvae to the same 3D coordinate system. We then calculated SPMs for whole-body expression patterns of up to 100 different genes assessed by whole-mount in situ hybridization. As our primary interest was in neural cell types, this gene set comprises 59 transcription factors covering most homeodomain families, bHLH factors, Sox, Fox, and T-box families, 39 neural differentiation genes, as well as 3 proliferation markers. To infer an average expression for a given gene from the SPM, we set out to determine the specific threshold up to which the experimentally determined expression probability in a given SPM would be interpreted as real expression. We found that a general thresholding method was not adequate because of gene- and probe-dependent technical and biological variation (Fig. 2 *D* and *E*). Instead, we chose an individual threshold for each gene, so that the volume of the resulting average expression approximated the expected expression volume (that is the mean of expression volume in the registered individuals). For increased accuracy, we determined this threshold independently for each spatially separable volume of the overall expression pattern (Fig. 3). We refer to the resulting objects as minimal expression domains (MEDs). Next, we automated the preregistration (Fig. S3*A*), registration, and postregistration (Fig. S3*B*) steps, to generate, from raw microscopy files, MEDs that can be easily transformed into full gene-expression maps. We thus generated standardized whole-body gene-expression maps for all genes at 6 dpf (Fig. S4). We refer to our automated pipeline as ProSPr (Fig. S3*C*).

Generating Cellular Models from the Expression Atlas. We next developed a method to reconstruct virtual cells from the gene-expression atlas (Fig. 4*A*). We reasoned that, given the cellular resolution that our gene-expression atlas provides, it should
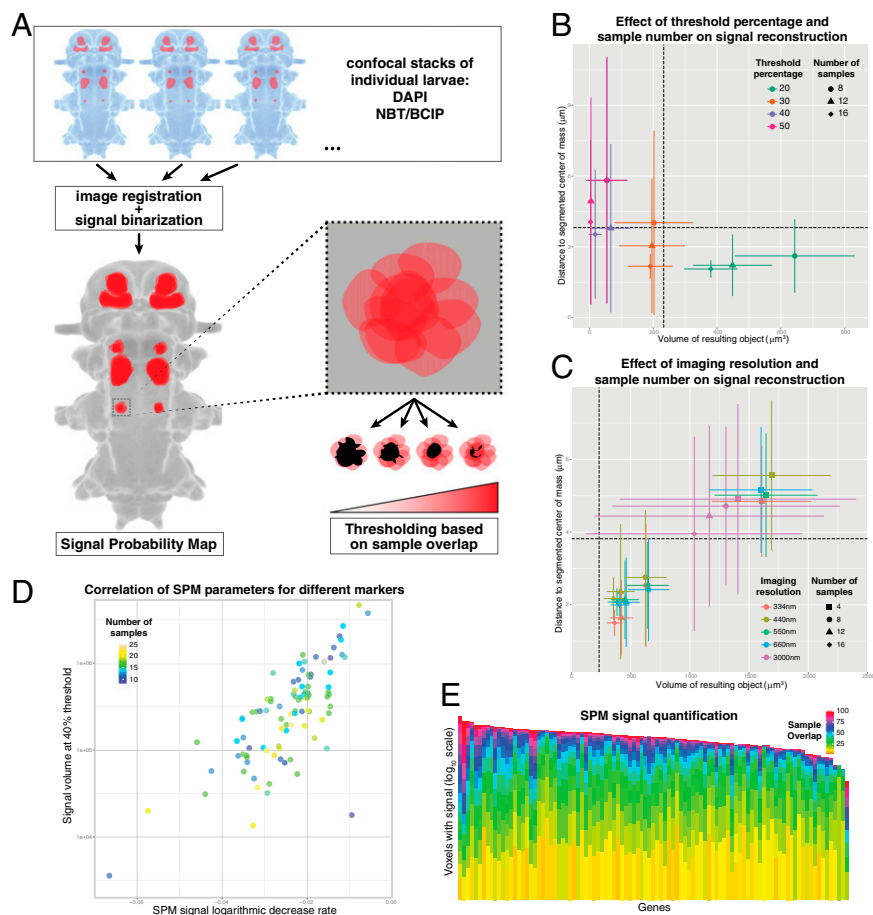
Vergara et al.

Fig. 2. SPMs can be used to achieve single-cell resolution after image registration at 6 dpf. (*A*) Scheme representing the signal reconstruction process, generating SPMs. Individual animals are registered to the reference, the signal is binarized, and overlapped to generate a probability map of the spatial distribution, which is later thresholded. (*B*) Effect of threshold percentage and sample size on signal reconstruction, scored as the distance of the resulting 3D object center of mass to the centroid of the cell distribution (calculated by manual segmentation), and the volume of the 3D object. From the initial dataset of 20 manually segmented cells (see text and Fig. S1), random picks were made for every imaging resolution. Each point represents the distribution of centroid-center distance and volume (mean and error bars) of 1,000 random iterations. Horizontal dotted line represents the average cell radius and vertical dotted line represents the average cell volume (calculated for the population of manually segmented cells). Results for the 550 nm per pixel imaging resolution dataset in *C*. (*C*) Effect of the imaging resolution and the number of samples on the performance of the signal reconstruction process, measured as in *B*. (*D*) Plot showing, for 99 different SPMs, the decrease rate of the signal volume in relation to the threshold, and its low correlation with the sample size and the coverage of the signal, making the use of a general thresholding method based on these parameters not adequate. (*E*) Quantification of SPMs based on the agreement between the samples for different molecular markers.

represent cells as units of homogenous gene expression. To reduce the complexity of the dataset of more than 15 million voxels, we grouped these into virtual cubes of 3^3 voxels that we refer to as supervoxels (SV), with each SV containing the average gene-expression information of the covered volume (Fig. 4*C*). Given an average cell diameter of 6 μm and the imaging resolution (isotropic 0.55 μm per pixel), each cell should thus be represented by 10–40 SV. To reduce the complexity of the dataset further, we subdivided the animal into distinct body parts enriched for neural tissue, hand-segmented from the reference. These included the VNC, brain, lateral ectoderm, cryptic segment (27), and the back end, the pygidium (Fig. 4*B*).

