PNAS RESEARCH ARTICLE | DEVELOPMENTAL BIOLOGY

Transcriptome age of individual cell types in *Caenorhabditis elegans*

Fuqiang Ma[a] and Chaogu Zheng[a,1]

Edited by Harmit Malik, Fred Hutchinson Cancer Research Center, Seattle, WA; received September 26, 2022; accepted January 27, 2023

The phylotranscriptomic analysis of development in several species revealed the expression of older and more conserved genes in midembryonic stages and younger and more divergent genes in early and late embryonic stages, which supported the hourglass mode of development. However, previous work only studied the transcriptome age of whole embryos or embryonic sublineages, leaving the cellular basis of the hourglass pattern and the variation of transcriptome ages among cell types unexplored. By analyzing both bulk and single-cell transcriptomic data, we studied the transcriptome age of the nematode *Caenorhabditis elegans* throughout development. Using the bulk RNA-seq data, we identified the morphogenesis phase in midembryonic development as the phylotypic stage with the oldest transcriptome and confirmed the results using whole-embryo transcriptome assembled from single-cell RNA-seq data. The variation in transcriptome ages among individual cell types remained small in early and midembryonic development and grew bigger in late embryonic and larval stages as cells and tissues differentiate. Lineages that give rise to certain tissues (e.g., hypodermis and some neurons) but not all recapitulated the hourglass pattern across development at the single-cell transcriptome level. Further analysis of the variation in transcriptome ages among the 128 neuron types in *C. elegans* nervous system found that a group of chemosensory neurons and their downstream interneurons expressed very young transcriptomes and may contribute to adaptation in recent evolution. Finally, the variation in transcriptome age among the neuron types, as well as the age of their cell fate regulators, led us to hypothesize the evolutionary history of some neuron types.

transcriptome age index | developmental hourglass pattern | evolutionary conservation | *Caenorhabditis elegans*

Significance

Previous studies of the whole embryo gene expression profiles found that older genes tend to be expressed in midembryonic stages, while younger genes are expressed in early and late embryos, suggesting that the midembryonic stage is the most conserved. We analyzed the transcriptome age of individual cell types using single-cell transcriptomic data in the nematode *Caenorhabditis elegans*. Our results confirmed the "hourglass" shape of the transcriptome age profile across development at the single-cell level and identified the cells and tissues that contribute to these changes. Moreover, by comparing the transcriptome ages of individual neuron types, we identified a potential coevolution of chemosensory neurons and their downstream interneurons and generated some hypotheses for the evolutionary relationship between some neuron types.

Measuring the age of genes expressed at various developmental stages provides an opportunity to understand development from an evolutionary perspective. Transcriptome age index (TAI) is such a measurement that combines phylostratigraphy with developmental stage-specific expression profiling to calculate the evolutionary age of the transcriptome (1). Higher TAI indicates younger transcriptomes. TAI calculated for different ontogenetic stages of several organisms, including zebrafish, *Drosophila*, *Arabidopsis thaliana*, and fungi, found that the oldest transcriptome was expressed at the phylotypic stage, whereas earlier and later stages of embryos expressed younger transcriptomes (1–3). These findings confirmed, at the molecular level, the hourglass model of embryonic evolution, which proposes that the morphological divergence in embryonic development is constrained at the midembryonic stages (4). Whether this pattern holds true for other organisms remains to be determined. Interestingly, in Spiralia, the midembryonic stage was found to be more divergent than early and late embryonic development, showing a reverse hourglass pattern (5).

Moreover, TAI has mostly been applied to the transcriptome of entire organisms or tissues/organs. Given the prevalence of single-cell transcriptomics, we reason that TAI could also be used to measure the transcriptome age of individual cell types, which may help understand the cellular basis of the hourglass pattern. For example, a detailed view of the cellular changes in transcriptome age during embryogenesis could help identify the tissues or lineages that are responsible for the expression of old and young transcriptomes at the conserved and divergent phases of development, respectively. One possibility is that the classical hourglass model of development can be captured by the single-cell transcriptomes in all lineages, and the other possibility is that tissue and cell types have significant variations in their transcriptome ages and contribute differently to the overall TAI profile. If the latter is true, what tissues drive the rise and fall of overall TAI in the embryos?

Single-cell transcriptome age can also be used to understand the evolutionary relationship between cell types under the assumption that cells that express younger transcriptomes may have evolved later than cells that express older transcriptomes. Together with the analysis of

Author affiliations: [a]School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China

Author contributions: F.M. and C.Z. designed research; F.M. performed research; F.M. and C.Z. analyzed data; and F.M. and C.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: cgzheng@hku.hk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2216351120/-/DCSupplemental.

Published February 22, 2023.

gene ages of the selectors that control cell fate specification, TAI of individual cell types may help deduce an ancestral state of the diversified cell fates and help understand how new cell types emerge upon genomic innovation.

With the above motivation, we studied the transcriptome age of the nematode *Caenorhabditis elegans* throughout embryonic and larval development by analyzing both bulk and single-cell (sc) RNA-seq data. Previously, Levin et al. compared the transcriptomes of five *Caenorhabditis* species during embryonic development and identified ventral enclosure as a developmental milestone that may be molecularly conserved in nematodes (6). Further pioneering work by Hashimshony et al. analyzed the transcriptomes of founder blastomeres and their descendants using CEL-seq and found that the endoderm expressed higher fraction of conserved genes than ectoderm and mesoderm during early time of germ layer induction, suggesting variations in transcriptome ages among the early embryonic cells (7). Building on these prior studies, we used refined phylostratification and transcriptomic data with high temporal resolution and identified a period of the lowest TAI during *C. elegans* midembryogenesis, which started at ventral closure and lasted to the beginning of elongation and may correspond to the phylotypic stage. Moreover, our analysis of single-cell transcriptomes revealed an older transcriptome age of germline precursors compared to somatic tissues in early embryos and a lower TAI in endodermal lineages compared to other germ layers during early induction. The variation in transcriptome ages among the cell and tissue types grew bigger at late embryonic and larval stages as cells differentiate. Tracking the TAI dynamics along lineages identified certain tissues (e.g., hypodermis) that contribute to the rise of TAI in late embryos. Further analysis of the variation in transcriptome ages among the 128 neuron types in *C. elegans* nervous system found that a specific group of chemosensory neurons and their downstream interneurons expressed very young transcriptomes and may contribute to adaptation in recent evolution. Finally, the variation in TAI among the neuron types, as well as the age of their fate regulators, led us to hypothesize the evolutionary history of these neuron types. In summary, using *C. elegans* as an example, we showcased how the transcriptome age at the single-cell level could provide insight into the cellular basis of developmental innovation and help understand the functional diversity and evolutionary origin of cell types.

Results

Transcriptional Age Index Identifies the Phylotypic Stage of Development in *C. elegans*. Since the *C. elegans* genome has expanded significantly in the nematode lineage (8), to sufficiently separate genes with different ages and generate TAI with high resolutions, we included 10 nematode species that have well-annotated genomes together with *Drosophila* and mouse in the phylogenetic analysis. We assigned the 19,997 *C. elegans* protein-coding genes to 12 phylostrata (PS) based on their positions in the phylogeny (Fig. 1*A*). About one-third (6,913) of the genes are from PS1-2, which are the oldest phylostrata and contain genes conserved in bilateria; another third (5,384) are nematode-specific genes from PS3-6; and the last third (6,511) contained genes that are originated in the *Caenorhabditis* genus (PS7-9) or more specifically in *C. elegans* and its closely related sister species (PS10-11) (Fig. 1*B*). Different phylostrata showed the enrichment of genes encoding different proteins. For example, PS1-2 were enriched with protein kinase genes, while the youngest PS10-12 were enriched with genes encoding G protein-coupled receptors (GPCRs), F-box proteins (FBPs), nuclear hormone receptors (NHRs), etc. (Fig. 1C and *SI Appendix,* Fig. S1*A*), which is consistent with previous findings on the recent expansion of these gene families in *C. elegans* (9, 10).

