PERSPECTIVES

Homeobox genes and the specification of neuronal identity

Oliver Hobert

Abstract | The enormous diversity of cell types that characterizes any animal nervous system is defined by neuron-type-specific gene batteries that endow cells with distinct anatomical and functional properties. To understand how such cellular diversity is genetically specified, one needs to understand the gene regulatory programmes that control the expression of cell-type-specific gene batteries. The small nervous system of the nematode *Caenorhabditis elegans* has been comprehensively mapped at the cellular and molecular levels, which has enabled extensive, nervous system-wide explorations into whether there are common underlying mechanisms that specify neuronal cell-type diversity. One principle that emerged from these studies is that transcription factors termed 'terminal selectors' coordinate the expression of individual members of neuron-type-specific gene batteries, thereby assigning unique identities to individual neuron types. Systematic mutant analyses and recent nervous system-wide expression analyses have revealed that one transcription factor family, the homeobox gene family, is broadly used throughout the entire *C. elegans* nervous system to specify neuronal identity as terminal selectors. I propose that the preponderance of homeobox genes in neuronal identity control is a reflection of an evolutionary trajectory in which an ancestral neuron type was specified by one or more ancestral homeobox genes, and that this functional linkage then duplicated and diversified to generate distinct cell types in an evolving nervous system.

Ramón y Cajal was the first person to appreciate the enormous diversity of cell types that characterizes all animal nervous systems[1]. Initial descriptions of neuronal diversity were based on anatomical features and, with the advent of electrophysiology, extended to the description of diverse biophysical features of neurons[2]. Over the past few years, orthogonal molecular profiling studies have substantially deepened our appreciation of the diversity of neuronal cell types, demonstrating that distinct neuronal cell types can be most precisely classified on the basis of neuron-type-specific gene batteries[3,4]. These gene batteries encode the many functional modules that define the discrete phenotypic features of a neuron[5]. To understand how the diversity of cell types is generated, one therefore needs to understand the nature of the gene regulatory programmes that specify the distinct transcriptional states

of individual neuronal cell types. In this Perspective article, I propose the hypothesis that one specific class of transcription factors, encoded by homeobox genes, plays a central role in defining the transcriptional states of neurons throughout the nervous system. Rather than trying to provide a comprehensive review of homeobox gene function in the nervous system in many animal species, I take a panoramic view of the entire nervous system of a simple model organism, the nematode *Caenorhabditis elegans*. The limited anatomical complexity and well-characterized cell types of *C. elegans* have enabled nervous system-wide differentiation studies in many different cellular contexts, resulting in the elucidation of common principles of neuronal cell-type specification. First, I briefly describe how homeobox genes were originally identified as regulators of neuronal identity, then how they were subsequently categorized

as 'terminal selectors' of neuronal identity and finally how recent comprehensive, nervous system-wide studies have revealed their importance in neuronal identity control in *C. elegans*. I propose that these findings provide a window into the evolution of neuronal cell-type diversity and that they may provide a guide for cell-type classification and functional genetic analysis in more-complex nervous systems.

Discovery of homeobox genes

Homeobox genes encode homeodomain proteins, which are transcription factors that are defined by the presence of the conserved 60 amino acid homeodomain that directly contacts DNA[6]. Homeobox genes were initially discovered in the fruit fly *Drosophila melanogaster*, in which mutations in a specific subset of homeobox genes, the Hox genes, result in defects in the organization and identity of specific segmented body structures[7–10]. These Hox genes were found to be organized into linear chromosomal clusters. Ensuing searches based on sequence homology revealed that, in addition to the clustered Hox genes, animal genomes contain numerous other homeobox genes that are distributed throughout the genome[6,7,11–13].

Homeodomain proteins are deeply conserved across phylogeny and can be classified into various subfamilies on the basis of sequence features within the homeodomain and/or the presence of additional domains, such as the PRD, SIX, CUT, LIM or POU domains, which are involved in DNA binding and/or protein–protein interactions[6] (FIG. 1). Protostomes contain about 100 homeobox genes (for example, *C. elegans* has 102 homeobox genes, of which six are Hox cluster genes, and *D. melanogaster* has 103 homeobox genes, of which eight are Hox cluster genes), whereas vertebrate genomes encode ~250–300 homeobox genes (for example, the mouse *Mus musculus* has 278 homeobox genes, of which 39 are Hox cluster genes[6] (FIG. 1). Within any animal genome, ~15% of all transcription factors are homeodomain proteins[14].

Homeobox genes were already present in unicellular organisms, but radiation of the gene family into many different classes occurred with the advent of

animal multicellularity[15]. During this radiation event, an ancestral Antennapedia homologue is thought to have undergone duplication to generate Hox cluster genes and to have been recruited into the specification of repeated body structures[16,17]. The last common ancestor of bilaterian animals is predicted to contain 56 homeobox genes (of which nine are Hox cluster genes; FIG. 1), including most major subfamilies of homeobox genes (such as LIM and POU subfamilies)[18].

Nervous system homeobox genes
The initial identification of homeobox genes that specify fate in the nervous system occurred just a few years after discovery of

the Hox cluster genes, again through genetic loss-of-function studies in *D. melanogaster* and *C. elegans*[19–31]. The *D. melanogaster* homeobox gene *cut* was discovered because of the effect of its loss on the specification of peripheral sensory neuron identity[19]. *ftz* and *eve*, which were initially discovered for their role in early embryonic segmentation, were found to control neuronal fate specification in the *Drosophila* central nervous system[20,21]. In parallel to these fly studies, homeobox genes were identified in *C. elegans* on the basis of specific behavioural and/or neuronal cell lineage defects associated with the loss of these genes[22–24]. The first two homeobox genes that were discovered through mutational analysis, *unc-86* and

mec-3, were identified on the basis of specific mechanosensory neuron defects resulting from loss of these genes[25,26]. The ensuing molecular analysis of locomotion-defective mutants (*unc-4*, *unc-30* and *unc-42*)[27–29] and thermotaxis mutants (*ttx-1* and *ttx-3*)[30,31] identified additional homeobox genes that are required for correct nervous system development and function. Following on from these initial studies in invertebrate model organisms, vertebrate homeobox genes were discovered through sequence homology to their invertebrate counterparts, and several of these genes were expressed highly selectively within the CNS[32,33]. The functions of these genes were then studied by the creation of mutant mice,
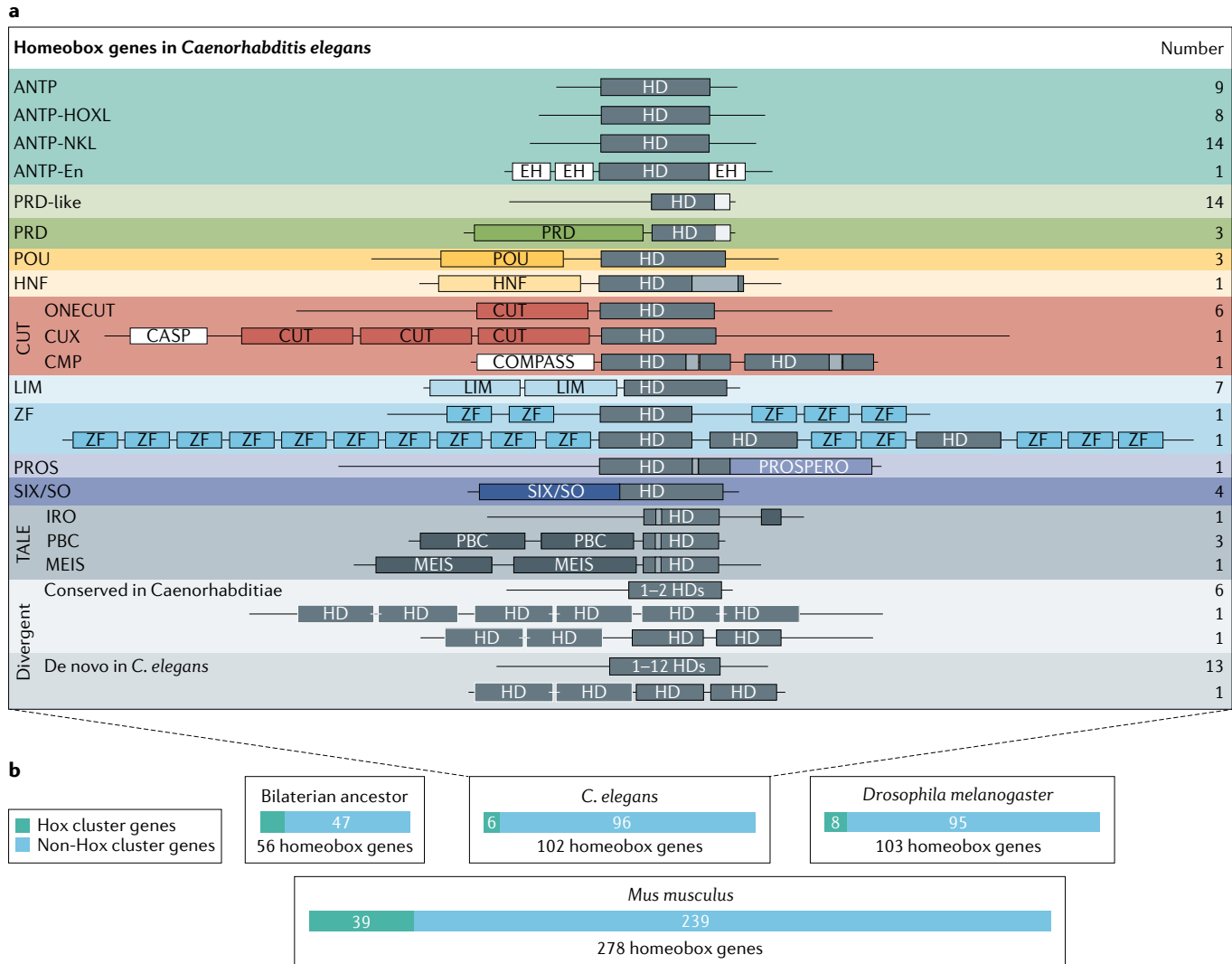