With the aim of generating a cellular model, we developed a strategy for the assembly of SV into groups representing virtual cells. First, we removed SV that did not show a high expression correlation with their neighbors, because these most likely represented cellular boundaries or noise (Fig. 4*D*). Second, we recursively used unsupervised hierarchical clustering to find groups of SV with similar expression profiles. Because cells belonging to the same cell type can be spatially separate (e.g., on the left and right body sides, or in consecutive segments), this similarity search was based on expression correlation only and did not take into account spatial location. As third step, however, we then assessed the resulting clusters of SV for spatial coherence. Clusters with largely scattered (spatially incoherent SV distributions) were eliminated, as they likely represented clustering artifacts (Fig. 4*A*). Clusters of SV concentrated around one or few locations were maintained, as they likely represented real cells. Note that in most cases such spatially coherent clusters represented bilateral pairs of cells. This recursive partition and clustering algorithm was run separately for each body part and recovered a total number of 4,315 spatially coherent,

minimal units of homogenous gene expression. These virtual units, mostly representing bilateral cell pairs, covered more than 80% of the total processed volume. Given that cell boundary volumes had been discarded, we are confident that our cellular atlas represents a molecular description of almost all cells of the *Platynereis* nervous system at 6 dpf.

Approximating Cell Types: Spatially Coherent Clusters of Cells with Similar Regulatory Signature. We next investigated how the 4,315 virtual cells would relate to each other. For a first coarse grouping, based on expression similarity, we used t-distributed stochastic neighbor embedding (t-SNE), which transfers high-dimensional data into a space of low dimensions (28). This revealed groups of VNC and head cells as coherent entities in the 6-dpf body (Fig. 4*E*). In subsequent analyses, we focused on the VNC cells and among these we selected, by visual inspection, a high-confidence set of cells based on their overall morphology, expression correlation, bilateral symmetry, and size. Fig. S5 shows these 605 reconstructed, bilateral pairs of cells in the *Platynereis* VNC. With the aim of identifying cell types within this dataset, we used t-SNE to group cells based on their combination of transcription factors. We focused our analysis on those cell groups with most expression information (groups I to XIV in Fig. 5*A*, composed of cells expressing four or more transcription factors) and discarded those for which little expression information was available. Each of these gene-rich groups homogeneously expressed a small set of transcription factors—such as *nk6*, *hb9*, *pax6*, and *islet* for group I (Fig. 5*B*)—many of which are known to play conserved roles in regional patterning (Fig. S6).

To further relate the molecular signatures of the cells composing each group, we performed unsupervised hierarchical clustering with all markers included in the atlas. This process yielded
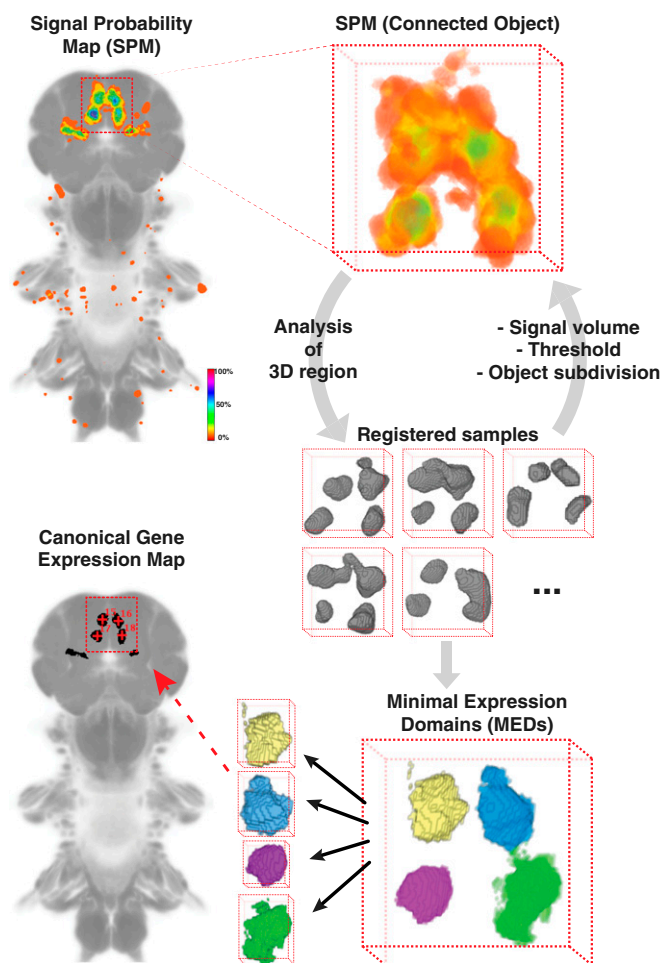
Fig. 3. Dissection of gene expression maps into MEDs. Schematic representation of the principle to threshold the SPMs (color-code indicates signal probability). Three-dimensional objects are automatically isolated and thresholded to approximate the amount of signal measured in the individual animals in that 3D space. As a consequence of the thresholding, new isolated 3D volumes can appear. This process is repeated for each isolated object until no further subdivision of the objects can be achieved, giving rise to MEDs.

clusters of virtual cells that expressed identical combinations of transcription factors and largely similar sets of neural effector genes (see, for example, type 1 to type 5 in Fig. 5*B*, and the sets of cells of similar color in Fig. 5 *C* and *D* and Fig. S7). Given the importance of the transcription factor regulatory signature in defining cell-type identity, we propose these core clusters may approximate cell types of the VNC (*Discussion*). We found all of these putative cell types to be spatially coherent, with specific ventral-dorsal, medial-lateral, and anterior-posterior coordinates (Fig. 5 *C* and *D*). In line with the segmented nature of the annelid nerve cord, we noted that many of them spread out over several segments in serially repeated arrangement (e.g., Fig. 5*C*). Interestingly, the first segment showed many unique types not iterated in the second and third segment (e.g., in groups III, IV, V).