With the gene ages, we calculated the TAI for different developmental stages of the embryos, larvae, and young adults using available bulk transcriptomic data; the embryonic data is from a single study with high temporal resolution (7), while the larval and adult data are aggregate expression across multiple studies (*Materials and Methods*). Strikingly, we observed the classical hourglass pattern of embryonic development in *C. elegans*, indicated by the high TAI at early and late embryogenesis and the low TAI in the midembryonic stages when the oldest genes are expressed. This period (about 280 to 380 min post-first cleavage; mpfc) of the lowest TAI roughly corresponds to the morphogenesis phase between the end of gastrulation and the start of elongation (11) and can be defined as the phylotypic stage of *C. elegans* at the molecular level. Interestingly, previous studies also identified the ventral enclosure milestone (at 290 mpfc) as a potential nematode phylotypic stage (6). Our analysis extended this earlier observation and identified morphogenesis (equivalent to early organogenesis in vertebrates) as the most genetically conserved stage of *C. elegans* embryogenesis. During this stage, cells began to adopt their terminal cell fate and integrate into tissues.

Moreover, we noted that the 2- to 8-cell embryos (0 to 60 mpfc) had relatively low TAI likely because maternally deposited messenger RNAs (mRNAs) are predominantly expressed from the PS1 genes (*SI Appendix,* Fig. S2). A similar pattern was observed in *Drosophila* (1). We also observed a decrease in TAI in early larval stages (L1 to L3) following hatching and then a peak of TAI at the fourth larval stage (L4) before adulthood (Fig. 1*D*). Interestingly, the expression of old and conserved genes (hence low TAI) during early larval development was also observed in zebrafish and *Drosophila* (1). The TAI of male adults is much larger than the TAI of hermaphrodites, which is then larger than females (made with *fog-2* mutants) (Fig. 1*D*), consistent with the male–female difference in *Drosophila*. Moreover, dauers appeared to express much younger transcriptomes than regular L3 animals, indicating that young genes might contribute to the formation of this alternative life stage (Fig. 1*D*).

Next, we analyzed the contribution of genes from different phylostrata to the general TAI profile and their relative expression across developmental stages. In general, the youngest genes specific to *C. elegans* and sister species (PS10-12) had peak expression in early embryos, *Caenorhabditis*-specific genes (PS7-9) had peak expression in late embryos, and both have a strong expression at L4 stage. The bilaterian genes (PS1-2) and nematode-specific genes (PS3-6) appeared to have broad expression in the embryos (*SI Appendix*, Fig. S1*B*). Based on the expression dynamics of genes in individual phylostrata, we found that PS9-12 (especially PS12) genes contributed strongly to the first TAI peak (at 120 mpfc, which corresponds to the end of proliferation and the start of gastrulation), PS7-11 genes to the second peak (at 770 mpfc, which corresponds to the quickening phase before hatching), and PS7-12 genes to the third peak (at L4 stage) (Fig. 1 *E* and *F*).

Further gene ontology analysis found that the expression of PS9-12 genes encoding F-box proteins (FBP), BTB/POZ domain-containing proteins (BTB), and DUF3557 domain proteins drives the first TAI peak in early embryos (Fig. 1*G* and *SI Appendix*, Figs. S3*A* and S4). Since FBP and BTB are both adaptors for protein ubiquitination and degradation, we suspect that the regulation of protein stability in early embryogenesis may have evolved recently in *C. elegans*. Expression of PS7-11 genes encoding membrane-associated proteins, such as G protein-coupled receptors (GPCRs), glycoproteins, and proteins with LEC, Claudin_3, or CUB_2 domains, are enriched at the second peak in late embryos. These young genes

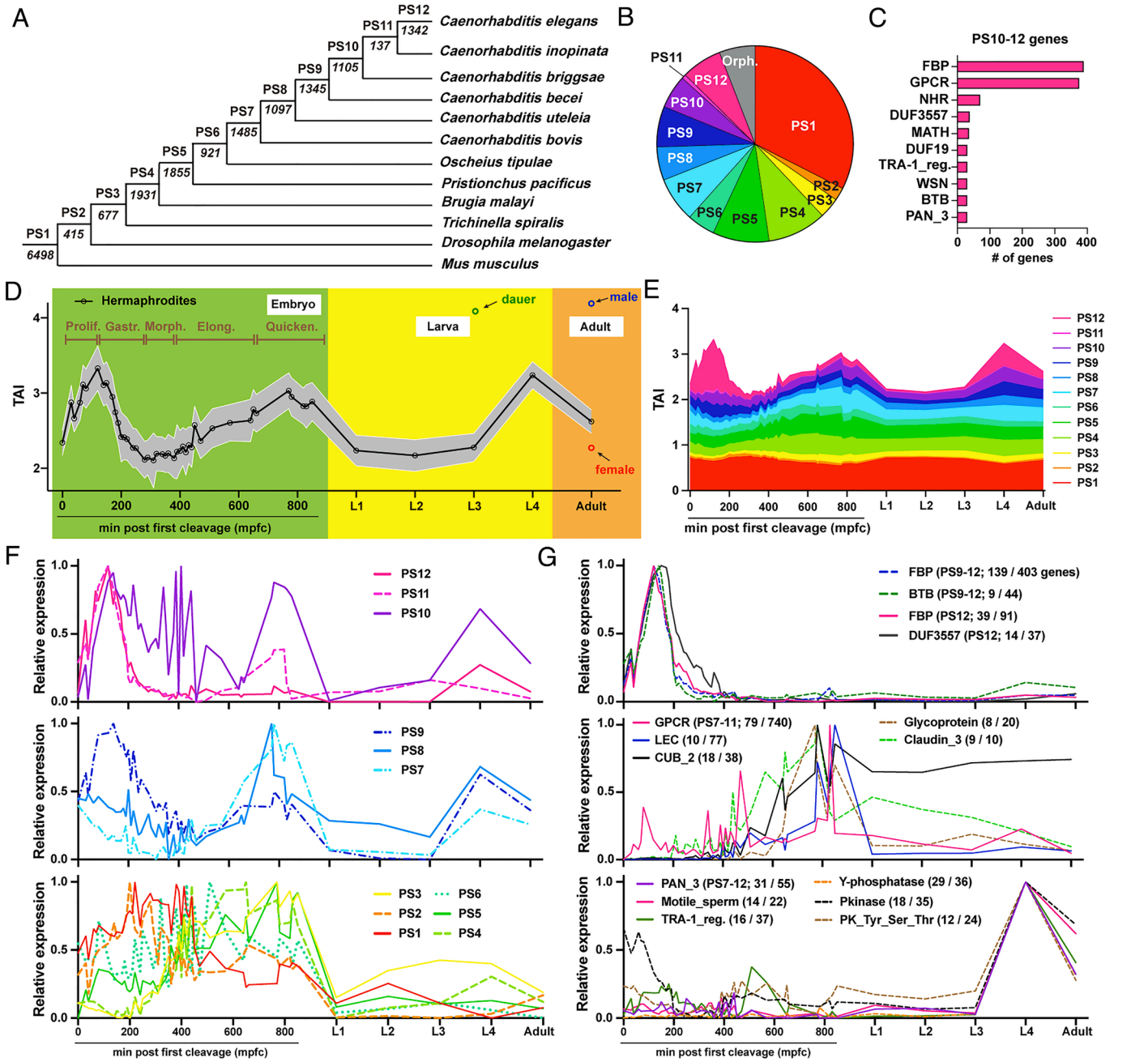

Fig. 1. Phylotranscriptomic analysis revealed an hourglass pattern of embryonic development in *C. elegans*. (*A*) Phylostratification of *C. elegans* genes. (*B*) Comparison for the number of genes in each phylostrata (PS). (*C*) Gene ontology analysis of the youngest genes (PS10-12 genes). (*D*) TAI profile of embryonic and larval development, dauers, and hermaphrodite, male and female adults using bulk RNA-seq data. The shaded areas represent the SD of TAI estimated by permutation analysis. The overall pattern is significant by a flat line test ($P = 0.0012$). (*E*) Contributions of genes originated from different PS to the overall TAI across developmental stages. (*F*) Relative expression of genes from different PS across development (calculated by setting the highest to 1 and lowest to 0 for easy comparison). (*G*) Relative expression of genes originated from PS9-12 (*Top*), PS7-11 (*Middle*), and PS7-12 (*Bottom*) with specific domains. The number of genes that have the strongest expression at the TAI peak was indicated by the number before "/" and the total number of genes in the family after "/".

that function in responding to extracellular cues might help the animal adapt to the external environment after hatching. Finally, the expression of PS7-12 genes that code for diverse proteins involved in intracellular signaling (e.g., protein kinases and phosphatases), metabolism (e.g., acyltransferases and glycoside hydrolase), sexual differentiation (e.g., proteins with Motile_sperm or TRA-1_regulated domains), and other pathways are enriched at the L4 stage before the animal matures into adults (Fig. 1*G* and *SI Appendix*, Figs. S3*B* and S4).