Fig. 1 | **Homeobox gene complement in some model animal genomes.** **a** | The domain structure of all homeobox genes identified in the nematode *Caenorhabditis elegans*[68,69], with different homeobox gene subfamilies indicated by coloured shading and the number of family members indicated on the right. HD indicates homeodomain; all other abbreviations are names of additional domains found in homeodomain proteins (for example, zinc finger (ZF))[6]. **b** | Comparison of the overall number of homeobox genes in three key model organisms: *C. elegans*, the fruit fly *Drosophila melanogaster* and the mouse *Mus musculus*. Bar length is proportional to the number of genes and illustrates that Hox cluster genes represent a small subset of all homeobox genes[6]. 'Bilaterian ancestor' refers to an extrapolation of the complement of homeobox genes in a last common ancestor of all bilaterians[18]. Part **a** adapted from REF.[69], Springer Nature Limited.

demonstrating, for example, the involvement of a mouse Engrailed homologue in cerebellar patterning[34], a Distal-less homologue in forebrain patterning[35] and LIM-homeodomain proteins in spinal cord patterning[36,37].

Homeobox genes as terminal selectors

For most of the *C. elegans* homeobox genes that were identified on the basis of behavioural defects of the respective mutants, two features were apparent from the outset: first, the behavioural defects resulting from loss of individual homeobox genes often precisely phenocopied the behavioural defects observed after microsurgical removal of specific neuron types. For example, microsurgical removal of light touch receptor neurons phenocopied loss of the homeobox gene *mec-3* (REF.[26]), loss of the homeobox gene *unc-30* resembled loss of function of the inhibitory D-type motor neurons[28,38], removal of the AFD thermosensory neuron was phenocopied in *ttx-1* mutants[30,39], removal of the AIY interneuron, which integrates thermosensory information, was phenocopied in *ttx-3* mutants[31,39] and loss of *unc-42* phenocopied removal of the ASH neuron[29]. Second, these homeobox genes were expressed in the same neuron types whose function they so profoundly affected, from the birth and throughout the life of these neurons[28,30,31,40]. These observations indicated that homeobox genes may have roles in initiating and perhaps also maintaining key functional features of specific neuron types.

This hypothesis was tested by leveraging a key toolkit that was developed by the *C. elegans* research community, namely, a large collection of reporter transgenes that mark the terminally differentiated state of a neuron[41]. Around 1,000 such reporter genes with known expression patterns in the nervous system exist, providing, on average, more than 20 molecular markers of the differentiated state of every neuron type[42]. These tools have been used extensively to precisely define the effect of loss of specific homeobox genes on neuronal differentiation[41]. In theory, one could envision two models for the roles of homeobox genes in neuronal differentiation (FIG. 2): homeobox genes could affect the expression of a small number of key functional features of a neuron or, alternatively, they could have very broad, coordinated effects on the entire differentiation programme of a neuron[41]. These two scenarios represent a general conceptual framework for how to think

about neuronal differentiation programmes: are these programmes controlled in a piecemeal manner with distinct regulatory factors controlling distinct phenotypic features of a neuron, or is the expression of distinct identity features coordinated via a common regulatory strategy (FIG. 2)?

The picture that emerged from analysing the expression of these molecular markers was consistent across many *C. elegans* homeobox gene mutants in many different cell types; individual homeobox genes indeed coordinate the expression of many different identity features of a neuron. For example, loss of *mec-3* affected the expression of all tested molecular markers of touch neuron identity[43,44], loss of *unc-30* affected all tested markers of D-type motor neuron identity[45–47], loss of *ttx-3* affected the expression of most, if not all, molecular markers of AIY neuron differentiation[48] and loss of

ceh-36 affected markers of AWC neuron differentiation[49,50]. Moreover, in several cases, the homeodomain proteins were found to bind directly to the *cis*-regulatory control region of molecular marker genes whose expression genetically required the respective homeobox gene[43,46–48]. In all mutants analysed, the respective neuron was still generated and expressed pan-neuronal markers (such as RAB-3 and other proteins involved in synaptic vesicle biology), but the expression of neuron-type-specific markers was significantly reduced or completely lost. Collectively, these studies led to the concept of 'terminal selectors', which are defined as transcription factors that coordinate the expression of the unique identity features of individual neuron types[51–53]. Recent bioinformatic analysis confirmed an enrichment of binding sites for validated or predicted terminal selectors in the *cis*-regulatory region of genes whose

a Scenario 1: piecemeal regulation by distinct TF complexes
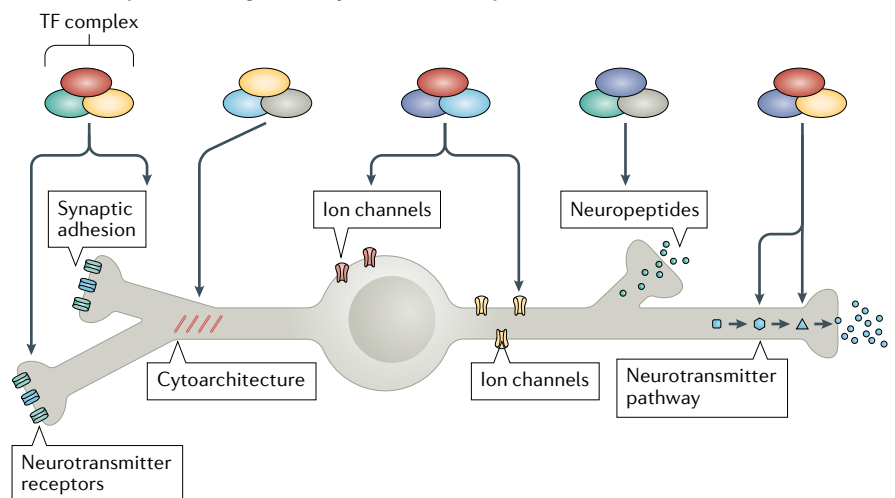
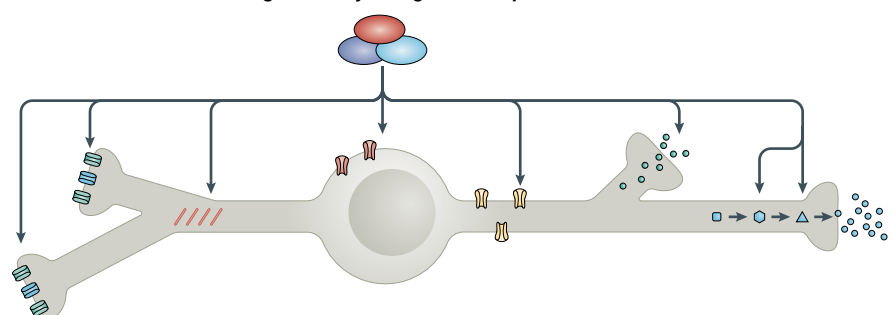**b** Scenario 2: coordinated regulation by a single TF complex