Molecular Anatomy of the Ventral Nerve Cord. We noted that the two axes of the t-SNE reflect the major anatomical axes of the VNC (ventral-dorsal and medial-lateral), illustrating its molecular anatomy. To explore this further, we subdivided the t-SNE plot into an arbitrary number of seven superclusters (Fig. 6*A*) using *k*-means classification (seven was found to produce consistent results). The resulting superclusters indeed demarcated medial, intermediate, and lateral parts in the nerve cord, as well as prominent dorsal

"extensions" of the nerve cord (Fig. 6*B*). We refer to these as the medial, intermediate, and lateral ventral ganglionic mass (VGMm, VGMi, VGMl) and dorsal ganglionic mass (DGM), respectively (Fig. 6*C*). In molecular terms, the VGMm-i boundary coincides with the medial expression boundary of *pax6*, and the VGMi-l boundary with the lateral extent of *hb9* and *nk6* expression. The DGM prominently and specifically expresses the *phox2* gene (Fig. 6*D*).

The partition of the 6-dpf VNC ganglia into *phox2*⁺ DGM and *pax6*⁺, *nk6*⁺, and *hb9*⁺ VGM relates back to mediolateral regions of the 48 hours post fertilization (hpf) ectoderm (Fig. S8*A*). At 48 hpf, *phox2* is expressed more laterally than *hb9* in the ventral neuroectoderm. During subsequent morphogenesis, these *phox2*⁺ lateral cells are folded inwards on each side of the neuroectoderm (Fig. S8 *B* and *C*). In consequence, the former lateral cells are now located inwards, in a more dorsal position, as part of the DGM. Fig. 1 explains and summarizes these morphogenetic movements, leading to the specific neuroarchitecture at 6 dpf (Fig. 6*C*).

Analyzing the molecular anatomy of the VNC further, we found two transcription factors, *tal* and *gata1/2/3*, specifically expressed in the VGMi (Fig. 6*E*), and determined their coexpression (Fig. 6*F*). This led us to identify nine *tal*⁺, *gata1/2/3*⁺ neuron types with distinct combinations of transcription factors distributed over several t-SNE groups (III, IV, V, VI, VII, XII, and XIV). Two of these cell types express serotonergic (coexpressing *sert* and *trpH*) and cholinergic markers, another three cholinergic only, and one GABA- and glutamatergic markers. These neuron types may relate to the *gata2/3*⁺, *tal1*⁺ V2b interneurons and the Kolmer-Agduhr cells of the ventral part of the spinal cord of the vertebrates (Fig. 1) (*Discussion*).

Discussion

Cellular-Resolution Expression Atlases for Complex Body Plans. We have demonstrated that whole-body gene-expression atlases with cellular resolution can be achieved for an advanced developmental stage of an entire animal, the three-segmented young worm of the annelid *P. dumerilii*. At this stage, the nervous system and locomotor apparatus is largely differentiated and fully functional. So far, in other animal models similar cellular resolution expression atlases exist for selected tissues only (29, 30). Building on previous image registration approaches (25, 26), ProSPr is a highly automated and easy-to-customize technique, running in free license software (Scripts available in prospr.embl.de). In an unbiased manner, it allows subdividing the animal into genetically defined regions for which quantitative information can be extracted. Our technique is transferable to other species that are transparent or amenable to whole-body imaging via tissue-clearing methods (31).

The high quality of the atlas is reflected by the successful reconstruction of 4,315 minimal spatial units: pairs (mostly) of virtual cells with expression information for up to 100 genes. This reconstruction is achieved by a recursive partition clustering algorithm that groups SV solely based on similar expression and then validates the resulting cells by spatial coherence. This way, a high percentage of SV corresponding to a large fraction of the total animal volume is successfully grouped as part of computer-defined cells. The bilateral symmetrical arrangement of the cells, reflecting the bilateral symmetry of the animal, makes it likely that such reconstructed cellular units correspond to real biological cells and thus validates our approach.

The Nature of *Platynereis* Ventral Nerve Cord Neural Cell Types. Having detected thousands of bilateral pairs of virtual cells, we ask how these assemble into cell types. In this report, we have focused our analyses on a subset of VNC cells, representing an important part of the trunk central nervous system. To identify possible cell types in the VNC, we have used hierarchical clustering of cells grouped via t-SNE, which yielded sets of cells expressing similar combinations of transcription factors. According to a recent evolutionary definition of cell types (2), this is good
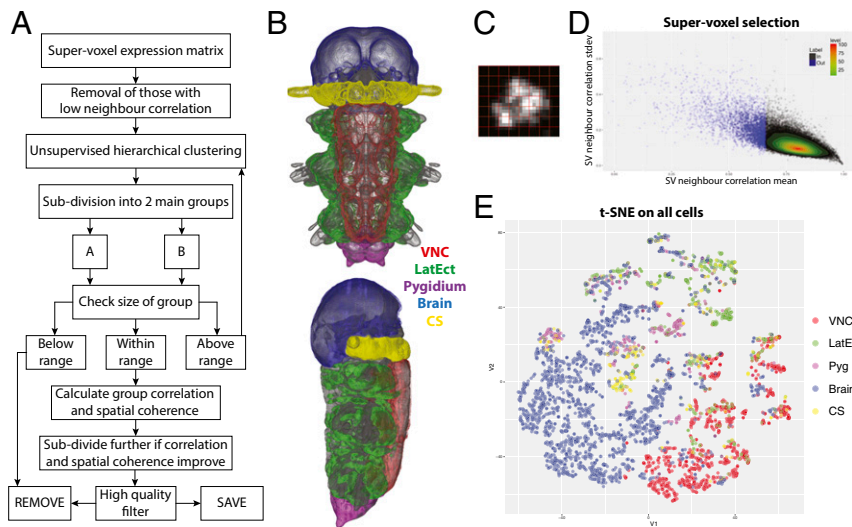
Vergara et al.