Overall, our results not only confirmed the existence of a phylotypic phase in midembryonic stages but also identified the young genes that potentially drive the evolutionary innovation or adaptation at three distinct developmental stages in *C. elegans*.

Variation of TAI at the Single Cell Level. Next, we attempted to understand the contribution of different tissues to the hourglass pattern of development by calculating the TAI of single-cell transcriptomes. We analyzed four sets of scRNA-seq data at different stages, including the zygote to 16-cell stages (12), embryonic stages (13), L2 stage (14), and L4 stage (15). First, to assess whether the scRNA-seq data could collectively recapitulate the TAI profile of development, we calculated the

whole-embryo transcriptome at different time points and the whole-animal transcriptome at L2 and L4 stages using the single-cell transcriptomes. (We accounted for the number of cells for each cell type; *Materials and Methods*.) Indeed, the TAI profile of the whole-embryo/organism transcriptome reconstituted from the sum of single-cell transcriptomes matched the profile of the bulk RNA-seq data (Fig. 2*A*), confirming that the scRNA-seq data are valid.

We then calculated the TAI of individual cells at each developmental stage using the single-cell transcriptomes. We found that the variation in TAI among the cell types appeared to be small and consistent in early and middle stages of embryogenesis but became bigger in late embryos when more cells became terminally differentiated (Fig. 2*B*). Cells at L4 stage also showed larger variation in their TAI than at L2 stage. When examining the TAI differences among early embryonic lineages (AB, MS, E, C, D, and P4), we found that the germline precursor P3 at the 8-cell stage and the P3 daughter cells (D and P4) at the 16-cell stage had lower TAI than the other blastomeres (Fig. 2*C*).

Another interesting observation is that during early gastrulation that gives rise to the three germ layers (120 mpfc), the endodermal (E) lineage appeared to have the lowest TAI among the somatic lineages (Fig. 2*C*). This result supported a previous study by Hashimshony et al. (7), who analyzed the transcriptome of in vitro descendants of blastomeres and found that the endoderm expressed a higher proportion of old genes that are conserved in opisthokont eukaryotes than the mesoderm and ectoderm at early time points of germ layer induction. Nevertheless, that study was based on the origin of a small number of genes. We reanalyzed the Hashimshony et al. dataset by including all genes in the TAI calculation and confirmed that the endoderm indeed expressed the oldest transcriptome during early induction (*SI Appendix*, Fig. S5). Thus, our analysis at the single-cell level supported the notion that endoderm may be the oldest among the three germ layers.

Interestingly, the TAI difference among the germ layers largely disappeared towards the end of gastrulation (200 to 280 mpfc). At around the comma stage (380 mpfc), the MS lineage descendants [most of which differentiate into pharyngeal and body wall muscles] and the P4 descendants (which become the germline) showed significantly lower TAI than the other lineages. The TAI variation within the same lineage (e.g., AB) also became bigger as more cells are generated from the lineage (Fig. 2*C*), and this variation persisted within sublineages (*SI Appendix*, Fig. S6*A*).

In late embryonic and larval stages when terminal cell types can be identified from the single-cell transcriptomes, we observed a significant difference in transcriptome ages among the tissue types. For example, hypodermis showed much higher TAI than neurons, body wall muscles (BWM), and intestine in late embryos, although the difference became smaller in larvae (Fig. 2*D*). The BWM cells had the lowest TAI among the four tissue types throughout development. Interestingly, although the TAI of the same neuron is highly correlated between L2 and L4 stages (*SI Appendix*, Fig. S8*B*), we observed marked increase in TAI variations among the individual neuron types from L2 to L4 (Fig. 2*D*), which indicate further transcriptomic divergence in the nervous system as the animals mature. Notably, among other tissues, the arcade cells (which are interfacial epithelial cells derived from the same lineages as hyp1-3) and the pharyngeal gland cells had very young transcriptome ages, suggesting possible recent evolution. Sperms also had a high TAI, whereas the germline that developed into oocytes had a much lower TAI (Fig. 2*D*), which is consistent with the male-biased sex difference in TAI (Fig. 1*D*) and the old age of maternal genes (*SI Appendix*, Fig. S2).

Cellular Basis of the Hourglass Pattern. To understand how the transcriptome age changes along development at the cellular level, we analyzed the single-cell transcriptomes of lineage precursors that lead to the terminally differentiated cells, as well as the terminal cells at different time points. In all 170 lineages we tracked, we were able to observe a peak around 60 to 100 mpfc and a peak at L4 stage, which correspond to the first and third peaks in the whole-organism TAI profile, respectively. However, the second peak in the late embryo was only recapitulated in some lineages (Fig. 2 *E*–*H*). For example, hypodermal cells, but not BWM and intestine cells, showed a peak at 600 to 700 mpfc. In fact, the hypodermal TAI values (~5), which is partly fueled by the strong expression of collagen genes from PS7 (*SI Appendix*, Fig. S6*E*), are higher than the whole-embryo average (~3) at the second peak, suggesting that the hypodermis may help drive the rise of TAI in late embryos. Similarly, seam cells, arcade cells, and pharyngeal and rectal gland cells also contribute to the high TAI in late embryogenesis (*SI Appendix*, Fig. S6 *B*–*D*). In contrast, pharyngeal muscles, enteric muscles, and glia cells appeared to express older genes and had below-average TAI values in the same period.

Moreover, among the 71 neurons or neuron types whose lineages we tracked, 12 (e.g., RIS and ADL) showed the typical TAI profile with three peaks and two valleys, 15 (e.g., AVH and CAN) had a similar profile except that the second peak appeared at an earlier time (375 to 550 mpfc), and the others (e.g., AIB and AVE) had no peak in late embryos, although many of them had no transcriptomic data at 700 mpfc (Fig. 2 *F* and *H*). In some neurons, the rise of TAI in late embryos is caused by the expression of young genes associated with their functions. For example, the expression of many GPCR genes that originated from PS12 appeared to cause the high TAI in the chemosensory ADL and ASJ neurons (*SI Appendix*, Fig. S6*F*). The variation of the TAI profiles among the neurons are somewhat expected because *C. elegans* neurons arise nonclonally from many different lineages (16).

In summary, our analysis provided a granular view of the hourglass pattern of development and found that the "W-shaped" TAI profile with younger genes expressed at the early and late embryonic and L4 stages and older genes expressed at the midembryonic and early larval stages can be faithfully reproduced at the single-cell level for some but not all the cell lineages. Although all lineages share the same pattern of expressing young genes in early embryos and then switching to express more conserved batteries of genes, some tissues (such as muscles and intestines) kept the old transcriptomes throughout the rest of embryonic development, whereas other tissues (such as hypodermis, seam cells, and some neurons) started to express another set of young genes that may be related to their differentiation and functions. We suspect the former tissues are more functionally conserved across species, while the latter ones are more subjected to evolutionary innovation.

Robustness of the TAI Measurements for scRNA-seq Data. Given the difference between scRNA-seq and the traditional bulk RNA-seq methods, we assessed the robustness of the TAI score against the variation of single-cell transcriptomes. We found that the TAI scores calculated using the mean expression across cells of the same cell type were almost identical to the mean TAI of single cells and were highly similar to the TAI calculated using the mean of pseudo-bulk replicates (*SI Appendix, SI Results* and Fig. S7). The SD of TAI estimated by the permutation of gene age was within the same range as the variation of the TAI values of individual cells. We also found that different normalization methods had little effect on the TAI score and the TAI of the same cell types in different datasets were highly correlated (*SI Appendix, SI Results* and Figs. S7 and S8), suggesting the robustness of the measurement.