Fig. 2 | **Models for the regulation of neuron-type-specific gene batteries by transcription factors.** In scenario 1, distinct phenotypic features of a differentiated neuron are regulated by distinct cohorts or complexes of transcription factors (TFs). In scenario 2, a cohort of transcription factors together co-regulates all identity features. Ample evidence exists to support scenario 2, mostly from work in *Caenorhabditis elegans*, and the transcription factors involved have been called 'terminal selectors'. Combinations of terminal selectors have also been called 'core regulatory complexes'. Homeodomain transcription factors are most prominently, although not exclusively, used as terminal selectors. Adapted from REF.[54], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

a

- ● Homeobox terminal selector (86 classes)
- ● Homeobox subtype diversification (6 classes)
- ● No yet fully analysed for homeobox involvement (26 classes)
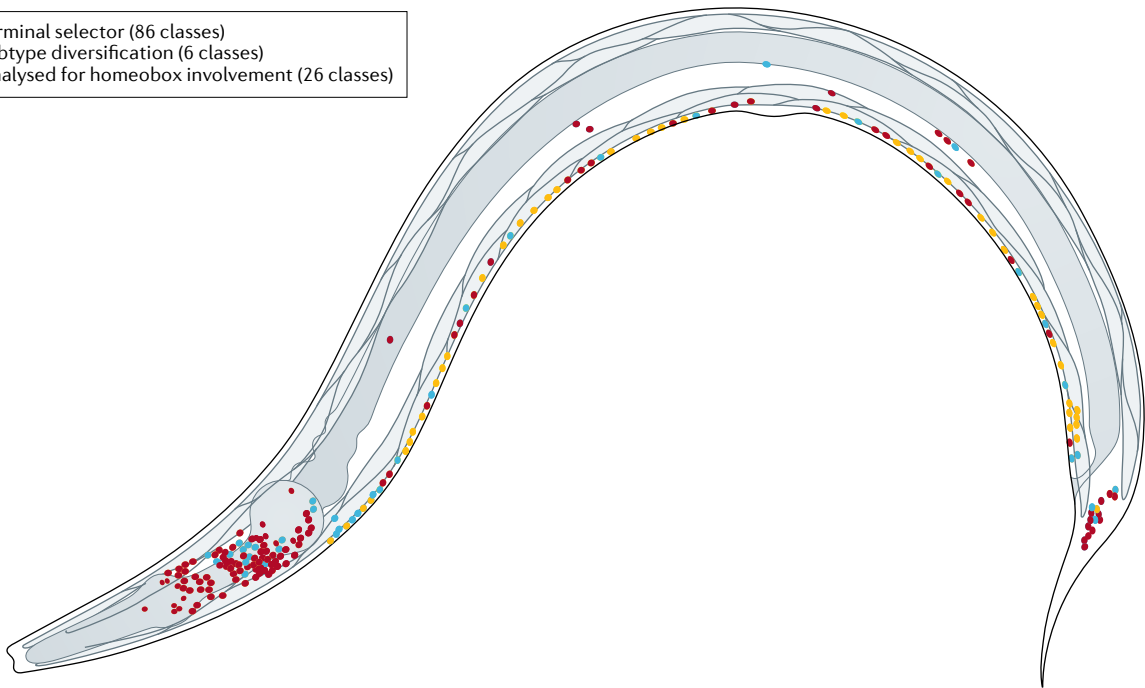
b
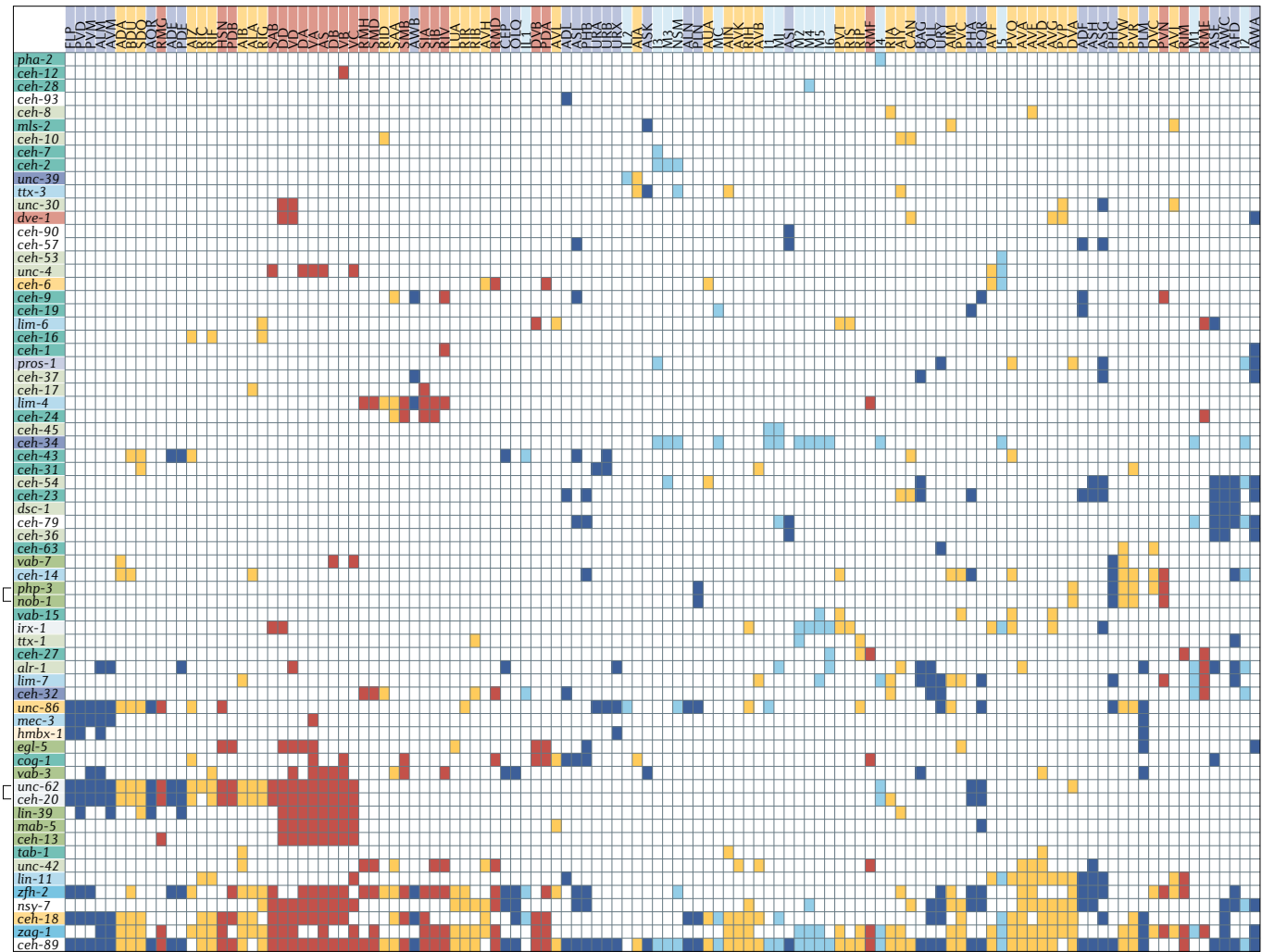
◀ Fig. 3 | **Homeobox gene expression in the nervous system** *of Caenorhabditis elegans.* **a** | The current (presently incomplete) status of functional analysis of homeobox gene expression in neurons of the nematode hermaphrodite (based on data reviewed in REF.[41] and additional updates from REF.[69]). Homeobox genes have been identified as terminal selectors in some neurons through genetic loss-of-function analysis; that is, multiple terminal markers fail to be expressed in the respective neuron class in at least one homeobox gene mutant. In other neurons, a homeobox gene is required for correct identity specification but does not act as a terminal selector; instead, the homeobox gene either acts as a repressor to diversify ('individuate') neuronal identities or controls only a limited number of identity features. For some neurons, the function of homeobox genes has not yet been analysed. **b** | A matrix indicating the expression patterns of *Caenorhabditis elegans* homeobox genes that show neuron-type-specific expression[69]. Non-neuronally, pan-neuronally or ubiquitously expressed homeobox genes are not shown. Homeobox genes (rows) are coloured by subfamily and clustered by similarity of expression pattern. Neuron classes (columns) are coloured by neuron type (sensory (purple), motor (red), interneuron (yellow) and pharyngeal (blue)) and are clustered by similarity of homeobox gene expression. The matrix illustrates that each neuron expresses a unique combination of homeobox genes. Expression levels of homeobox genes differ across cell types, but the extent to which these different levels have a functional meaning is presently unclear. Part **a** adapted from REF.[66], Springer Nature Limited. Part **b** adapted from REF.[69], Springer Nature Limited.

transcripts are selectively expressed in specific neuron types, thereby corroborating the hypothesis that terminal gene batteries are indeed coordinately controlled by neuron-type-specific terminal selectors[54].