Fig. 4. Reconstruction of a cellular model from the 6 dpf *Platynereis* atlas. (*A*) Workflow of the recursive partition of hierarchical clustering to group SV into cells. (*B*) Manual segmentation of the reference to profile different tissues. (*C*) Illustration of the cellular grid to build SV (maximum projection of NBT/BCIP reflection signal in a single neuron). (*D*) Removal of SV with low correlation with their neighbors. (*E*) t-SNE of the 4,315 cells recovered by the algorithm in the segmented tissues, using all genes in the atlas. CS, cryptic segment; latEct, lateral ectoderm; Pyg, pygidium; VNC, ventral nerve cord.

indication that these cells might belong to the same type. We reasoned that cells within the same cell type might still vary in gene expression, because of distinct states of differentiation, or local somatic environment. (Note that variation because of technical or biological noise, such as bursts in transcription, is less likely because expression has been averaged over many individuals.) Using these criteria, we ended up defining 41 tentative cell types for those parts of the VNC with good expression coverage (Fig. 5 *B–D* and Fig. S7). Taking into account that some of these might represent quiescent progenitors or unidentified glial or epidermal cells, a careful estimate would extrapolate around 200 distinct types of neurons in the *Platynereis* nerve cord at this stage (relating the volume contained in the 41 cell types to the total volume of the VNC). However, given the restrictions of our dataset with a limited number of genes per cell assayed for only one stage, it is still difficult to draw a line between cells belonging to the same cell type and cells belonging to related, so-called sister-cell types (1).

We noted that a considerable number of our tentative cell types are unique to the first segment. This is in line with the observation that, while the second and third segments take part in crawling locomotion, the parapodia of the first segment do not play a major role—which would imply divergent sets of neuron types. Later, in the course of the second metamorphosis (24), segment one transforms into a head segment bearing long sensory cirri. Our results indicate that this specific fate is already foreshadowed in the rather distinct cell-type complement of segment one at 6 dpf.

Molecular Anatomy of the Annelid Ventral Nerve Cord. The subdivision of the nervous system into genetically defined superclusters reveals its molecular neuroanatomy in an unbiased manner. Of 59 transcription factors, we identify only a handful that define coherent anatomical regions. Among them, *pax6* plays the most prominent role, as it is absent from the VGMm, coexpressed with *nk6* and *hb9* in the VGMi, and coherently expressed in the VGMl. In addition, *phox2* expression defines the DGM (Fig. 6D). Overall, this resembles the vertebrate situation, where Nk6 together with Hb9 define ventral interneuron populations in the basal plate, whereas Phox2 is found in several types of dorsal interneurons (Fig. 1) (15). Notably, however, vertebrates differ from the annelids in that Phox2 is also specifying visceral motor neurons more ventrally (15).

Our unbiased approach reveals another important feature of the annelid VNC. As apparent in Fig. 6 *B* and *C*, the intermediate and lateral parts of the ganglionic mass (VGMi and VGMl) take the rough shape of concentric rings that cross the midline along the intersegmental boundaries. This finding reflects the segmental nature of the VNC and its anatomical subdivision into

spherical ganglia. It appears that the evolution of ganglia from a homogenous ventral neural mass involved some kind of "constriction" of the mediolateral architecture, moving lateral parts closer to the midline. This idea finds support for example in the medially constricting *pax6* expression in *Drosophila* (32), which likewise possesses a ganglionic VNC.