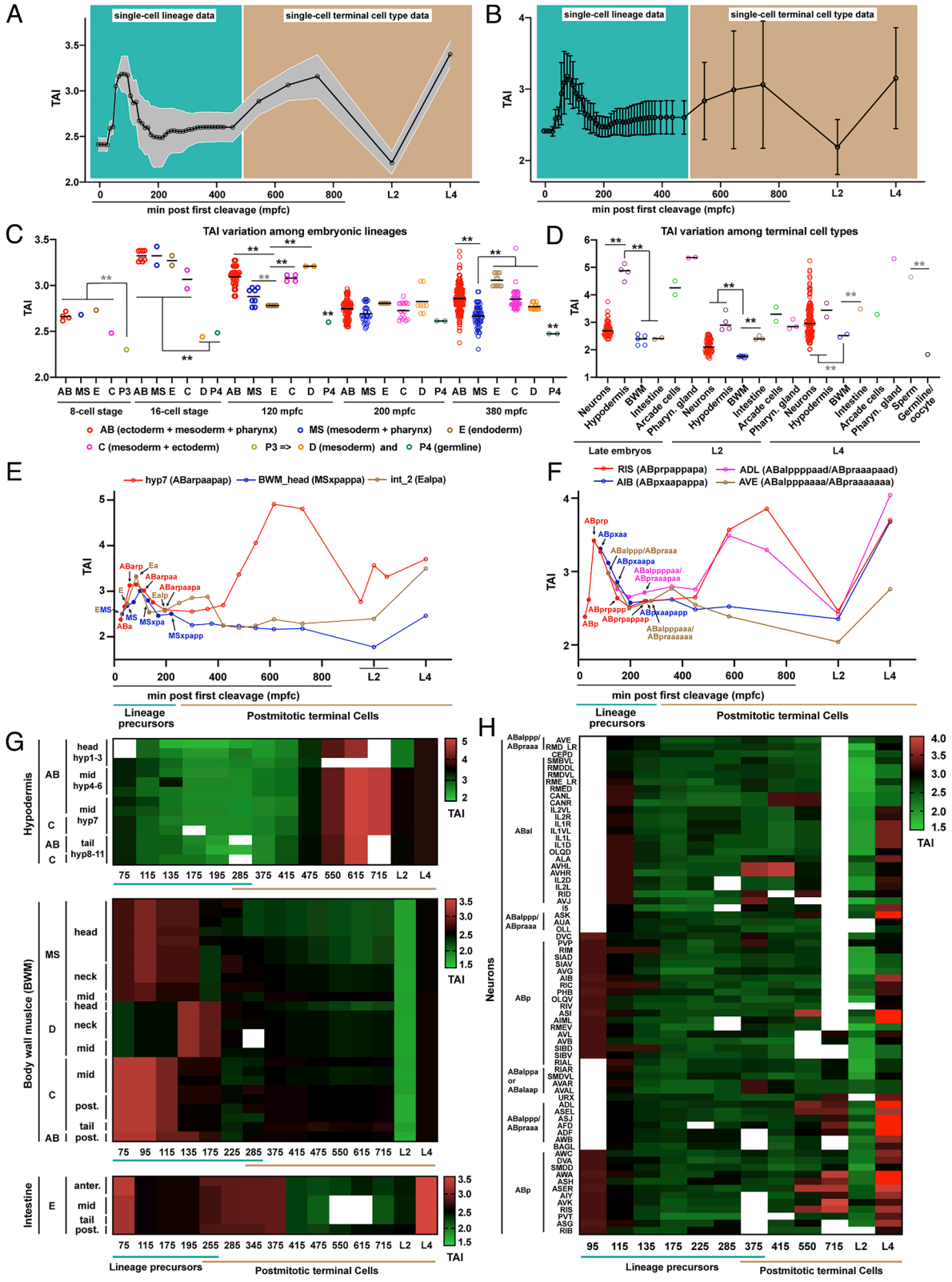

Fig. 2. Variation of transcriptome ages among cell types and the hourglass pattern at single-cell level. (*A*) The TAI profile of whole embryo/animal transcriptome assembled in silico from single-cell transcriptomic data (P = 0.016 in a flat line test). (*B*) The mean ± SD TAI of transcriptomically unique cell types across developmental stages. (*C*) The TAI of the descendants of specific lineages at 16-cell stage, 200 mpfc, and 380 mpfc in the embryos (note that the last round of cell division for most embryonic lineages occurred before 400 mpfc). (*D*) The TAI of annotated terminal cell types at late embryonic (645 mpfc), L2 and L4 stages. The results of two statistical analyses were shown. Grey asterisks indicate significant difference ($P < 0.01$) in an unpaired t test using the SD estimated by permutation. Black asterisks indicate significant difference ($P < 0.01$) in both the unpaired t test and a post-ANOVA Tukey's test. (*E* and *F*) The TAI profile of representative lineages that give rise to terminal cells. Lineage precursors were represented as filled circles, and the postmitotic terminal cells as empty circles. Flat line tests for hyp7 and ADL neuron showed statistical significance ($P < 0.05$). (*G* and *H*) Heatmaps showing the TAI dynamics across development of the lineages that generate hypodermis, BWM, intestine, and neurons. The empty blocks indicate that the data for the time point is not available. Bright red in (*H*) for ASK, ASI, AIML, ADL, ASJ, AFD, ADF, AWA, and ASH at L4 stage means that their TAI scores are above 4 and out of range.

As expected from the TAI formula, we found that highly and reliably expressed genes contributed much more to the TAI score than weakly or unreliably expressed genes, and the TAI difference among cell types is mainly driven by highly variable genes instead of the ubiquitously or broadly expressed genes (*SI Appendix, SI Results* and Figs. S9–S11). Lastly, the TAI score is also robust against the potential ambiguity in the phylostratification of genes (*SI Appendix, SI Results* and Fig. S12). Despite the above points, we caution that TAI may have an inherent bias toward highly expressed young genes.

Specific Chemosensory Neurons and their Downstream Interneurons Have Younger Transcriptomes than Other Neurons. We next focused on understanding the dramatic variation in TAI among the neurons using the L4 scRNA-seq data (15), which described the transcriptomes of 128 individual neuron types in their terminally differentiated state. Based on the TAI values (Fig. 3*A*), the oldest transcriptomes came from the mechanosensory touch receptor neurons (TRNs) ALM (TAI = 2.01) and PLM (2.02), while the youngest transcriptomes came from the chemosensory neurons ASJ (5.23), ASI (4.90), and ASK (4.85). Although neurons that share similar cell fate generally have similar TAI [e.g., IL2_LR (3.15) and IL2_DV (3.21)], some exceptions were observed [e.g., RMD_DV (2.27) and RMD_LR (2.99)]. In another example, AVM (2.61) and PVM (3.02) are also touch receptor neurons but have much younger transcriptomes than ALM and PLM.

We compared the TAI of individual neuron types based on neuronal function, anatomical location, birth time, lineage history, neurotransmitter identity, and connectivity. Sensory neurons and interneurons express significantly younger transcriptomes than motor neurons (Fig. 3*B*), suggesting stronger adaptive evolution of the sensory neurons and the interneurons (which acquire and integrate sensory information, respectively), compared to the motor neurons (which execute motor command and control muscle contraction). Being consistent with this finding, neurons located in the head express younger transcriptomes than neurons located in the body (Fig. 3*C*), since most sensory neurons and interneurons are in the head and most motor neurons are located along the ventral nerve cord. Similarly, neurons born in the embryos have higher TAI than postembryonically born neurons, because 59 of the 80 postembryonic neurons are motor neurons (Fig. 3*D*); nevertheless, the birth time does not correlate directly with TAI (*SI Appendix*, Fig. S13 *A* and *B*). Distance in cell lineage does not correlate with the difference in TAI either (*SI Appendix*, Fig. S13*C*), suggesting that lineage history is not related to the transcriptome age. Moreover, neurons that release different neurotransmitters do not show significant differences in their transcriptome ages (Fig. 3*E*), which echoes the parallel evolution of a diverse array of neurotransmitters (18).