As more regulators of neuronal identity were discovered over the years in the *C. elegans* nervous system, it also became clear that terminal selectors usually work in a combinatorial manner as heteromeric assemblies (also known as 'core regulatory complexes'[5]) to initiate and maintain the expression of terminal identity features[53]. The first well-characterized example of such a combinatorial code involved the POU-homeodomain transcription factor UNC-86 and the LIM-homeodomain transcription factor MEC-3, which bind cooperatively to *cis*-regulatory motifs in touch neuron-specific genes[43,44,55]. UNC-86 is now known to act as a terminal selector to specify the identity of at least 20 different neuron classes in *C. elegans*, and in each case, it is thought to act in the context of distinct combinations of transcription factors[56]. Apart from its collaboration with MEC-3 in specifying touch receptor neurons[44,55], UNC-86 also cooperates with, for example, the zinc-finger transcription factor PAG-3 to specify the fate of the BDU neurons[57], the LIM-homeodomain protein CEH-14 to specify AIM and PHC neurons[58,59], the LIM-homeodomain protein TTX-3 to specify NSM neuron identity[60], and the ETS-domain transcription factor AST-1 to specify HSN neuron identity[61]. In turn, transcription factors such as TTX-3 and CEH-14, which are each expressed in additional neuron classes, act in combination with additional factors to control the identity of other neuron classes[41]. This combinatorial code of neuron specification is a common theme of transcription factor activity in cell-type

specification[62], providing a conceptually simple solution for how a limited cohort of transcription factors encoded in any animal genome can control the expression of many different gene batteries in distinct cell types. Importantly, combinatorial codes of neuronal identity specifiers are not limited to two factors. For example, in dopaminergic neurons, the combinatorial signature of identity specifiers includes four transcription factors, and in HSN neurons, the combinatorial signature of identity specifiers includes as many as six transcription factors[61].

The underlying mechanism by which ensembles of terminal selectors cooperate to activate terminal gene batteries can be quite diverse. In some cases, it is known that combinatorially acting factors physically interact to bind cooperatively to DNA to then activate target gene expression; examples include the UNC-86–MEC-3 and TTX-3–CEH-10 heteromeric complexes[48,55]. In other cases, terminal selectors act as so-called transcription factor collectives that independently assemble on a given target gene promoter to then jointly recruit the basal transcriptional machinery to activate gene expression[63,64].

Reducing neuronal complexity to homeoboxes alone
How far can one take the involvement of homeobox genes in neuronal identity specification? *C. elegans* offers a unique opportunity to address this question because anatomical studies, complemented by more recent molecular studies, have precisely defined the composition of its entire nervous system. *C. elegans* hermaphrodites contain 302 neurons that fall into 118 anatomically distinct classes, on the basis of cell body position, axodendritic projections and synaptic connectivity[65]. Some of these

neuron classes can be further divided into subclasses on the basis of subtle anatomical and molecular differences between individual members of the class[42,65–67]. To what extent can homeobox genes be implicated in the specification of the entire nervous system of the nematode? As mentioned earlier, the *C. elegans* genome encodes 102 homeobox genes, most of which are phylogenetically conserved[68,69] (FIG. 1). A recent analysis of the expression pattern of all 102 *C. elegans* homeodomain proteins throughout the entire nervous system revealed a striking picture[69] (FIG. 3): about 80% of the homeodomain proteins are expressed in the nervous system. Strikingly, on aggregate, every single neuron class in the hermaphrodite not only expresses multiple homeodomain proteins but also expresses a unique combination of homeodomain proteins[69] (FIG. 3). Most of the individual homeodomain proteins are expressed in only a few of the 118 neuron classes, but every individual neuron class expresses on average five homeobox genes. Thus, the diversity of all anatomically defined cell types in the entire nervous system can be captured by one class of transcription factors.

The classification of individual neuron classes into subclasses on the basis of functional or anatomical criteria can also be mirrored by distinct patterns of expression of homeodomain proteins. For example, the six IL2 sensory neurons are composed of three bilaterally symmetrical, highly similar neuron pairs (a dorsal, a lateral and a ventral pair) that share their overall anatomy and synaptic connectivity. However, the lateral pair has some slight differences in synaptic connectivity and dendritic morphology[65]. This subclassification is mirrored by the expression of distinct sets of homeodomain proteins. All six IL2 neurons share expression of the homeodomain protein UNC-86, but the lateral pair expresses one additional homeodomain protein, UNC-39 (REF.[69]). In the case of ventral nerve cord motor neurons, homeodomain protein expression patterns even revealed subtypes that were not predicted on the basis of anatomical features, illustrating the impressive granularity with which homeodomain proteins can define neuronal subtypes[69]. The notion that homeodomain protein expression patterns may distinguish all neuronal cell types is an extension of the previously proposed operational definition of a cell type as "a set of cells accessing the same regulatory program driving differentiation"[70].

The ability of homeobox genes to precisely demarcate individual neuron

classes may extend to other animals besides *C. elegans*. Recent transcriptomic studies in flies and mice, which focused on specific parts of the nervous system, have also found that combinatorial homeobox gene expression patterns represent the best classifiers of distinct neuron cell populations[71–73]. Vertebrate nervous systems contain a greater number of neuronal cell types than *C. elegans* but, at least in theory, building greater complexity does not require novel players or mechanisms; instead, complexity can arise from the diversity of interactions between individual components (homeobox genes) of a system.

Does the functional analysis of homeobox genes support the notion that, in *C. elegans*, every single neuron class is indeed specified by one or a combination of homeobox genes? Functional analysis

of individual homeobox genes in *C. elegans* has progressed significantly since the initial isolation of homeobox gene mutants from behavioural screens[41,69]. As illustrated in FIGS 3,4, extensive analysis of homeobox gene mutants that were generated by directed gene removal has so far implicated at least one homeobox gene in the identity specification of 98 of the 118 different neuron classes (the remaining neuron classes have not yet been analysed for the involvement of homeobox genes)[41,69]. The theme uncovered by the first in-depth analyses of homeobox genes such as *mec-3*, *ttx-3* and *unc-30* (*LHX1* and *LHX5*, *LHX2* and *LHX9*, and *PITX1–PITX3*, respectively, in vertebrates) seems to hold up in other cases as well; these genes usually control the expression of many key identity features, but they are not required for the neurons to be

specified as neurons per se. Loss-of-function studies that implicated homeobox genes in neuronal identity specification have in some select cases been further bolstered by ectopic expression studies showing that misexpression of specific homeobox genes can reprogramme the identity of specific neuron types[28,44,48,49,57,74].

Many studies in flies and vertebrates underscore the function of conserved homeodomain proteins as potential terminal selectors of neuronal identity. Classic examples include the *Drosophila* LIM-homeodomain proteins Apterous and Islet[75,76] and, in vertebrates, the PRD-type homeodomain protein CRX in photoreceptors[77], the LIM-homeodomain protein LHX2 in olfactory neurons[78,79], the POU-homeodomain protein BRN3 in several parts of the CNS[56], the DLX family homeodomain proteins in forebrain GABAergic interneurons[80,81] and the homeodomain protein PITX3 in dopaminergic neurons[82]. These factors probably act in combinatorial partnerships with other transcription factors, but the identification of more comprehensive combinatorial cohorts requires extensive, time-consuming genetic mutant analysis that is not as facile in flies and mice as it is in *C. elegans*. The ease, speed and depth with which such mutant analysis can be conducted in *C. elegans* seems to be the key reason why deciphering the regulatory logic of terminal differentiation programmes in the nervous system is comparatively far advanced in *C. elegans*.

Homeobox genes and subtype diversification

Homeobox genes do not act exclusively as terminal selectors to activate neuron-type-specific gene batteries; they also play roles in making highly related neurons become dissimilar from one another, a process that I term here 'subtype diversification'. In the nervous system of *C. elegans* as well as in that of other organisms, neuronal cell types are not equally dissimilar from one another but differ from each other to some extent, as revealed by both anatomical classification and hierarchical clustering of molecular profiles[42,65]. For example, cholinergic ventral nerve cord motor neurons are classified by anatomical and molecular features into A-type and B-type classes. UNC-3 is a zinc-finger transcription factor and non-homeodomain terminal selector for both of these cholinergic motor neuron types[83], but the difference between these two classes arises through the action of the
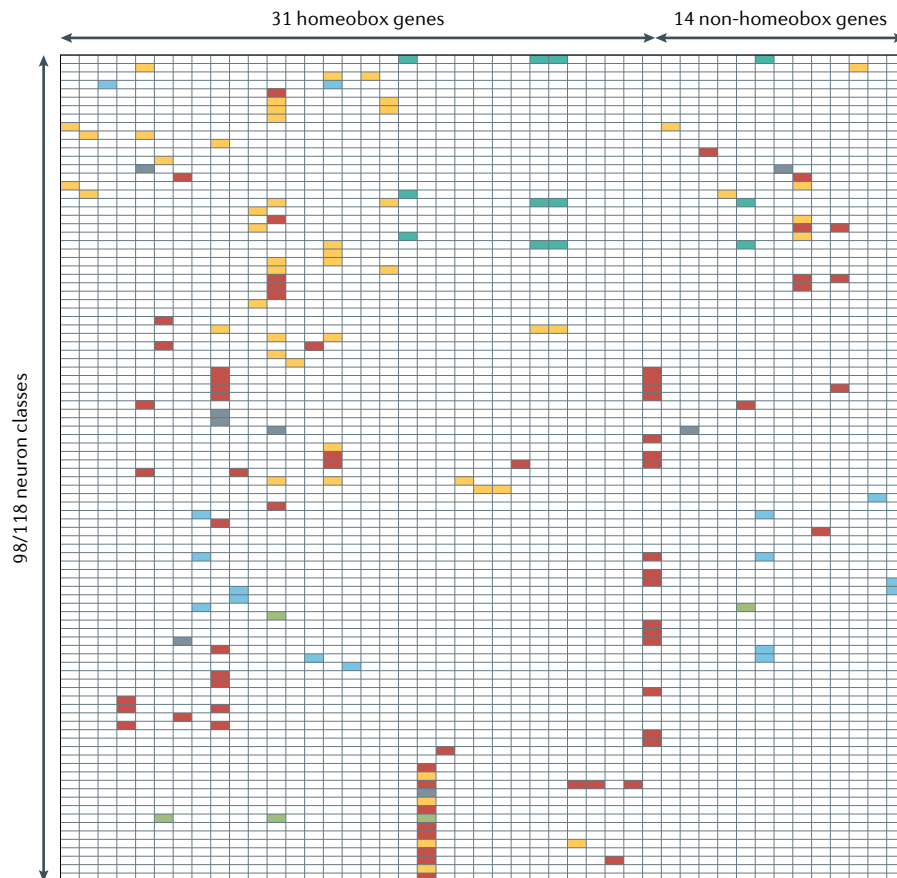

