bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Real age prediction from the transcriptome with RAPToR

Romain Bulteau[1] and Mirko Francesconi [1,*]

[1] : *Laboratoire de Biologie et Modélisation de la Cellule, Université de Lyon, ENS, UCBL, CNRS, INSERM, UMR5239, U 1210, F-69364 Lyon, France.*
** : mirko.francesconi@ens-lyon.fr*

Abstract

Genome-wide gene expression profiling is a powerful tool for exploratory analyses, providing a high dimensional picture of the state of a biological system. However, uncontrolled variation among samples can obscure and confound the effect of variables of interest. Uncontrolled developmental variation is often a major source of unknown expression variation in developmental systems. Existing methods to sort samples from transcriptomes require many samples to infer developmental trajectories and only provide a relative pseudo-time.

Here we present RAPToR (**R**eal **A**ge **P**rediction from **T**ranscriptome staging **o**n **R**eference), a simple computational method to estimate the absolute developmental age of even a single sample from its gene expression with up to minutes precision. We achieve this by staging samples on high-resolution reference developmental expression profiles we build from existing time series data. We implemented RAPToR for the most common animal model systems: nematode, fruit fly, zebrafish, and mouse, and demonstrate application for non-model organisms. We show how developmental variation discovered by RAPToR can be exploited to increase power to detect differential expression and to untangle the signal of perturbations of interest even when it is completely confounded with development. We anticipate our RAPToR post-profiling staging strategy will be especially useful in large scale single organism profiling because it eliminates the need for synchronization or for a tedious and potentially difficult step of accurate staging before profiling.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Genome-wide gene expression profiling is a powerful exploratory technique that provides a highly multidimensional and systematic view of the system under study. However, the analysis of gene expression data can be complicated by uncontrolled and unknown sources of variance – that can be technical but also biological in nature[1] – that can mask or confound the effects of variables of interest.

To tackle this problem, several methods have been developed to learn and remove hidden covariates (or surrogate variables) from the data, such as *Remove Unwanted Variance* (RUV)[2], *Surrogate Variable Analysis* (SVA)[3], or *Probabilistic Estimation of Expression Residuals* (PEER)[4]. However, a drawback of these methods is that the sources of variance usually remain obscure, therefore potentially interesting biological variance might also be removed.

A major source of unintended variance when profiling developing and differentiating systems is often developmental progression. This is especially true in organisms with rapid life cycles and highly variable growth speed such as worms, fruit fly or zebrafish, where numerous factors like genetic background, temperature, diet, crowding [5–9], or even the physiological state of the previous generation[9] substantially impact developmental speed. Carefully controlling for all conditions influencing development is therefore particularly challenging, but failing to do so can strongly impact gene expression. For example, in *C. elegans* even a few hours of development may result in 10,000 differentially expressed genes[10]. Hence, it is not surprising that around 50% of gene expression variance in the profiling of a large panel of *C. elegans* recombinant inbred lines[11] is due to unintended developmental variation and that almost 38% of the datasets that did not intend to include development in a *C. elegans* gene expression database[12] show substantial developmental variation in gene expression[10].

Identifying hidden developmental variation and estimating developmental time of samples is important first to quantify the impact of the perturbation of interest on developmental speed (Fig. 1a); second, to distinguish perturbation-specific from unspecific gene expression changes due to development (Fig. 1b); third, to uncover time specific effects of the perturbations under study[13] by including estimated age as a covariate in differential expression analyses (Fig. 1c). In yeast, analogous ideas successfully identified genetic and environmental perturbations impacting specific phases of the cell cycle [14] and direct and specific effects of 700 gene

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deletions on gene expression after removing the main source of variance (25%): a shared expression signature of cell cycle and growth rate [15].

Extracting developmental progression from transcriptomes has recently become a topic of intense research, especially after the advent of single-cell RNASeq. Many algorithms have been developed that learn developmental trajectories from large scale bulk, single cell, or whole-organism transcriptomic data and sort samples along those trajectories (e.g. Slingshot[16], DPT[17], Monocle[18], BLIND[19]). However, a major drawback of these trajectory-learning algorithms is that they require large amounts of samples to learn the developmental trajectory from the data. Moreover, they only provide dataset specific ranks or pseudo-times, making it difficult to compare results across datasets or conditions.

To overcome these limitations, we developed RAPToR (Real Age Prediction from Transcriptome staging on Reference), a computational method that, instead of learning developmental trajectory from the data, exploits available time series gene expression data as reference to determine the absolute developmental age of even a single sample from its transcriptome with high precision. We implemented RAPToR in R (available at https://github.com/LBMC/RAPToR) providing references to stage *C. elegans*, *D. melanogaster, D. rerio,* and *M. musculus* development from gene expression.