We then asked whether the position of the neurons in the connectome explains the TAI variation. Although the overall connectedness of individual neurons is not correlated with TAI (Fig. 3*F* and *SI Appendix*, Fig. S13 *D–F*), we found that category 6 sensory neurons (SN6), based on the connectome description by Cook et al. (17), expressed much younger transcriptomes than all other five categories of sensory neurons (SN1-5; Fig. 3*H*). Different groups of sensory neurons respond to different types of environmental stimuli with SN6 mostly detecting chemical and thermal cues (Fig. 3*G*). Interestingly, SN2 (PHA, PHB, and PHC neurons) are also chemosensory neurons but have much lower TAI than SN6. Since SN2 respond to noxious and harmful chemicals and drive protective escape behaviors (19, 20), whereas SN6 respond to a wide range of environmental chemicals and odors (both attractants and repellents) (21), we suspect that SN2 is involved in more conserved nociception (22), while SN6 is subjected to evolutionary changes that allow the animals to better sense and adapt to the environment.

Downstream of the sensory neurons, interneurons fall into four categories (IN1-4) according to Cook et al. (17). IN1-3 form three interneuron layers in a feedforward loop with IN3 targeting both IN2 and IN1, and IN2 targeting IN1. The category 4 interneurons (IN4) interact across all layers of interneurons, as well as many sensory neurons. Interestingly, among the interneurons, IN4 expresses younger transcriptomes than IN1, suggesting that IN4 might have evolved after the establishment of the three layers (Fig. 3*H*). We also noticed a large variation in TAI among the IN3 neurons. Given that SN6 neurons preferentially target IN3, we identified the IN3 interneurons (AIA, AIY, AIZ, and PVQ) that are highly connected to the SN6 and found that their transcriptome ages are much younger than the IN3 neurons that have low connectivity with SN6 (Fig. 3 *I* and *J*). The same holds true when we broadened the SN6 downstream interneurons to include IN1 and IN2 targets (RIA and AIB, respectively) that have high connectivity with SN6 (Fig. 3*K*). Importantly, high connectedness between IN1-3 and sensory neurons per se does not predict high TAI in interneurons (Fig. 3*L*), but the connectedness with SN6 does. Moreover, the transcriptome age differences between groups of neurons were supported not only by the TAI of mean expression (Fig. 3) but also by the TAI of single cells and the TAI calculated from the average expression of pseudobulk replicates (*SI Appendix*, Figs. S14 and S15).

The young transcriptome age of both chemosensory neurons and their downstream interneurons indicate that the chemosensory circuit may have coevolved to adapt to environmental conditions. We found that a large number of GPCR and NHR genes that emerged from PS7-12 were expressed specifically or at a much higher level in SN6 neurons compared to the other sensory neurons (*SI Appendix*, Fig. S16 *A* and *B*). These young GPCR and NHR genes are involved in sensory perception and their positive selections may have driven the adaptation of SN6 chemosensory neurons (23, 24). The adaptation of SN6 downstream interneurons is more puzzling. We found that among the 147 PS7-12 genes that had significant expression (TPM > 100) in IN3 interneurons with high SN6 connectivity, 29 genes encode neuropeptides (13-fold enrichment) and majority of them had higher expression in IN3 neurons highly connected to SN6 than the lowly connected ones (*SI Appendix*, Fig. S16*C*). Interestingly, we also found that a different set of PS7-12 neuropeptides were more strongly expressed in SN6 compared to SN1-5 neurons (*SI Appendix*, Fig. S16*D*). Since neuropeptides can modulate synaptic transmission and regulate neuronal activity (25), the expression of newly emerged neuropeptide genes in SN6 and downstream interneurons may be involved in the adaptation of sensory response.

The Age of Cell Fate Regulators May Explain the Evolutionary History of Neuron Types. A compelling hypothesis in the field of cell type evolution is that the extraordinary diversity of cell types is generated by stepwise diversification of a handful of existing cell types by the addition of cell fate regulators that arise through genomic evolution (26, 27). However, very few comparative studies could capture the emergence of new cell types and pinpoint its genetic basis. We reason that the transcriptome age of neuron types and the gene age of their cell fate regulators may provide insights into the evolutionary relationship between closely related cell types.

Focusing on the *C. elegans* hermaphrodite nervous system, we first compiled a list of 60 neuronal fate regulators that control the fate specification of 93 (out of 128) transcriptionally distinct neuron types (*SI Appendix*, Table S1). 52 (87%) of the 60 fate

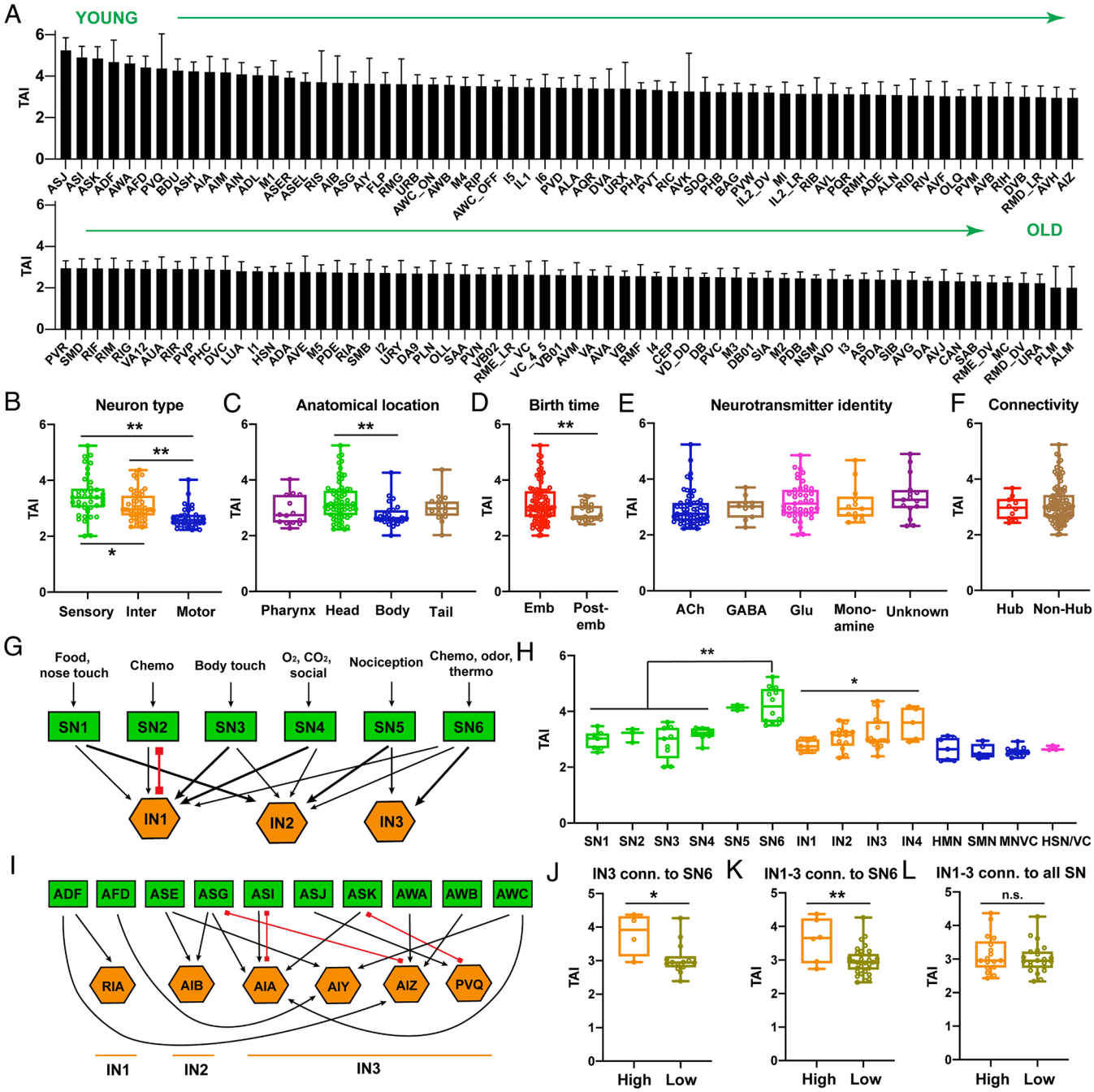

Fig. 3. TAI variation among the 128 neuron types in *C. elegans* nervous system. (*A*) TAI of annotated neuron types from the scRNA-seq data at the L4 stage. SD was estimated by a permutation analysis. (*B–F*) Comparison of neurons based on their general type, location, birth time, neurotransmitter identity, and connectedness. Single and double asterisks indicate statistical significance ($P < 0.05$ or 0.01, respectively) in a post-ANOVA Tukey's multiple comparison or in an unpaired *t* test for two samples. (*G*) A schematic representation of the major connectivity between SN1-6 and IN1-3 according to Cook et al. (17). Black arrows indicate chemical synapses (larger arrows means more synapses), and red lines with squared heads indicate gap junctions. (*H*) Comparison of the TAI of neurons belonging to different groups in the connectome. (*I*) Connection of SN6 neurons with the IN1-3 neurons. (*J* and *K*) Comparison of the TAI of IN3 or IN1-3 interneurons based on their connectivity to SN6 neurons. High means the interneuron has in total over 50 serial electron microscopic (EM) sections showing connectivity with SN6 neurons according to Cook et al. (17). (*L*) The TAI of IN1-3 interneurons based on their connectivity to all sensory neurons (high means over 100 sections of connectivity).