Fig. 4 | **Terminal selectors in *Caenorhabditis elegans*.** This matrix summarizes the current status of identifying terminal selectors in *Caenorhabditis elegans* (data are from REFS[41,69]), irrespective of whether they are homeodomain proteins or other types of transcription factors, such as C2H2 zinc-fingers, orphan nuclear receptors and ETS-type transcription factors. Each coloured box represents a gene (rows) shown to be involved, through genetic loss-of-function analysis, in the specification of the identity of a neuron class (columns), as defined by the loss of expression of multiple marker genes. The neurotransmitters released by neurons are indicated by the colour of the box: cholinergic (red), glutamatergic (yellow), dopaminergic (dark green), serotonergic (light green) and GABAergic (blue). 'Orphan' neurons for which no neurotransmitter has yet been identified are in grey. In most cases in which a non-homeobox gene has been identified as a terminal selector, it cooperates with a homeobox-type terminal selector.

homeodomain transcription factor UNC-4, which is a transcriptional repressor that is expressed only in A-type but not B-type ventral motor neurons. Loss of *unc-4* leads to a conversion of A-type to B-type neuron identity[84–86]. UNC-4 is thought to antagonize the ability of UNC-3 to activate subsets of target genes, thereby giving this homeodomain protein the ability to diversify neuronal identity[83,87]. Other homeobox genes, such as the *eve* homologue *vab-7*, the *MNX1* homologue *ceh-12* and the Iroquois-class homeodomain gene *irx-1*, also play roles in distinguishing other classes of ventral cord motor neurons[86–89]. In the context of diversifying neuronal identity across the left–right axis, homeobox genes have also been found to act downstream of a terminal selector to antagonize its activating effect on specific downstream target genes[90,91].

Motor neuron classes in the *C. elegans* ventral nerve cord, as well as those in the spinal cord of vertebrates, are organized along the anterior–posterior axis and, in this context, Hox cluster genes have taken on important roles in specifying different neuronal identities along that axis[92–95]. In *C. elegans*, Hox proteins interact with the terminal selector of cholinergic neuron identity, UNC-3, to specify cholinergic motor neuron identity along the anterior–posterior axis[92,93]. Specifically, motor neurons at the most posterior end of the ventral nerve cord display several distinct molecular and anatomical features that distinguish them from more anteriorly located motor neurons. The Hox gene product EGL-5 cooperates with UNC-3 to activate these features, and in the absence of *egl-5*, homeotic transformation of these neurons to more anterior neurons is observed[93]. In peripheral touch receptor neurons, Hox gene products also serve to diversify the identity of these neurons along the anterior–posterior axis via cooperation with the UNC-86–MEC-3 terminal selector complex[96].

Neuronal identity beyond homeobox genes

Neuronal identities are not determined by homeodomain transcription factors alone. A plethora of genetic loss-of-function studies in different model organisms, including *C. elegans*, have shown that non-homeodomain transcription factors also operate as terminal selector-type specifiers of neuronal identity[41]. For example, the ARID-type transcription factor CFI-1, the transcription factor SOX-2, the ETS-domain transcription factor AST-1 and the zinc-finger protein PAG-3 operate as terminal selectors in several *C. elegans* neuron types[57,60,97,98]. However, within each specific cellular context, each one of these factors cooperates with the same homeodomain transcription factor, UNC-86 (REFS[49,57,60,61,64]). Similarly, the aforementioned non-homeodomain terminal selector UNC-3 controls the identity of several different types of command interneurons and ventral cord motor neurons and operates together with different homeodomain transcription factors (for example, UNC-42, the *C. elegans* homologue of vertebrate PROP1) in head command interneurons or CEH-14 (LHX3 and LHX4 in vertebrates) in tail command interneurons) to activate different gene batteries[99,100]. From the findings taken together, core regulatory complexes[5] comprising combinations of terminal selectors are by no means always exclusively composed of homeodomain proteins, but they most often (if not always) do contain a homeodomain component[41] (see FIG. 4, which summarizes our current understanding of neuronal identity regulators in *C. elegans*).

Clearly, the current state of analysis of transcription factor function in *C. elegans* indicates that there is no other single family of transcription factors that is as broadly used as homeodomain proteins in neuronal identity specification. This is not due to the size of the homeodomain transcription factor family, which constitutes only ~15% of all transcription factors encoded by the *C. elegans* genome. For example, although there are twice as many C2H2 zinc-finger transcription factors as homeodomain proteins, they seem to be much less frequently used as terminal selectors[41]. Moreover, a recent comprehensive single-cell RNA sequencing transcriptome analysis of all *C. elegans* neurons revealed features of homeobox genes that are not shared by other transcription factor families[67]. First, homeodomain transcription factors are expressed throughout the entire nervous system, but the vast majority of them are expressed in a very restricted, cell-type-specific manner, unlike transcription factors such as T-box or basic helix–loop–helix-type transcription factors, which are expressed much more sparsely in the terminally differentiating nervous system. Second, the expression of members of larger transcription factor families, such as the C2H2 zinc-finger family or the C4-type orphan nuclear receptor family, covers the entire nervous system, but each member tends to be much more broadly expressed

throughout the nervous system than are individual homeodomain proteins[67], most of which are expressed in less than 10% of all neuron classes. The restricted expression of members of a relatively large protein family such as the homeobox family therefore provides the ability to generate unique codes established by the expression of a limited number of proteins per neuron type (FIG. 3).

Evolutionary implications

Following Dobzhansky's dictum that "nothing makes sense in biology except in the light of evolution"[101], I propose to consider the preponderance of homeodomain proteins in neuronal identity specification in *C. elegans* in an evolutionary context. Several excellent essays and reviews have provided thoughtful ideas and conceptual frameworks for how neuronal cell types came into being and how they evolved[5,70,102–104]. Below, I rephrase several of the basic concepts discussed in these reviews by specifically including homeobox genes as the key linchpin in these proposed evolutionary processes.

One previously proposed key concept is that cell-type identity is defined by the regulatory mechanisms that specify and maintain the distinct gene expression programme of a cell type within the organism[5]. Since these regulatory mechanisms involve terminal selector complexes (core regulatory complexes), it follows that the evolution of new cellular identities must involve the evolution of novel, uniquely cell-type-specific regulatory signatures comprising novel combinations of terminal selectors[5,70]. In *C. elegans* at least, homeodomain proteins are key components of these complexes, and thus I propose that changes in homeobox gene expression are key drivers of evolutionary novelty of cell types in the nervous system. This proposal parallels previous suggestions about how Hox cluster genes drive novelties in body plans during evolution[105].

Gains and losses of homeobox gene expression on an evolutionary, phylogenetic timescale may resemble the ontogenetic process that instructs the differential induction of homeobox genes during development. Homeobox genes are usually expressed in multiple distinct neuron types, and thus multiple distinct mechanisms must be operational to direct homeobox gene expression to specific neuron types (for example, see FIG. 4). Thus, as much as *cis*-regulatory elements of terminal selector-controlled 'function genes' are thought to be substrates for evolutionary changes in phenotypic features of a neuron class[5,70],

changes in *cis*-regulatory elements of homeobox genes may result in even more dramatic changes in cellular identity. Analogously to 'homeotic' mutations in Hox cluster genes that result in alteration in the identity of segmented body structures[106], alterations in the sites of expression of neuronal homeobox genes could be thought of as resulting in homeotic transformation of neuronal identity[107]. Starting with Bateson's original description of homeotic processes, such homeotic changes have a long history in evolutionary thought[106–109].