Evolutionary Conservation of Motor and Interneurons. Many of the transcription factors defining the presumed cell types and groups of cell types in the *Platynereis* VNC have previously been shown to control specification of neural cell types in the vertebrates [Tal, Gata1/2/3 (18); Beta3 (33); Brn3, Hb9, HNF6, Lhx1/5, Nk6, Pax6, Phox2 (14, 15); Dach (34)]. This information indicates that at least some of the *Platynereis* cell types may be specified by similar CoRCs and thus be evolutionarily related (1, 2). Four possible cases of conservation are apparent from our study. First, the nine neuron types in the VGMi that coexpress *gata1/2/3* and *tal (scl)* (Fig. 6 *E* and *F* and Fig. S7) may relate to V2b interneurons and Kolmer-Agduhr sensory neurons in the vertebrate basal plate, which are specified by the same transcription factors. However, although the vertebrate V2b CoRC also contains Lmo4, *Pdu-lmo4* is not prominently expressed in the 6-dpf VNC (Fig. S4). In addition, vertebrate V2b cells are GABAergic, whereas the *Platynereis gata1/2/3*⁺, *tal*⁺ cells show multiple transmitters including GABA. Second, VGMi contains segmentally reiterated cells coexpressing *hb9*, *nk6*, and *islet* (Figs. 5 *B* and *C* and 6 *E* and *F*) that may represent conserved motor neurons, given that the same three transcription factors specify motor neurons in the vertebrate neural tube (14). Third, VGMl contains *pax6*⁺, *dbx1*⁺ neurons (Fig. 6 *E* and *F*) that, by molecular identity, may relate to GABAergic commissural interneurons in the vertebrate neural tube (14). Fourth, DGM contains *phox2*⁺, *brn3*⁺, *isl*⁺ neurons (Fig. 6 *E* and *F*) that may correspond to sensory interneurons with similar identity in the vertebrate hindbrain (15). Notably, in none of these cases other cells were detected outside of the central nervous system showing similar combinations of transcription factors. Because at this state of analysis not much is known about the axonal projections of the *Platynereis* neuron types, further morphological and physiological characterization and comparison with vertebrate motor and interneuron types would be desirable. In addition, because phenotypic similarities—and thus, similarities in gene expression—can also arise by convergent or concerted evolution (2), validation of the proposed homologies will require more extensive molecular characterization, including the functional validation of transcription factor involvement in cell-type specification and, ideally, single-cell transcriptomics for an unbiased account of all cellular modules used by each type (35).
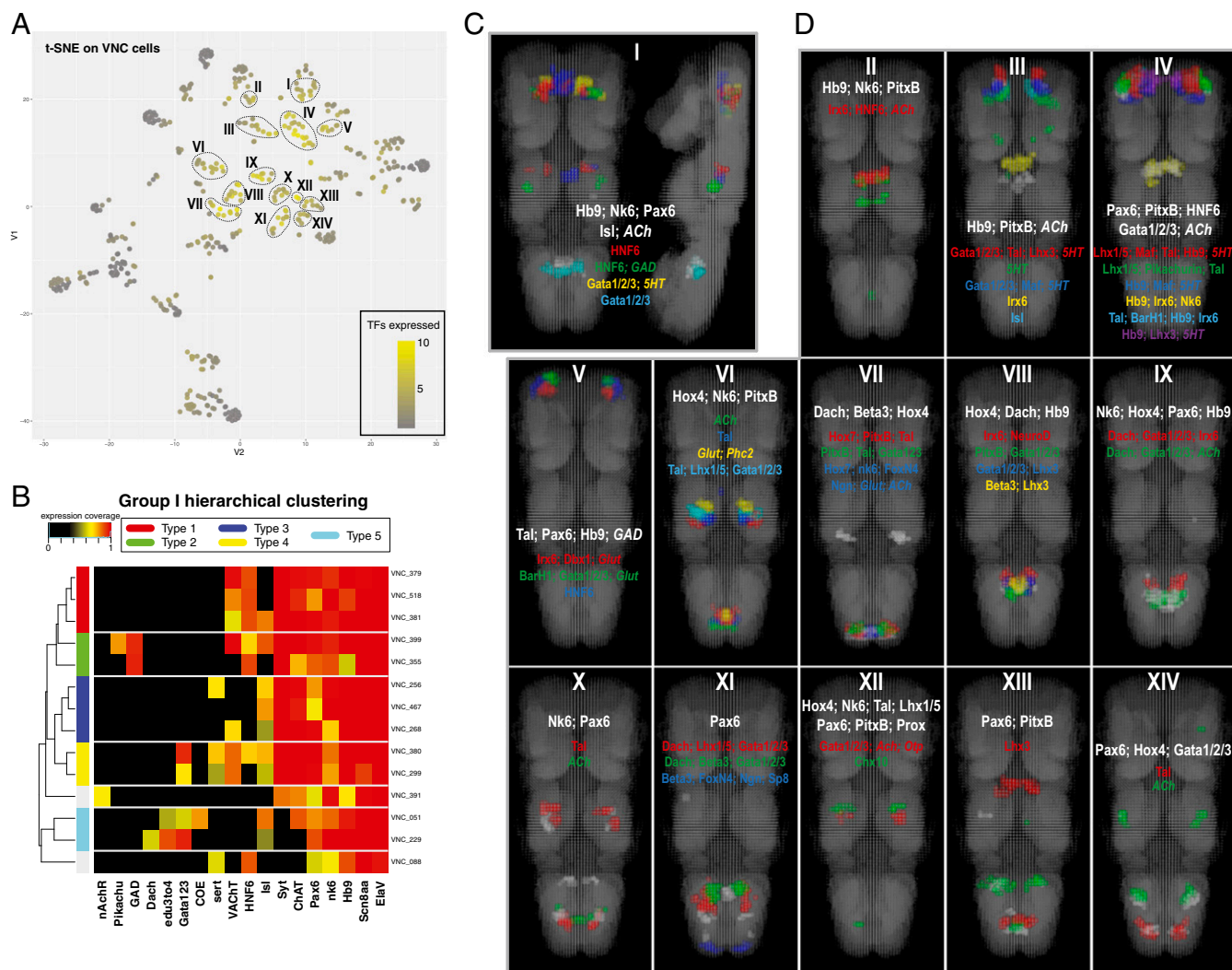
Fig. 5. Cell-type analysis on the 6 dpf *Platynereis* ventral nerve cord. (*A*) t-SNE of the VNC cells using transcription factor expression, showing cells color-coded according to how many transcription factors they show expression for. Groups of cells with high expression information have been manually selected for further analysis. (*B*) Hierarchical clustering and heatmap showing the genes expressed in cells in group I. Cell types have been manually defined based on the clustering and are highlighted in different colors. (*C*) Spatial location (frontal and lateral views of the VNC) of the cell types in group I. Transcription factors and neurotransmitters (italics) expressed in these cells are indicated in the figure. In white, genes common to the entire group; in color, those specific for the different cell types. (*D*) Frontal views of the VNC showing the location of the different cell types identified from the groups in *A*. Genes expressed in these cell types are indicated as in *D*.

Kolmer Agduhr Cells: An Ancient Cell Type of the Mucociliary Sole?