regulators were from the oldest phylostratum (PS1), which supported the idea that the genetic mechanism that regulates neuronal differentiation is highly conserved throughout evolution. These fate regulators and more broadly the homeodomain transcription factors (TFs), which are over-represented among the TFs that regulate neuronal fate specification (28), are considerably older than other TFs (Fig. 4*A*). Interestingly, eight fate regulators originated from more recent phylostrata ranging from PS2 to PS6 (Fig. 4*B*), which provided a window to understand the evolutionary history of a few neuron types controlled by these regulators.

For example, the differentiation of two sister cells, the mechanosensory neuron ALM and the peptidergic interneuron BDU, is controlled by a regulatory switch between MEC-3 (a LIM homeodomain TF) and PAG-3 (a zinc finger TF) (30). In ALM, UNC-86 (a POU homeodomain TF) preferably binds to MEC-3 and promotes the touch receptor neuron (TRN) fate, whereas in BDU, which is the posterior sister cell, Wnt signals and PAG-3 act together

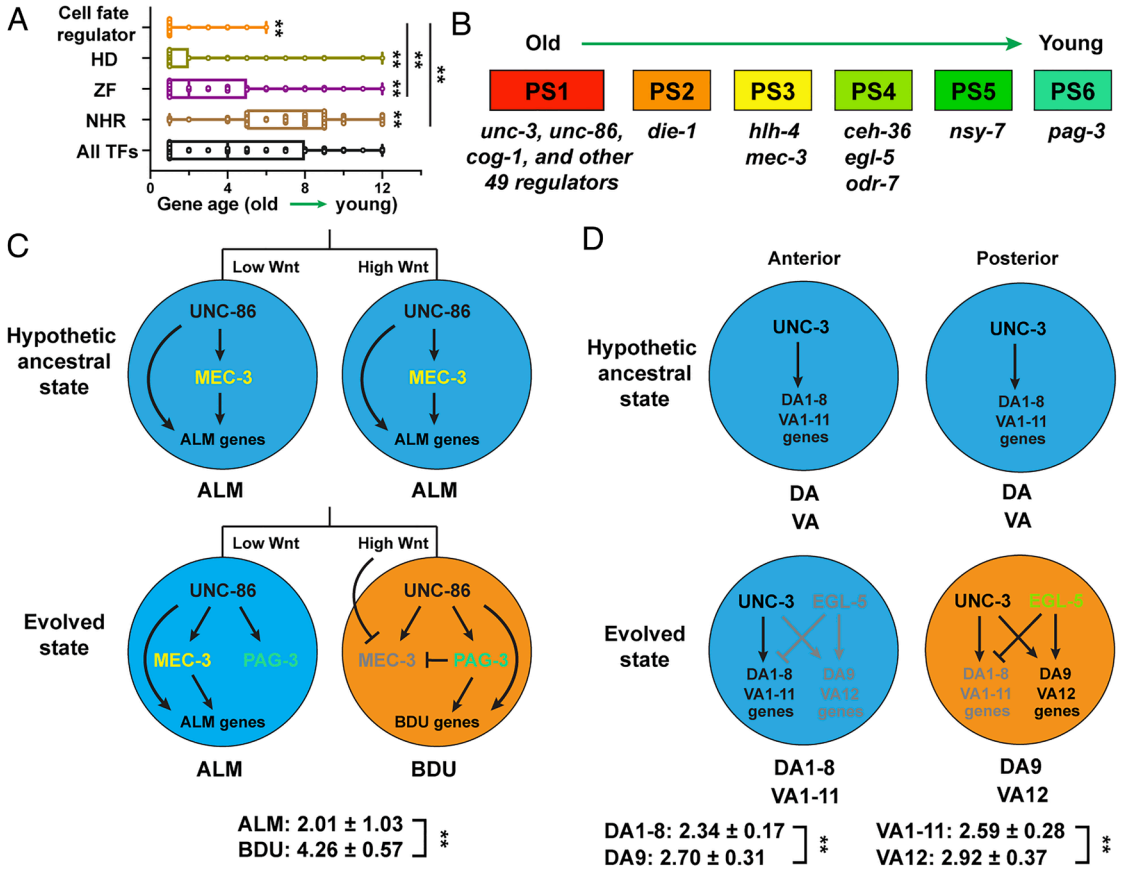

Fig. 4. The gene age of cell fate regulators helps understand the evolutionary relationship between neuron types. (*A*) The comparison of the gene age for transcription factors (TF) based on the phylostrata of their origin. TF classification is based on the Vermeirssen et al. study (29). (*B*) The gene ages of the cell fate regulators (*SI Appendix*, Table S1). (*C*) In a hypothetic ancestral state before the emergence of *pag-3*, both sister cells adopt the ALM fate and express the ALM genes. In the evolved state, PAG-3 inhibits the expression of *mec-3* and causes the posterior sister cell to develop into the new BDU fate. (*D*) The emergence of the *egl-5* changes the most posterior neuron of the DA and VA motor neurons and causes them to differentiate away from the general DA and VA fate and become specific subtypes different from the anterior ones. Double asterisks indicate statistically significant differences in an unpaired *t* test using the SD estimated by permutation.

to repress *mec-3* expression, leading to the alternative BDU fate (Fig. 4*C*). Since *pag-3* originated from PS6, whereas *mec-3* is a PS3 gene, we hypothesize that in an ancestral state both sister cells would adopt an "ALM" fate, and the emergence of *pag-3* led to the differentiation of BDU as a new type of neuron (Fig. 4*C*). The difference in TAI (ALM: 2.01 vs. BDU: 4.26) supported this hypothesis, as BDU expressed a much younger transcriptome.

Another example is the diversification of motor neurons along the anterior–posterior axis. The general cholinergic motor neuron fate in DA and VA neurons is induced by the conserved COE domain TF *unc-3* (a PS1 gene), but the most posterior subtypes (DA9 and VA12) undergo further specification induced by the Hox gene *egl-5*, which originated from PS4 (31, 32). Thus, we suspect that the DA9 and VA12 fates might have evolved after the emergence of *egl-5* from an ancestral state where all the DA and VA neurons adopted the present "DA1-8" and "VA1-11" fates, respectively (Fig. 4*D*). This hypothesis is supported by the younger transcriptome in DA9 and VA12 compared to their anterior counterparts.

The age of fate regulators also shed light on the rise of bilateral asymmetry in the nervous system. The left and right chemosensory ASE neurons had symmetric cell shapes but diversified at the molecular and functional level through a bistable switch between DIE-1 (a C2H2 zinc finger TF that controls the ASEL fate) and COG-1 (a homeodomain TF that promotes the ASER fate) (33). Since *die-1* (a PS2 gene) emerged later than *cog-1* (a PS1 gene), we suspect that ASER is the ancestral neuron type before the evolution of left-right asymmetry (*SI Appendix*, Fig. S17*A*). Similarly, because *nsy-7* (encoding a homeodomain-like TF that controls the AWC^{ON} fate) arose later than *die-1* (which controls the AWC^{OFF} fate), the AWC^{OFF} fate may be the ancestral state (*SI Appendix*, Fig. S17*B*). Nevertheless, the TAI values for the left and right subtypes of ASE and AWC were very close.