One particularly striking example of a homeotic neuron identity transformation that could be placed in an evolutionary context is observed for the ALM and BDU mechanosensory neurons in *C. elegans* (FIG. 5). These neurons derive from the same mother neuroblast and both require the POU-homeodomain transcription factor UNC-86 as a terminal selector[57]. ALM identity is specified by the combination of UNC-86 and the ALM-restricted LIM-homeodomain transcription factor MEC-3 (REF.[44]), whereas BDU identity is specified by the combination of UNC-86 and the zinc-finger transcription factor PAG-3, which is expressed in both ALM and BDU neurons. In *mec-3* mutants, the identity of ALM transforms to that of BDU[26,57], and both the 'new' BDU neuron and the original BDU neuron now require UNC-86 and PAG-3 for their specification[57]. Thus, one can envision an evolutionary scenario in which two identical BDU sister cells constituted the ancestral state

(FIG. 5), and the gain of MEC-3 expression in one BDU sister cell reroutes UNC-86 regulatory activity to a different cohort of target genes that lead to the establishment of a novel neuronal identity (ALM). The generation of two different cells from ancestral, indistinguishable sister cells has been called 'genetic individuation'[5,110] and the ALM–BDU case may provide an example for such individuation, in which homeobox genes may play a critical role. Future mutant analysis in *C. elegans* will reveal the extent to which such homeotic identity transformation can be observed in homeobox gene mutants. The conceptual similarity between the neuronal cell identity transformation in non-Hox cluster homeobox gene mutants and the segmental transformation of body parts in Hox cluster gene mutants is notable.

Evolutionary processes that give rise to novel neuronal identities owing to changes in regulatory programmes should not be viewed only from the perspective of an isolated, individual cell. New neurons provide adaptive value only if they become integrated into functional circuits with adaptive behavioural outcome. Potential mechanisms of circuit evolution have been discussed previously elsewhere[111]. Interestingly, individual homeobox genes have been noted to be expressed in synaptically connected neurons[99,112–116], suggesting a potential role for these genes in circuit wiring. Hence, alterations in homeobox gene expression are predicted to have profound effects on circuit evolution as well.

The pervasiveness of homeobox-mediated control of neuronal identity in *C. elegans* suggests not only that homeobox expression pattern changes may be important drivers in generating novel cell types and neuronal circuits but also that homeobox genes may have an ancestral role in a key evolutionary transition, the generation of the 'first' neuron type[104]. A number of different cellular specializations are thought to have paved the way for generating such a primitive 'protoneuron': first and foremost, the ability of a cell to perceive signals from the environment and the ability of this signal perception event to be transformed into a signal to 'effector' cells, such as movement-generating myoepithelial cells[70,104]. Such protoneurons probably organized into primitive neuronal nets that enabled signal propagation over greater distances and the ability to integrate signals[70,104]. Intercellular signalling modules, composed of receptor molecules and signalling machinery (such as peptides or the secretory machinery that enabled the secretion of amino acids as transmitters), and/or homophilic cell adhesion that assembled protoneurons into networks, may have been under coordinated control of a homeodomain transcription factor. Since even unicellular organisms already contain more than one homeobox gene[15], such control may not have been exerted by a single homeodomain protein but by a heteromeric complex of several interacting homeodomain proteins. The radiation of homeobox genes that paralleled the advent of multicellularity and the origin of nervous systems may then have enabled the generation of multiple different types of homeodomain protein complexes, each with subtle differences in target gene specificity. With the acquisition of interaction surfaces (such as LIM, POU or SIX domains) in addition to their DNA-binding domain, homeodomain proteins may have had a greater propensity than other transcription factors to form multiple distinct heteromeric complexes. Similarly, effector modules (that is, genes involved in signalling (receptors, peptides and presynaptic, vesicular machinery) and also in cell–cell recognition) also duplicated, thereby allowing distinct homeodomain protein complexes to control the expression of different flavours of signalling properties, as well as endowing them with different capacities to form specialized contacts with each other.

In the model proposed here, a neuron specification function is ancestral to all metazoan homeobox genes and preceded the role of homeobox genes in defining
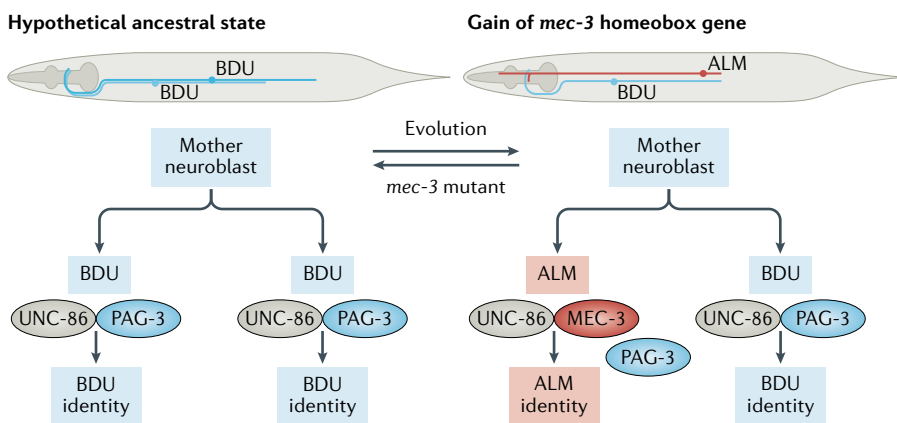

Hypothetical ancestral state   **Gain of *mec-3* homeobox gene**

Fig. 5 | **A hypothetical scenario for the individuation of two sister neurons.** In *Caenorhabditis elegans*, the ALM and BDU mechanosensory neurons arise by division of a mother neuroblast and diversification of the sister cells. In ALM, the homeodomain proteins UNC-86 and MEC-3 heterodimerize to cooperatively bind to DNA[55] to induce ALM identity[44]. However, in BDU neurons and *mec-3* mutants, UNC-86 interacts with the zinc-finger transcription factor PAG-3 to induce BDU identity. Mutant and gain-of-function studies have shown that MEC-3 'outcompetes' PAG-3 to direct UNC-86 to ALM-specific target genes[57]. These analyses can be extended to a potential evolutionary scenario in which the gain of *mec-3* expression 'individuates' (that is, diversifies) one sister cell from the other. Homeotic transformation of neuronal identity in *C. elegans* as well as other organisms has been discussed further elsewhere[107].

segmented identities of body plans. Such a body patterning role may have evolved only once a subset of the homeobox genes (the ANTP class) became organized into chromosomal clusters. Consistent with such an ancestral neuronal function, Hox cluster genes have been implicated in neuronal identity control across the animal kingdom[92–96,117,118].

Conclusions

By focusing on the extensive knowledge of neuronal cell-type specification in the simple nervous system of *C. elegans*, I propose here a central role of homeobox genes in specifying cell-type diversity in the nervous system and, furthermore, that these findings may provide an opportunity to better understand nervous system evolution, from the origin of neuronal cell types to their individuation into many different types.

To probe the hypothesis of the central importance of homeobox genes in neuronal identity control, careful examination of existing and forthcoming single-cell transcriptome datasets of various brain regions in different animal species is required. The classification power of homeobox gene combinations may not hold for all levels of cell-type taxonomy in more-complex brains, but may still define broader neuron classes. Such a scenario might reflect how initial homeobox codes were secondarily modified through the recruitment of additional gene-regulatory factors into neuron identity control. On a functional level, it will be important to move beyond the many studies that have delineated homeobox gene function in earlier patterning of the CNS (for example, see REFS[119–122]) and to consider more systematically the function of homeobox genes in initiating and maintaining terminal differentiation in the vertebrate CNS. Such 'late' phenotypes are often masked by additional functions of a gene at earlier stages of development, necessitating the use of conditional alleles to assess later, postmitotic functions in initiating and maintaining terminal differentiation programmes. Such late functions are rarely considered; for example, for the many vertebrate homeobox genes that are implicated in early spinal cord patterning, few published data exist about the expression, let alone the function, of these genes in the adult spinal cord. Similarly, the coupling of neuronal identity acquisition with neuronal survival in vertebrates, which is not observed in *C. elegans*, necessitates the use of additional genetic manipulation to prevent cell death so that the functions of late-acting homeobox genes in neuronal identity control are revealed[123]. Both expression and functional analysis should ideally be performed in as many animal model organisms as possible to probe the fundamental role of homeobox genes in neuronal identity specification.