The proposed ancient nature of a neuron-type clade specified by a Gata1/2/3-Tal CoRC finds support by recent studies showing that the same transcription factors are also expressed in Kolmer-Agduhr neurons in fish (23) and in the corresponding central spinal fluid-expressing neurons in mouse (36). Kolmer-Agduhr neurons are hair cell-like sensory neurons that contact the central canal and send out ipsilateral ascending axons (37). Interestingly, the sensory apparatus of Kolmer-Agduhr cells directly contacts Reissner's fiber, a mucous strand in the neural canal in all vertebrates and in the chordate amphioxus (38), indicative of an ancient origin of Kolmer-Agduhr–like neurons. We have recently proposed that Reissner's fiber is a remnant of an ancient mucociliary sole (39). Finally, both Kolmer-Agduhr cells and some *Platynereis* g*ata1/2/3+*, *tal+* neuron types (Fig. 6) share the synthesis and release of serotonin, at least transitorily (ref. 40 and this study). Taken together, these observations indicate that today's *gata1/2/3-tal*–expressing neuron types in annelids and vertebrates may stem from Kolmer-Agduhr–like sensory serotoninergic precursors involved in the control of mucociliary locomotion.

Outlook. ProSPr is a highly valuable method to study the cell-type complement across different developmental stages and across different species that has the potential to significantly change the direction and throughput of research in the field of evolution and development. Once established, cellular expression atlases can be constantly expanded with additional gene sets fed into the resource. Although we exemplify and illustrate the value of our resource with the inferred molecular neuroanatomy of the annelid VNC and the identification of a subset of putative cell types, a similar analysis and comparison is currently carried out for other parts of the nervous system and the entire body across different developmental stages. In addition, spatial mapping of single-cell transcriptomes (41) onto the cellular atlas has huge potential in complementing this resource with full expression profiles for all cell types, and guide the inclusion of new genes to the atlas for refined mapping. Because ProSPr can easily be applied to other stages, this will allow tracking cell types through development. Furthermore, the small size of *Platynereis* larvae and the relatively small overall cell number enable whole-body connectomics (42), and thus the integration of ProSPr and connectomics data
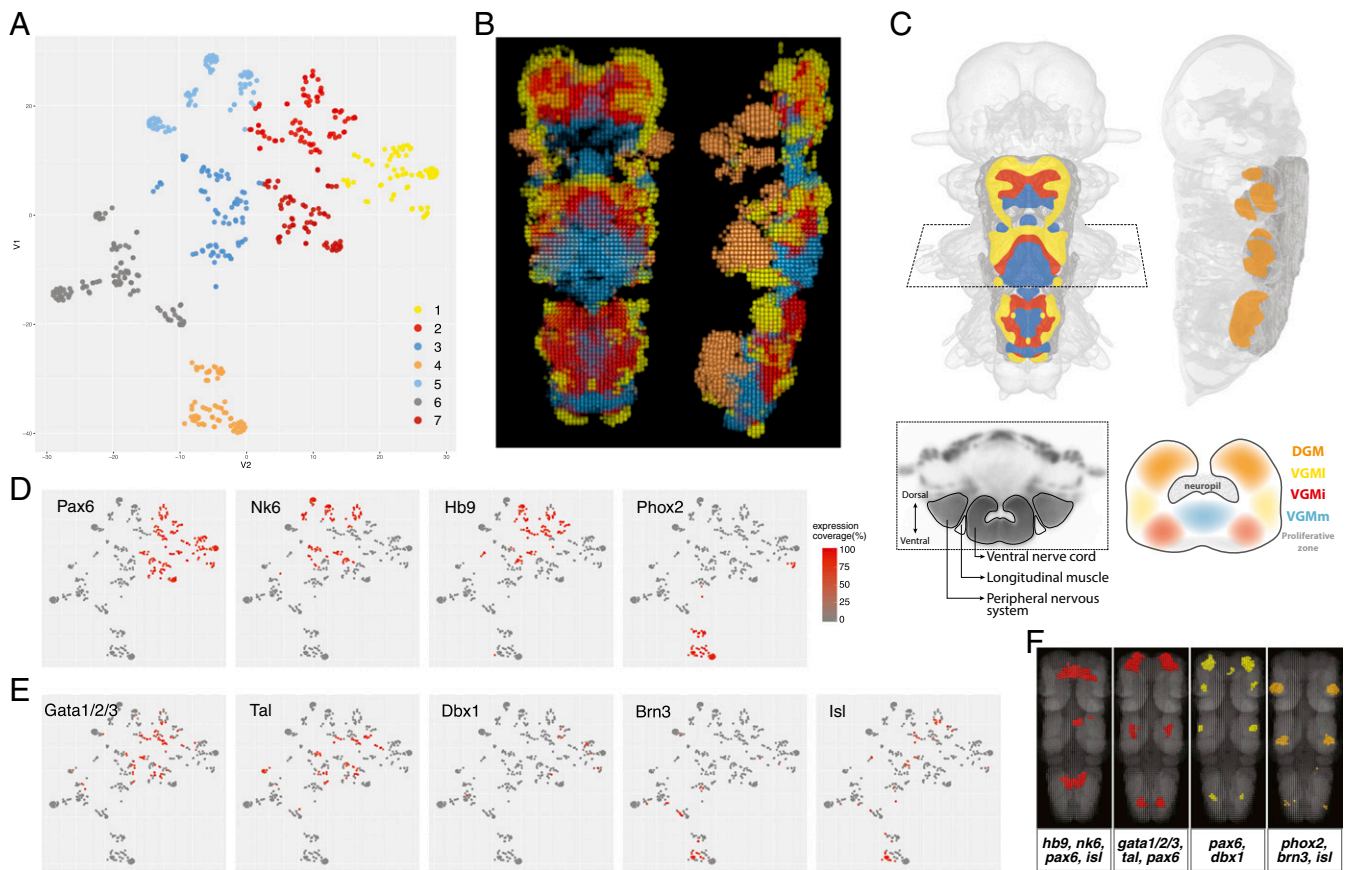
Vergara et al.