Finally, intermediate TFs appeared to evolve later than the upstream cell fate regulator. For example, *odr-7* (an NHR-type TF that controls the AWA fate), *ceh-36* (a homeodomain TF that regulates AWC and ASE fates), and *mec-3* (a LIM homeodomain TF that control the TRN fate), which are all derived from PS3 or PS4, were activated by more conserved TFs that are derived from PS1 (*SI Appendix*, Fig. S17 *C* and *D*). In another example, the later arisen HLH-4 (an NHR-type TF) acts in parallel to the more conserved regulator to promote ADL fate (*SI Appendix*, Fig. S17*E*). Thus, we hypothesize that in the ancestral state with the absence of these intermediate or secondary fate regulators the robustness of neuronal fate specification may be poor due to inefficient gene activation. The addition of these regulators helps strengthen the gene regulatory programs. Indeed, experimental evidence supporting this hypothesis can be found in TRN (34, 35), ASE (36), and ADL (37) neurons. An alternative explanation is that the later arisen genes facilitate the differentiation of specific neuron fates among the broader spectrum of cell fates regulated by the more conserved

TFs. This idea is supported by the much narrower expression pattern of *mec*-3 and *odr*-7 compared to their upstream regulators (38–40).

In addition to the above examples, we also compared the transcriptome age of neurons that were regulated by the oldest (PS1) TF only and the ones regulated by at least one younger (PS2-6) TF and found a small but statistically insignificant difference in TAI scores (*SI Appendix*, Fig. S18). We reason that, although in some cases (e.g., BDU and PVQ neurons) the young transcriptome age may be related to the young age of fate regulators (e.g., PS6 TF *pag-3*), it is also possible to recruit old TFs for the differentiation of a new cell type by genomic innovation in *cis*-regulatory elements or by the evolution of new protein–protein interactions.

Discussion

In this study, we analyzed the transcriptome age of *C. elegans* using the TAI score of whole-organism and single-cell transcriptomes. Using the bulk RNA-seq data, we found that *C. elegans* has a classical phylotypic stage in midembryogenesis when the embryo expressed the oldest genes. Comparing the TAI profiles of zebrafish, *Drosophila*, and *C. elegans* revealed surprising conservation in the timing of the phylotypic stages, which all occurred shortly after gastrulation. In zebrafish, the period of lowest TAI in embryos occurred after the gastrula phase and covered the segmentation and pharyngula, while in *Drosophila* the period started from germ band elongation after gastrulation and lasted until the beginning of germ band retraction (1). Similarly, in *C. elegans*, we identified the morphogenesis (or early organogenesis) phase between the end of gastrulation and the start of elongation to be the period with the oldest transcriptome in the embryos. In all three organisms, the TAI score peaks at the start of gastrulation in early embryogenesis, then declines throughout gastrulation, and reaches the lowest point in the phase following gastrulation. In addition to midembryogenesis, all three organisms showed another period of conserved transcriptomes in early larval development. The similarity among the TAI profiles suggests deep conservation in the pattern of organismal development across animal species.

Morphological and genetic studies largely support each other in identifying the most conserved period of development in nematodes. For example, the variation in early embryonic development across nematode species (41–43) is consistent with the TAI peak in early embryogenesis. Gastrulation imposes stronger constraints for nematodes to adopt similar morphology, which explains the decline of the TAI during gastrulation, but gastrulation also showed variation among nematodes (44) and may be less molecularly conserved than the morphogenesis phase. However, we note a potential complication in the interpretation of the transcriptomic data. A delayed degradation of the mRNA that functioned at an earlier time may cause potential discrepancy for the timing of morphologically and molecularly defined phylotypic stages. This is particularly relevant for early embryogenesis, since maternal genes tend to be older and more conserved than zygotic genes.

By analyzing single-cell transcriptomes, we found that the changes in transcriptome age of the whole organism throughout development can be recapitulated at the single-cell level by the TAI dynamics during the progression of some (but not all) cell lineages, which provided insights into the cellular basis of the hourglass pattern of development. Moreover, our analyses also revealed the variation in transcriptome ages among cell types. We found that the blastomere that gives rise to the germline had older transcriptome age than the somatic tissues, which is consistent with the very low TAI of the oocytes. Sperms, on the other hand, expressed young transcriptomes, causing the males to show higher TAI than hermaphrodites. During early induction of the germ layers, endodermal lineages appeared to express the oldest transcriptome among the three germ layers. Among the differentiated tissues, hypodermis had higher TAI than muscles, and a big variation of transcriptome ages was observed among the neuron types. This variation, together with the difference in the phylostratific age of cell fate regulators, was exploited to understand the evolutionary relationship among neuron types.

The application of transcriptome age to probe cell type evolution operates under two important assumptions. First, the cells that express older transcriptomes (e.g., body wall muscles and motor neurons) are more likely to be functionally conserved and/or evolutionarily constrained, whereas the cells (e.g., hypodermis and chemosensory neurons) that express younger transcriptomes are more likely to evolve new functions and contribute to adaptation. Second, cell types that have older transcriptome ages emerged earlier, whereas cell types that have younger transcriptome ages arose later (e.g., ALM vs. BDU neurons). Although it is debatable whether these two assumptions always hold true, the TAI score provides clues for the evolutionary mechanisms that generate cell type diversity.

Materials and Methods

Phylostratigraphy and the Calculation of Transcriptome Age Index (TAI). The age of *C. elegans* genes was assigned based on previous phylostratigraphic approach (45). Genomes of *Drosophila Melanogaster* and *Mus musculus* were downloaded from Ensembl Web site. The nematode genomes were selected from Wormbase ParaSite (WBPS16) based on the quality of the assembly (Benchmarking Universal Single-Copy Orthologs and annotation scores are both above 90%) (46). For *Pristionchus pacificus*, we used a recently annotated genome with high quality (47). We then applied OrthoFinder v2.5.4 with default parameters (48) to construct orthogroups. Phylostrata (PS) value for each *C. elegans* gene was assigned based on its location related to the most distant outgroup species (Dataset S1). The smaller the PS value, the older the gene. Orphan genes were not included in our analysis. Protein domain for each gene was identified using PfamScan to search against Pfam-A protein database.

Transcriptome age index (TAI) was calculated using the myTAI package (http://CRAN.R-project.org/package=myTAI) as the weighted mean of PS value by the expression level:

$$TAI = \frac{\sum_{i=1}^{n} PS_i E_i}{\sum_{i=1}^{n} E_i},$$

where PS_i is the phylogenetic rank of gene i, n is the total number of genes, and E_i represents the expression value [Transcripts Per Million (TPM) or Fragments Per Kilobase of exon per Million mapped fragments (FPKM)] of gene i. High TAI indicates young transcriptome, whereas low TAI indicates ancient transcriptome. SD of each TAI value was computed by random permutation (1,000 times) of gene ages. Flat line tests were used to calculate the significance of the TAI profile across development.

Bulk and Single-Cell Transcriptomic Data Analysis. Bulk transcriptomic data for the embryos at different time points were obtained from Hashimshony et al. (7) and for larvae, adults, and dauer were derived from the aggregate mean expression on WormBase (WS284). Male and *fog-2(-)* female expression data were downloaded from Thomas et al. (49). Single-cell transcriptomic data for the embryos were obtained from Tintori et al. (12) and Packer et al. (13). The former provided single-cell transcriptomes from zygote to 16-cell stage, whereas the latter provided the transcriptomes of 1,040 (87% of all) nonapoptotic embryonic lineage cells (including both precursors and terminal cells) and 154 terminal cell types. We also obtained the transcriptomes (profiled using CEL-seq) of the founder blastomeres and their in vitro descendants from Hashimshony et al. (7). For differentiated cells, we obtained single-cell transcriptomic data for L2 animals (14), which contained the transcriptomes of 116 cell types, and for L4 animals (15), which contained the transcriptomes

of 168 cell types. We calculated the TAI for all cell types from the five sets of scRNA-seq data (Datasets S2 and S3).