Oliver Hobert [ID]

Department of Biological Sciences, Columbia University, Howard Hughes Medical Institute, New York, NY, USA.

e-mail: or38@columbia.edu

1. y Cajal, R. *Histologie du Système Nerveux de L'homme et des Vertébrés* (Maloine, 1911).
2. Bota, M. & Swanson, L. W. The neuron classification problem. *Brain Res. Rev.* **56**, 79–88 (2007).
3. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
4. Tasic, B. Single cell transcriptomics in neuroscience: cell classification and beyond. *Curr. Opin. Neurobiol.* **50**, 242–249 (2018).
5. Arendt, D. et al. The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
6. Burglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125**, 497–521 (2016).
7. Gehring, W. J. *Master Control Genes in Development and Evolution: The Homeobox Story* (Yale Univ. Press, 1998).
8. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308**, 428–433 (1984).
9. Scott, M. P. & Weiner, A. J. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc. Natl Acad. Sci. USA* **81**, 4115–4119 (1984).
10. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
11. Carrasco, A. E., McGinnis, W., Gehring, W. J. & De Robertis, E. M. Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell* **37**, 409–414 (1984).
12. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–408 (1984).
13. Burglin, T. R., Finney, M., Coulson, A. & Ruvkun, G. *Caenorhabditis elegans* has scores of homoeobox-containing genes. *Nature* **341**, 239–243 (1989).
14. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
15. Derelle, R., Lopez, P., Le Guyader, H. & Manuel, M. Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evol. Dev.* **9**, 212–219 (2007).
16. Holland, P. W. Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 31–45 (2013).
17. Garcia-Fernandez, J. The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* **6**, 881–892 (2005).
18. Ryan, J. F. et al. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biol.* **7**, R64 (2006).
19. Blochlinger, K., Bodmer, R., Jack, J., Jan, L. Y. & Jan, Y. N. Primary structure and expression of a product from cut, a locus involved in specifying sensory organ identity in *Drosophila*. *Nature* **333**, 629–635 (1988).
20. Doe, C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. Expression and function of the segmentation gene fushi tarazu during *Drosophila* neurogenesis. *Science* **239**, 170–175 (1988).
21. Doe, C. Q., Smouse, D. & Goodman, C. S. Control of neuronal fate by the *Drosophila* segmentation gene even-skipped. *Nature* **333**, 376–378 (1988).
22. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
23. Hedgecock, E. M. & Russell, R. L. Normal and mutant thermotaxis in the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **72**, 4061–4065 (1975).
24. Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
25. Finney, M., Ruvkun, G. & Horvitz, H. R. The *C. elegans* cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell* **55**, 757–769 (1988).
26. Way, J. C. & Chalfie, M. mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**, 5–16 (1988).
27. Miller, D. M. et al. *C. elegans* unc-4 gene encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. *Nature* **355**, 841–845 (1992).
28. Jin, Y., Hoskins, R. & Horvitz, H. R. Control of type-D GABAergic neuron differentiation by *C. elegans* UNC-30 homeodomain protein. *Nature* **372**, 780–783 (1994).
29. Baran, R., Aronoff, R. & Garriga, G. The *C. elegans* homeodomain gene unc-42 regulates chemosensory and glutamate receptor expression. *Development* **126**, 2241–2251 (1999).
30. Satterlee, J. S. et al. Specification of thermosensory neuron fate in *C. elegans* requires ttx-1, a homolog of otd/Otx. *Neuron* **31**, 943–956 (2001).
31. Hobert, O. et al. Regulation of interneuron function in the *C. elegans* thermoregulatory pathway by the ttx-3 LIM homeobox gene. *Neuron* **19**, 345–357 (1997).
32. Rubenstein, J. L. & Puelles, L. Homeobox gene expression during development of the vertebrate brain. *Curr. Top. Dev. Biol.* **29**, 1–63 (1994).
33. Tsuchida, T. et al. Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
34. Joyner, A. L., Herrup, K., Auerbach, B. A., Davis, C. A. & Rossant, J. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science* **251**, 1239–1243 (1991).
35. Qiu, M. et al. Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain. *Genes Dev.* **9**, 2523–2538 (1995).
36. Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320 (1996).
37. Hobert, O. & Westphal, H. Function of LIM homeobox genes. *Trends Genet.* **16**, 75–83 (2000).
38. McIntire, S. L., Jorgensen, E. & Horvitz, H. R. Genes required for GABA function in *Caenorhabditis elegans*. *Nature* **364**, 334–337 (1993).
39. Mori, I. & Ohshima, Y. Neural regulation of thermotaxis in *Caenorhabditis elegans*. *Nature* **376**, 344–348 (1995).
40. Way, J. C. & Chalfie, M. The mec-3 gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes Dev.* **3**, 1823–1833 (1989).
41. Hobert, O. A map of terminal regulators of neuronal identity in *Caenorhabditis elegans*. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 474–498 (2016).
42. Hobert, O., Glenwinkel, L. & White, J. Revisiting neuronal cell type classification in *Caenorhabditis elegans*. *Curr. Biol.* **26**, R1197–R1203 (2016).
43. Zhang, Y. et al. Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* **418**, 331–335 (2002).
44. Duggan, A., Ma, C. & Chalfie, M. Regulation of touch receptor differentiation by the *Caenorhabditis elegans* mec-3 and unc-86 genes. *Development* **125**, 4107–4119 (1998).
45. Eastman, C., Horvitz, H. R. & Jin, Y. Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. *J. Neurosci.* **19**, 6225–6234 (1999).
46. Cinar, H., Keles, S. & Jin, Y. Expression profiling of GABAergic motor neurons in *Caenorhabditis elegans*. *Curr. Biol.* **15**, 340–346 (2005).
47. Yu, B. et al. Convergent transcriptional programs regulate cAMP levels in *C. elegans* GABAergic motor neurons. *Dev. Cell* **43**, 212–226.e7 (2017).

48. Wenick, A. S. & Hobert, O. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev. Cell* **6**, 757–770 (2004).

49. Alqadah, A. et al. Postmitotic diversification of olfactory neuron types is mediated by differential activities of the HMG-box transcription factor SOX-2. *EMBO J.* **34**, 2574–2589 (2015).

50. Lanjuin, A., VanHoven, M. K., Bargmann, C. I., Thompson, J. K. & Sengupta, P. Otx/otd homeobox genes specify distinct sensory neuron identities in *C. elegans*. *Dev. Cell* **5**, 621–633 (2003).

51. Hobert, O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl Acad. Sci. USA* **105**, 20067–20071 (2008).

52. Hobert, O. Regulation of terminal differentiation programs in the nervous system. *Annu. Rev. Cell Dev. Biol.* **27**, 681–696 (2011).

53. Hobert, O. Terminal selectors of neuronal identity. *Curr. Top. Dev. Biol.* **116**, 455–475 (2016).

54. Glenwinkel, L. et al. In silico analysis of the transcriptional regulatory logic of neuronal identity specification throughout the *C. elegans* nervous system. *eLife* **10**, e64906 (2021).

55. Xue, D., Tu, Y. & Chalfie, M. Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science* **261**, 1324–1328 (1993).

56. Leyva-Diaz, E., Masoudi, N., Serrano-Saiz, E., Glenwinkel, L. & Hobert, O. Brn3/POU-IV-type POU homeobox genes — paradigmatic regulators of neuronal identity across phylogeny. *Wiley Interdiscip Rev. Dev. Biol.* **9**, e374 (2020).

57. Gordon, P. M. & Hobert, O. A competition mechanism for a homeotic neuron identity transformation in *C. elegans*. *Dev. Cell* **34**, 206–219 (2015).

58. Serrano-Saiz, E. et al. Modular control of glutamatergic neuronal identity in *C. elegans* by distinct homeodomain proteins. *Cell* **155**, 659–673 (2013).

59. Serrano-Saiz, E., Oren-Suissa, M., Bayer, E. A. & Hobert, O. Sexually dimorphic differentiation of a *C. elegans* hub neuron is cell autonomously controlled by a conserved transcription factor. *Curr. Biol.* **27**, 199–209 (2017).

60. Zhang, F. et al. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Development* **141**, 422–435 (2014).

61. Lloret-Fernandez, C. et al. A transcription factor collective defines the HSN serotonergic neuron regulatory landscape. *eLife* **7**, e32785 (2018).

62. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).

63. Junion, G. et al. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**, 473–486 (2012).

64. Doitsidou, M. et al. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev.* **27**, 1391–1405 (2013).

65. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).

66. Cook, S. J. et al. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**, 63–71 (2019).

67. Taylor, S. R. et al. Molecular topography of an entire nervous system. *Cell* **184**, 4329–4347.e23 (2021).

68. Hench, J. et al. The homeobox genes of *Caenorhabditis elegans* and insights into their spatio-temporal expression dynamics during embryogenesis. *PLoS ONE* **10**, e0126947 (2015).

69. Reilly, M. B., Cros, C., Varol, E., Yemini, E. & Hobert, O. Unique homeobox codes delineate all the neuron classes of *C. elegans*. *Nature* **584**, 595–601 (2020).