Fig. 6. Medio-lateral arrangement of VNC superclusters. (*A*) *k*-means clustering (seven clusters) on the VNC cells t-SNE. (*B*) Spatial location of the *k*-means superclusters found in *A* (ventral and lateral views of the VNC). (*C*, *Upper*) Illustration of the molecular structure of *Platynereis* VNC at 6 dpf mapped on the entire body plan (ventral and lateral views). (*Lower*, *Left*) Transversal section (see *Top*) of the reference at the level of the second VNC ganglia, indicating the ventral structures. (*Lower*, *Right*) Depiction of the major genetically defined territories in the VNC. (*D*) Cell expression of transcription factors Pax6, Nk6, Hb9, and Phox2 overlaid on the t-SNE graph. (*E*) Cell expression of transcription factors Gata1/2/3, Tal, Dbx1, Brn3, and Isl overlaid on the t-SNE graph. (*F*) Spatial location of four major groups of cell types compared with those in vertebrates in this study. Frontal views of the VNC showing the cells that coexpress the transcription factors indicated at the bottom.

into multidimensional, cellular whole-body atlases, which can finally be complemented by functional characterization of key genes and cell types using Crispr-Cas9. The possible transfer of ProSPr to other model species has the potential to boost comparisons between species and phyla, and to thus contribute to unraveling the evolutionary history of cell types.

Materials and Methods

***Platynereis* Breeding.** *P. dumerilii* larvae were obtained from an established culture at the European Molecular Biology Laboratory Heidelberg. Larvae were raised in natural sea water, at 18 °C under a 16-h light/8-h dark cycle.

Whole-Mount in Situ Hybridization. Available transcriptome and EST data were used to design PCR primers. HotStart Taq Polymerase (Qiagen) or the Phusion polymerase (New England BioLabs) was used for amplification. PCR fragments were cloned into the pCRII-Topo vector (Invitrogen) and validated by sequencing. Animal fixations and whole-mount in situ hybridizations were performed as described previously (43).

Imaging. Six days post fertilization, *Platynereis* larvae were mounted in 2,2′-thiodiethanol (TDE) according to Asadulina et al. (25), and imaged in a Leica SP8 confocal microscope using the following parameters: 40× objective, 0.75 zoom, 696 × 696 pixels, and *z*-step size of 0.55 μm, achieving an isotropic resolution of 550 nm per voxel. DAPI channel was imaged using the 405-nm laser and the NBT/BCIP signal using the 633 laser in reflection mode. Bright field and autofluorescence channels were acquired exciting with the 488 or the 633 laser.

Image Registration. Image registration was performed using the DAPI stain as the reference signal, following the protocol described in Asadulina et al. (25).

Reference. The reference was built using the DAPI channels of 153 individual larvae in an iterative registration procedure. First, a rigid transformation was applied to the individuals to position them in the same orientation, and a "rigid average" was calculated. Rigidly oriented animals were registered to this average using affine transformation, and an "affine average" was calculated. Rigidly oriented animals were registered to the affine average using an affine, followed by a deformable registration procedure, and a "deformable average" was calculated. This last step was repeated one more time to generate the final reference as the average result. The reference file can be downloaded from prospr.embl.de. Different tissues have been manually segmented from the reference using TrackEM (44).

ProSPr Pipeline. The scripts composing the full preregistration and postregistration steps of ProSPr pipeline, coded in Fiji (45), are explained in Fig. S3 and distributed in prospr.embl.de. All gene phylogenetic analyses and raw microscopy files are available on demand. Gene expression maps and the complete atlas can be downloaded from prospr.embl.de.

Bioinformatics Analysis. The full analysis of the gene-expression atlas has been performed using R Bioconductor. The SV generation script and the recursive partition pipeline for cellular reconstruction, are distributed in prospr.embl.de. For hierarchical clustering and t-SNE, "jaccard" distance from the "vegan" R-package was used. R-package "Rtsne" was used to run the t-SNE algorithm on the data using the following parameters: initial dimensions = 10; perplexity = 31. The method "complete" was used in the hierarchical clustering function.

Vergara et al.