Other Analysis. Detailed information about relative expression, the reconstitution of whole organism transcriptome from scRNA-seq data, TAI calculation for single-cell transcriptomes and pseudo-bulk mean expression, TAI profiles across development for individual cells and lineages, neuron type characteristics, statistical analysis, etc., can be found in *SI Appendix*.

Data, Materials, and Software Availability. This study used previously published datasets (7, 12, 13, 15, 46). The results of our analyses are all included in the article and the *SI Appendix*.

ACKNOWLEDGMENTS. We thank the CeNGEN project, the Goldstein, Murray, Shendure, and Yanai labs, and others for sharing RNA-seq datasets. This study was supported by funds from the National Natural Science Foundation of China (Excellent Young Scientists Fund for Hong Kong and Macau, 32122002), the Research Grant Council of Hong Kong [GRF 17107021, GRF 17106322, and CRF C7026-20G], the Health Bureau of Hong Kong [HMRF 07183186 and 09201426], and the seed fund from the University of Hong Kong [201910159087, 202011159053, and Strategic Interdisciplinary Research Scheme 2019/20] to C.Z. Computational works were performed using research computing facilities offered by Information Technology Services at the University of Hong Kong.

1. T. Domazet-Loso, D. Tautz, A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
2. M. Quint *et al.*, A transcriptomic hourglass in plant embryogenesis. *Nature* **490**, 98–101 (2012).
3. X. Cheng, J. H. Hui, Y. Y. Lee, P. T. Wan Law, H. S. Kwan, A "developmental hourglass" in fungi. *Mol. Biol. Evol.* **32**, 1556–1566 (2015).
4. N. Irie, S. Kuratani, The developmental hourglass model: A predictor of the basic body plan? *Development* **141**, 4649–4655 (2014).
5. L. Wu, K. E. Ferger, J. D. Lambert, Gene expression does not support the developmental hourglass model in three animals with Spiralian development. *Mol. Biol. Evol.* **36**, 1373–1383 (2019).
6. M. Levin, T. Hashimshony, F. Wagner, I. Yanai, Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* **22**, 1101–1108 (2012).
7. T. Hashimshony, M. Feder, M. Levin, B. K. Hall, I. Yanai, Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* **519**, 219–222 (2015).
8. A. Coghlan, Nematode genome evolution. *WormBook*, 1–15 (2005), 10.1895/wormbook.1.15.1.
9. J. H. Thomas, Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res.* **16**, 1017–1030 (2006).
10. J. H. Thomas, H. M. Robertson, The *Caenorhabditis* chemoreceptor gene families. *BMC Biol.* **6**, 42 (2008).
11. S. Wang *et al.*, A high-content imaging approach to profile *C. elegans* embryonic development. *Development* **146**, dev174029 (2019).
12. S. C. Tintori, E. Osborne Nishimura, P. Golden, J. D. Lieb, B. Goldstein, A transcriptional lineage of the early *C. elegans* embryo. *Dev. Cell* **38**, 430–444 (2016).
13. J. S. Packer *et al.*, A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019).
14. J. Cao *et al.*, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
15. S. R. Taylor *et al.*, Molecular topography of an entire nervous system. *Cell* **184**, 4329–4347.e23 (2021).
16. O. Hobert, Neurogenesis in the nematode *Caenorhabditis elegans*. *WormBook*, 1–24 (2010), 10.1895/wormbook.1.12.2.
17. S. J. Cook *et al.*, Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**, 63–71 (2019).
18. L. L. Moroz, D. Y. Romanova, A. B. Kohn, Neural versus alternative integrative systems: Molecular insights into origins of neurotransmitters. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20190762 (2021).
19. M. A. Hilliard, C. I. Bargmann, P. Bazzicalupo, *C. elegans* responds to chemical repellents by integrating sensory inputs from the head and the tail. *Curr. Biol.* **12**, 730–734 (2002).
20. W. Zou *et al.*, Polymodal responses in *C. elegans* phasmid neurons rely on multiple intracellular and intercellular signaling pathways. *Sci. Rep.* **7**, 42295 (2017).
21. C. I. Bargmann, Chemosensation in *C. elegans*. *WormBook*, 1–29 (2006), 10.1895/wormbook.1.123.1.
22. L. U. Sneddon, Evolution of nociception and pain: Evidence from fish models. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190290 (2019).
23. F. Ma, C. Y. Lau, C. Zheng, Large genetic diversity and strong positive selection in F-box and GPCR genes among the wild isolates of *Caenorhabditis elegans*. *Genome Biol. Evol.* **13**, evab048 (2021).
24. S. Sural, O. Hobert, Nematode nuclear receptors as integrators of sensory information. *Curr. Biol.* **31**, 4361–4366.e2 (2021).
25. C. Li, K. Kim, Neuropeptides. *WormBook*, 1–36 (2008), 10.1895/wormbook.1.142.1.
26. D. Arendt *et al.*, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
27. D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
28. M. B. Reilly *et al.*, Unique homeobox codes delineate all the neuron classes of *C. elegans*. *Nature* **584**, 595–601 (2020).
29. V. Vermeirssen *et al.*, Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res.* **17**, 1061–1071 (2007).
30. P. M. Gordon, O. Hobert, A competition mechanism for a homeotic neuron identity transformation in *C. elegans*. *Dev. Cell* **34**, 206–219 (2015).
31. C. Zheng, H. M. T. Lee, K. Pham, Nervous system-wide analysis of Hox regulation of terminal neuronal fate specification in *Caenorhabditis elegans*. *PLoS Genet.* **18**, e1010092 (2022).
32. P. Kratsios *et al.*, An intersectional gene regulatory strategy defines subclass diversity of *C. elegans* motor neurons. *Elife* **6**, e25751 (2017).
33. O. Hobert, Development of left/right asymmetry in the *Caenorhabditis elegans* nervous system: From zygote to postmitotic neuron. *Genesis.* **52**, 528–543 (2014).
34. C. Zheng, F. Q. Jin, M. Chalfie, Hox proteins act as transcriptional guarantors to ensure terminal differentiation. *Cell Rep.* **13**, 1343–1352 (2015).
35. A. Duggan, C. Ma, M. Chalfie, Regulation of touch receptor differentiation by the *Caenorhabditis elegans* mec-3 and unc-86 genes. *Development* **125**, 4107–4119 (1998).
36. S. Chang, R. J. Johnston Jr., O. Hobert, A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* **17**, 2123–2137 (2003).
37. N. Masoudi *et al.*, Unconventional function of an Achaete-Scute homolog as a terminal selector of nociceptive neuron identity. *PLoS Biol.* **16**, e2004979 (2018).
38. R. Baumeister, Y. Liu, G. Ruvkun, Lineage-specific regulators couple cell lineage asymmetry to the transcription of the *Caenorhabditis elegans* POU gene unc-86 during neurogenesis. *Genes Dev.* **10**, 1395–1410 (1996).
39. T. R. Sarafi-Reinach, T. Melkman, O. Hobert, P. Sengupta, The lin-11 LIM homeobox gene specifies olfactory and chemosensory neuron fates in *C. elegans*. *Development* **128**, 3269–3281 (2001).
40. J. C. Way, M. Chalfie, 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. M. Brauchle, K. Kiontke, P. MacMenamin, D. H. Fitch, F. Piano, Evolution of early embryogenesis in rhabditid nematodes. *Dev. Biol.* **335**, 253–262 (2009).
42. E. Schierenberg, Embryological variation during nematode development. *WormBook*, 1–13 (2006), 10.1895/wormbook.1.55.1.
43. B. Goldstein, On the evolution of early development in the Nematoda. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 1521–1531 (2001).
44. B. G. Chitwood, M. B. Chitwood, *Introduction to Nematology* (University Park Press, Baltimore, Maryland, U.S., 1974).
45. T. Domazet-Loso, J. Brajkovic, D. Tautz, A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
46. K. L. Howe, B. J. Bolt, M. Shafie, P. Kersey, M. Berriman, WormBase ParaSite–a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10 (2017).
47. M. Athanasouli *et al.*, Comparative genomics and community curation further improve gene annotations in the nematode Pristionchus pacificus. *BMC Genom.* **21**, 708 (2020).
48. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
49. C. G. Thomas *et al.*, Simplification and desexualization of gene expression in self-fertile nematodes. *Curr. Biol.* **22**, 2167–2172 (2012).