70. Arendt, D., Bertucci, P. Y., Achim, K. & Musser, J. M. Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).

71. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).

72. Davis, F. P. et al. A genetic, genomic, and computational resource for exploring neural circuit function. *eLife* **9**, e50901 (2020).

73. Sugino, K. et al. Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. *eLife* **8**, e38619 (2019).

74. Sagasti, A., Hobert, O., Troemel, E. R., Ruvkun, G. & Bargmann, C. I. Alternative olfactory neuron fates are specified by the LIM homeobox gene lim-4. *Genes Dev.* **13**, 1794–1806 (1999).

75. O'Keefe, D. D., Thor, S. & Thomas, J. B. Function and specificity of LIM domains in *Drosophila* nervous system and wing development. *Development* **125**, 3915–3923 (1998).

76. Thor, S. & Thomas, J. B. The *Drosophila* islet gene governs axon pathfinding and neurotransmitter identity. *Neuron* **18**, 397–409 (1997).

77. Swaroop, A., Kim, D. & Forrest, D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat. Rev. Neurosci.* **11**, 563–576 (2010).

78. Zhang, G., Titlow, W. B., Biecker, S. M., Stromberg, A. J. & McClintock, T. S. Lhx2 determines odorant receptor expression frequency in mature olfactory sensory neurons. *eNeuro* **3**, ENEURO.0230-16.2016 (2016).

79. Monahan, K. et al. Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *eLife* **6**, e28620 (2017).

80. Pla, R. et al. Dlx1 and Dlx2 promote interneuron GABA synthesis, synaptogenesis, and dendritogenesis. *Cereb. Cortex* **28**, 3797–3815 (2018).

81. Lindtner, S. et al. Genomic resolution of DLX-orchestrated transcriptional circuits driving development of forebrain GABAergic neurons. *Cell Rep.* **28**, 2048–2063.e8 (2019).

82. Smidt, M. P. & Burbach, J. P. Terminal differentiation of mesodiencephalic dopaminergic neurons: the role of Nurr1 and Pitx3. *Adv. Exp. Med. Biol.* **651**, 47–57 (2009).

83. Kratsios, P., Stolfi, A., Levine, M. & Hobert, O. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat. Neurosci.* **15**, 205–214 (2011).

84. White, J. G., Southgate, E. & Thomson, J. N. Mutations in the *Caenorhabditis elegans* unc-4 gene alter the synaptic input to ventral cord motor neurons. *Nature* **355**, 838–841 (1992).

85. Winnier, A. R. et al. UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in *Caenorhabditis elegans*. *Genes Dev.* **13**, 2774–2786 (1999).

86. Von Stetina, S. E. et al. UNC-4 represses CEH-12/HB9 to specify synaptic inputs to VA motor neurons in *C. elegans*. *Genes Dev.* **21**, 332–346 (2007).

87. Kerk, S. Y., Kratsios, P., Hart, M., Mourao, R. & Hobert, O. Diversification of *C. elegans* motor neuron identity via selective effector gene repression. *Neuron* **93**, 80–98 (2017).

88. Esmaeili, B., Ross, J. M., Neades, C., Miller, D. M. 3rd & Ahringer, J. The *C. elegans* even-skipped homologue, vab-7, specifies DB motoneurone identity and axon trajectory. *Development* **129**, 853–862 (2002).

89. Petersen, S. C. et al. A transcriptional program promotes remodeling of GABAergic synapses in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 15362–15375 (2011).

90. Hsieh, Y. W., Alqadah, A. & Chuang, C. F. Asymmetric neural development in the *Caenorhabditis elegans* olfactory system. *Genesis* **52**, 544–554 (2014).

91. Hobert, O. Development of left/right asymmetry in the *Caenorhabditis elegans* nervous system: from zygote to postmitotic neuron. *Genesis* **52**, 528–543 (2014).

92. Feng, W. et al. A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. Preprint at *bioRxiv* https://doi.org/10.1101/643320 (2019).

93. Kratsios, P. et al. An intersectional gene regulatory strategy defines subclass diversity of *C. elegans* motor neurons. *eLife* **6**, e25751 (2017).

94. Dasen, J. S. & Jessell, T. M. Hox networks and the origins of motor neuron diversity. *Curr. Top. Dev. Biol.* **88**, 169–200 (2009).

95. Jung, H. et al. The ancient origins of neural substrates for land walking. *Cell* **172**, 667–682.e15 (2018).

96. Zheng, C., Diaz-Cuadros, M. & Chalfie, M. Hox genes promote neuronal subtype diversification through posterior induction in *Caenorhabditis elegans*. *Neuron* **88**, 514–527 (2015).

97. Flames, N. & Hobert, O. Gene regulatory logic of dopamine neuron differentiation. *Nature* **458**, 885–889 (2009).

98. Vidal, B. et al. *C. elegans* SoxB genes are dispensable for embryonic neurogenesis but required for terminal differentiation of specific neuron types. *Development* **142**, 2464–2477 (2015).

99. Berghoff, E. et al. The Prop1-like homeobox gene unc-42 specifies the identity of synaptically connected neurons. *eLife* **10**, e64903 (2021).

100. Pereira, L. et al. A cellular and regulatory map of the cholinergic nervous system of *C. elegans*. *eLife* **4**, e12432 (2015).

101. Dobzhansky, T. Biology, molecular and organismal. *Am. Zool.* **4**, 443–452 (1964).

102. Arendt, D. The evolutionary assembly of neuronal machinery. *Curr. Biol.* **30**, R603–R616 (2020).

103. Achim, K. & Arendt, D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* **27**, 102–108 (2014).

104. Arendt, D. Elementary nervous systems. *Phil. Trans. R. Soc. B* **376**, 20200347 (2021).

105. Holland, P. W. Did homeobox gene duplications contribute to the Cambrian explosion? *Zool. Lett.* **1**, 1 (2015).

106. Akam, M. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *Int. J. Dev. Biol.* **42**, 445–451 (1998).

107. Arlotta, P. & Hobert, O. Homeotic transformations of neuronal cell identities. *Trends Neurosci.* **38**, 751–762 (2015).

108. Bateson, W. *Materials for the Study of Variation, Treated with Especial Regard to Discontinuity in the Origin of Species* (Macmillan, 1894).

109. Sattler, R. Homeosis in plants. *Am. J. Bot.* **75**, 1606–1617 (1988).

110. Wagner, G. P. *Homology, Genes and Evolutionary Innovation* (Princeton Univ. Press, 2014).

111. Tosches, M. A. Developmental and genetic mechanisms of neural circuit evolution. *Dev. Biol.* **431**, 16–25 (2017).

112. Velten, J. et al. The molecular logic of synaptic wiring at the single cell level. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.30.402057 (2021).

113. Dauger, S. et al. Phox2b controls the development of peripheral chemoreceptors and afferent visceral pathways. *Development* **130**, 6635–6642 (2003).

114. Shin, M. M., Catela, C. & Dasen, J. Intrinsic control of neuronal diversity and synaptic specificity in a proprioceptive circuit. *eLife* **9**, e56374 (2020).

115. Ha, N. T. & Dougherty, K. J. Spinal Shox2 interneuron interconnectivity related to function and development. *eLife* **7**, e42519 (2018).

116. Ruiz-Reig, N. et al. Developmental requirement of homeoprotein Otx2 for specific habenulo-interpeduncular subcircuits. *J. Neurosci.* **39**, 1005–1019 (2019).

117. Parker, H. J. & Krumlauf, R. A Hox gene regulatory network for hindbrain segmentation. *Curr. Top. Dev. Biol.* **139**, 169–203 (2020).

118. Zheng, C., Jin, F. Q. & Chalfie, M. Hox proteins act as transcriptional guarantors to ensure terminal differentiation. *Cell Rep.* **13**, 1343–1352 (2015).

119. Urbach, R. & Technau, G. M. Dorsoventral patterning of the brain: a comparative approach. *Adv. Exp. Med. Biol.* **628**, 42–56 (2008).

120. Briscoe, J., Pierani, A., Jessell, T. M. & Ericson, J. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**, 435–445 (2000).

121. Zhang, Q. & Eisenstat, D. D. Roles of homeobox genes in retinal ganglion cell differentiation and axonal guidance. *Adv. Exp. Med. Biol.* **723**, 685–691 (2012).

122. Lichtneckert, R. & Reichert, H. Anteroposterior regionalization of the brain: genetic and comparative aspects. *Adv. Exp. Med. Biol.* **628**, 32–41 (2008).

123. Serrano-Saiz, E., Leyva-Diaz, E., De La Cruz, E. & Hobert, O. BRN3-type POU homeobox genes maintain the identity of mature postmitotic neurons in nematodes and mice. *Curr. Biol.* **28**, 2813–2823.e2 (2018).