We show that RAPToR can stage samples of one species using another species as reference, and can also capture tissue-specific development from whole-organism data. Finally, we show that inferred age allows quantification of a perturbation effect on developmental speed and of perturbation-specific effects on gene expression even when the perturbation is completely confounded by development.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

RAPToR Design

We set out to develop a strategy to stage development from gene expression that would be effective even for limited number samples when trajectory learning methods are not applicable. We reasoned that we could exploit existing developmental time series data as reference to estimate the age of even a single sample by simply taking the time point of the reference with maximum correlation with the sample transcriptome as the age estimate. In this way, not only the age of each sample is inferred independently from others, but age estimates of samples from different experiments, conditions and genetic backgrounds are comparable when acquired on the same reference.

However, one drawback of this approach would be that the precision of age estimates depends on the temporal resolution of the reference. To overcome this limitation, we interpolate reference gene expression (Fig. 1d) with respect to time in a dimensionally reduced space (Fig. 1e, Sup. Note 1), generating interpolated expression profiles between original reference time points (Fig. 1f).

The sample age estimate is simply the time point of maximum Spearman correlation between the interpolated reference and the sample gene expression (Fig. 1g). We then compute the estimate confidence interval by bootstrapping on genes (Fig. 1h, methods).

We implemented this strategy in RAPToR, an R package where we provide functions to interpolate references and stage samples. Moreover, we already provide high resolution interpolated references to stage the most commonly used animal model organisms exploiting existing time series data on roundworm embryonic and larval development[20–22] zebrafish embryonic and larval development[23], mouse[24], and fly[25] embryonic development (Sup. Table 1).

Evaluating RAPToR's performance

Reference interpolation dramatically increases temporal resolution and accuracy of age estimates

To evaluate RAPToR performance, we staged independent time-series data of *C. elegans* late-larval development[26] and zebrafish[27,28], mouse[29], and fly[27] embryonic development.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We found RAPToR age estimates accurately match chronological age for both *C. elegans* and zebrafish (R^2>0.99, Fig. 2a, 2b), as well as morphological staging (somite number) for mouse (R^2=0.95, Fig. 2c) while in fly our age estimates less accurately match chronological age especially for later stages (R^2=0.74, Fig. 2d). However, RAPToR estimates rank the samples similarly to BLIND (ρ>0.99, Sup. Fig. 1) – a trajectory-learning method[19] used by the authors27 – which unlike RAPToR only provides ranks. Furthermore, RAPToR estimates enhance both detection of expression dynamics captured by principal components (Fig. 2e, 2f, Sup. Fig. 2) and model goodness of fit for the majority of genes (Sup. Fig. 1) compared to chronological age (see methods) suggesting that RAPToR staging provides more accurate estimates of physiological age than chronological time.

Crucially, staging a dense zebrafish developmental time course[28] shows that RAPToR accurately stages time series with over 40 times higher resolution than the reference data, demonstrating that reference interpolation effectively increases temporal resolution of age estimates (Sup. Note 1, Sup. Fig. 3, methods). RAPToR estimates also stays remarkably accurate and precise even when staging samples using only a fraction of available genes (Sup. Note 1, Sup. Fig. 4, 5, 6) and are robust to both the choice of dimension-reduction method and the number of components used for reference interpolation (Sup. Note 1, Sup. Figure 7, Sup. Table 2).

RAPToR correctly infers developmental speed scaling factors

RAPToR estimates are relative to the reference chronological age. This means that one can use RAPToR to stage samples with known chronological age to estimate developmental speed differences or scaling factors with a reference. For example, staging a *C. elegans* developmental time series grown at 25°C[26] on the reference grown at 20°C[20] recapitulates the expected 1.5 fold increase in developmental speed[20] due to temperature increase (Fig. 2a, Sup. Note 1).

RAPToR stages dissected tissue samples well

We tested RAPToR performance on expression data from dissected tissues – where variation in cell type composition and relative amount might potentially confound staging – using time-series of *M. musculus* upper and lower-jaw first molar embryonic development[30,31]. Since these two organs have very similar transcriptomic signatures[30], we built a lower jaw reference to stage the upper-jaw (see methods). RAPToR not only accurately estimates age (R^2>0.99, Fig. 2g), but also correctly estimates the known developmental delay of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

upper molars compared to lower molars[30,31]. Thus, despite potential confounders, RAPToR is effective and precise on dissected tissue samples.

RAPToR age estimates are robust to genetic variation in gene expression.

Variable genetic background is another potential confounder for RAPToR so we tested RAPToR performance on expression data for over 200 *C. elegans* recombinant inbred lines (RILs) that shows extensive genetic variation in gene expression[11]. This dataset was already staged by a trajectory-learning approach and found to span mid-larval to young adult stage[13], a period with vast expression changes both in the soma (molting) and the germline (spermatogenesis, oogenesis).