www.manaraa.com

1. Arendt D (2008) The evolution of cell types in animals: Emerging principles from molecular studies. *Nat Rev Genet* 9:868–882.
2. Arendt D, et al. (2016) The origin and evolution of cell types. *Nat Rev Genet* 17: 744–757.
3. Tessmar-Raible K, et al. (2007) Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* 129:1389–1400.
4. Fritzsch B, Beisel KW, Pauley S, Soukup G (2007) Molecular evolution of the vertebrate mechanosensory cell and ear. *Int J Dev Biol* 51:663–678.
5. Jacobs DK, et al. (2007) Evolution of sensory structures in basal metazoa. *Integr Comp Biol* 47:712–723.
6. Baker CV, Modrell MS, Gillis JA (2013) The evolution and development of vertebrate lateral line electroreceptors. *J Exp Biol* 216:2515–2522.
7. Flames N, Hobert O (2009) Gene regulatory logic of dopamine neuron differentiation. *Nature* 458:885–889.
8. Thor S, Thomas JB (2002) Motor neuron specification in worms, flies and mice: Conserved and 'lost' mechanisms. *Curr Opin Genet Dev* 12:558–564.
9. Denes AS, et al. (2007) Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell* 129:277–288.
10. Cornell RA, Ohlen TV (2000) Vnd/nkx, ind/gsh, and msh/msx: Conserved regulators of dorsoventral neural patterning? *Curr Opin Neurobiol* 10:63–71.
11. Lowe CJ, et al. (2006) Dorsoventral patterning in hemichordates: Insights into early chordate evolution. *PLoS Biol* 4:e291.
12. Pani AM, et al. (2012) Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* 483:289–294.
13. Goulding M (2009) Circuits controlling vertebrate locomotion: Moving in a new direction. *Nat Rev Neurosci* 10:507–518.
14. Lu DC, Niu T, Alaynick WA (2015) Molecular and cellular development of spinal cord locomotor circuitry. *Front Mol Neurosci* 8:25.
15. Gray PA (2008) Transcription factors and the genetic organization of brain stem respiratory neurons. *J Appl Physiol* 104:1513–1521.
16. Hobert O (2008) Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc Natl Acad Sci USA* 105:20067–20071.
17. Lee S, et al. (2008) A regulatory network to segregate the identity of neuronal subtypes. *Dev Cell* 14:877–889.
18. Joshi K, Lee S, Lee B, Lee JW, Lee SK (2009) LMO4 controls the balance between excitatory and inhibitory spinal V2 interneurons. *Neuron* 61:839–851.
19. Tosches MA, Bucher D, Vopalensky P, Arendt D (2014) Melatonin signaling controls circadian swimming behavior in marine zooplankton. *Cell* 159:46–57.
20. Lauri A, et al. (2014) Development of the annelid axochord: Insights into notochord evolution. *Science* 345:1365–1368.
21. Conzelmann M, et al. (2013) Conserved MIP receptor-ligand pair regulates *Platynereis* larval settlement. *Proc Natl Acad Sci USA* 110:8224–8229.
22. Williams EA, Jékely G (2016) Towards a systems-level understanding of development in the marine annelid *Platynereis dumerilii*. *Curr Opin Genet Dev* 39:175–181.
23. Batista MF, Jacobstein J, Lewis KE (2008) Zebrafish V2 cells develop into excitatory CiD and Notch signalling dependent inhibitory VeLD interneurons. *Dev Biol* 322:263–275.
24. Fischer AH, Henrich T, Arendt D (2010) The normal development of *Platynereis dumerilii* (Nereididae, Annelida). *Front Zool* 7:31.
25. Asadulina A, Panzera A, Verasztó C, Liebig C, Jékely G (2012) Whole-body gene expression pattern registration in *Platynereis* larvae. *Evodevo* 3:27.
26. Tomer R, Denes AS, Tessmar-Raible K, Arendt D (2010) Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* 142:800–809.
27. Steinmetz PR, Kostyuchenko RP, Fischer A, Arendt D (2011) The segmental pattern of *otx*, *gbx*, and *Hox* genes in the annelid *Platynereis dumerilii*. *Evol Dev* 13:72–79.
28. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9: 2579–2605.
29. Heckscher ES, et al. (2014) Atlas-builder software and the eNeuro atlas: Resources for developmental biology and neuroscience. *Development* 141:2524–2532.
30. Peng H, et al. (2011) BrainAligner: 3D registration atlases of *Drosophila* brains. *Nat Methods* 8:493–500.
31. Roberts DG, Johnsonbaugh HB, Spence RD, MacKenzie-Graham A (2016) Optical clearing of the mouse central nervous system using passive CLARITY. *J Vis Exp* 112: e54025.
32. Kammermeier L, et al. (2001) Differential expression and function of the *Drosophila Pax6* genes eyeless and twin of eyeless in embryonic central nervous system development. *Mech Dev* 103:71–78.
33. Adolf B, Bellipanni G, Huber V, Bally-Cuif L (2004) atoh1.2 and beta3.1 are two new bHLH-encoding genes expressed in selective precursor cells of the zebrafish anterior hindbrain. *Gene Expr Patterns* 5:35–41.
34. Davis RJ, Shen W, Sandler YI, Heanue TA, Mardon G (2001) Characterization of mouse Dach2, a homologue of *Drosophila* dachshund. *Mech Dev* 102:169–179.
35. Achim K, Arendt D (2014) Structural evolution of cell types by step-wise assembly of cellular modules. *Curr Opin Genet Dev* 27:102–108.
36. Petracca YL, et al. (2016) The late and dual origin of cerebrospinal fluid-contacting neurons in the mouse spinal cord. *Development* 143:880–891.
37. Dale N, Roberts A, Ottersen OP, Storm-Mathisen J (1987) The morphology and distribution of 'Kolmer-Agduhr cells', a class of cerebrospinal-fluid-contacting neurons revealed in the frog embryo spinal cord by GABA immunocytochemistry. *Proc R Soc Lond B Biol Sci* 232:193–203.
38. Vigh-Teichmann I, Vigh B (1983) The system of cerebrospinal fluid-contacting neurons. *Arch Histol Jpn* 46:427–468.
39. Arendt D, Benito-Gutierrez E, Brunet T, Marlow H (2015) Gastric pouches and the mucociliary sole: Setting the stage for nervous system evolution. *Philos Trans R Soc Lond B Biol Sci* 370(1684):20150286.
40. Montgomery JE, Wiggin TD, Rivera-Perez LM, Lillesaar C, Masino MA (2016) Intraspinal serotonergic neurons consist of two, temporally distinct populations in developing zebrafish. *Dev Neurobiol* 76:673–687.
41. Achim K, et al. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* 33:503–509.
42. Randel N, et al. (2014) Neuronal connectome of a sensory-motor circuit for visual navigation. *eLife* 3:3.
43. Tessmar-Raible K, Steinmetz PRH, Snyman H, Hassel M, Arendt D (2005) 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.
44. Cardona A, et al. (2012) TrakEM2 software for neural circuit reconstruction. *PLoS One* 7:e38011.
45. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9:676–682.
46. Dichmann DS, Harland RM (2011) Nkx6 genes pattern the frog neural plate and Nkx6.1 is necessary for motoneuron axon projection. *Dev Biol* 349:378–386.
47. Talikka M, Stefani G, Brivanlou AH, Zimmerman K (2004) Characterization of *Xenopus* Phox2a and Phox2b defines expression domains within the embryonic nervous system and early heart field. *Gene Expr Patterns* 4:601–607.

www.manaraa.com