RAPToR age estimates closely match those previously found (R^2=0.94, Sup. Fig. 8). However, we noticed that some gene expression dynamics are advanced and others delayed compared to the reference (Fig. 3a, Sup. Note 1). Shifts between soma and germline developmental time (soma-germline heterochrony) are easily induced by environmental and physiological changes in *C. elegans*[9,32]. Indeed the advanced and delayed dynamics are consistently enriched in soma and germline genes respectively (Fig. 3d, Sup. Fig. 9) suggesting soma-germline heterochrony between the reference and the RILs.

Tissue specific staging enables quantification of heterochrony

To confirm this we used germline- and somatic-specific gene sets[22,26] to separately stage the germline and soma in the RILs (see methods, Sup. Fig. 8). Indeed, we find germline- and soma-specific dynamics align better on the reference when staged with the corresponding gene set (Fig. 3b, 3c) while they are otherwise shifted, confirming heterochrony between the reference and the RILs. Thus tissue specific staging outperforms global staging in case of heterochrony between the reference and the samples to stage.

We also noticed that tissue specific staging not only corrects the heterochrony between the RILs and reference but also decreases heterochrony variance among the RILs. Indeed, germline genes are better fit by germline than soma age and vice versa, suggesting soma-germline heterochrony among the RILs (Sup. Fig. 10). However, when we searched for the genetic bases of this heterochrony performing a multivariate QTL analysis, we found no significant genetic locus at an FDR of 0.5 and overall no significant amount of genetic variance in heterochrony (Sup. Note 1) which is therefore likely due to unknown and uncontrolled environmental variation or to a very complex genetic architecture which is not captured by the model.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In summary, RAPToR provides accurate tissue-specific age estimates from whole-organism expression despite varying genetic background.

Staging on references of a different species

Developmental time series data are often unavailable for non-model organisms. However, gene expression dynamics during development are often well-conserved across related species, especially during the phylotypic stage[33]. Encouraged by RAPToR robustness to genetic variation within species, we decided to test how well RAPToR can stage one species on a related species.

Staging time series of embryo development across 6 *Drosophila* species[33] on a *D. melanogaster* reference using orthologs indeed results in accurate age estimates (R^2=0.997, Fig. 4a) despite decreasing overall correlation with increasing phylogenetic distance (Fig. 4b). Moreover, we infer between species growth speed differences matching those calculated by the authors (Sup. Table 3). Importantly, we also detect small age differences between replicates of each time point, which refine expression dynamics (Sup Fig. 11), thus reducing unexplained variance in the data (Sup. Fig. 12).

Encouraged by this, we probed RAPToR limits by staging on a distant species reference. To our surprise, we could successfully stage *C. elegans* embryogenesis[27] on a *D. melanogaster* reference (R^2 = 0.958, Fig. 4c, Sup. Note 1, Sup. Fig. 13), two species separated by 600 million years of evolution[34].

Which biological processes with an extremely conserved dynamics during embryogenesis could account for this accurate staging? We found that a gene expression signature of decreasing cell proliferation shared across phyla[27] and a signature of muscle development are necessary and almost sufficient for accurate staging (Sup. Note 1, Sup. Fig. 13, Sup. Table 4, 5, 6, methods).

Thus RAPToR can stage non-model organisms using available close species data and perform well even in extremely distant species, at least when applied to developmental stages with highly conserved developmental dynamics.

To summarize, RAPToR performs well across the organisms, sample types, and diverging genetic backgrounds and species we tested, yielding estimates that are accurate, precise thanks to interpolation, and robust to gene set size changes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RAPToR provides biological interpretation of drug effects

RAPToR absolute age estimates are useful in many ways. First, instead of just obtaining a list of differentially expressed genes from expression profiling data, using RAPToR precisely quantifies the effect of variables of interest on developmental timing, including in a tissue-specific manner. For example, tissue-specific staging of *C. elegans* exposed to three concentrations of mefloquine, dichlorvos, and fenamiphos [35] found that all three drugs induce a similar germline-specific and dose-dependent developmental delay (Fig. 5a, Sup. Note 2, Sup. Fig. 14).

RAPToR increases statistical power in differential expression analyses

Even when chronological age is known, including RAPToR age estimates as a model covariate instead of chronological age increases power in differential expression (DE) analyses. For example, including RAPToR estimates instead of chronological age when analyzing expression changes in *C. elegans pash-1* vs wt [36] (Fig. 5b), detects up to 60% more DE genes in *pash-1* and 10% more DE genes across development thanks to overall better model fits (Fig. 5c, Sup. Fig. 15, Sup. Note 2).

Quantifying differential expression due to differences in development

Often, when perturbations strongly impact developmental speed and controlling for age between experimental groups is challenging, development and variable of interest are completely confounded. In this scenario, detecting perturbation specific effects by including age as a model covariate is not feasible. However, not accounting for confounding developmental variation can lead to misleading conclusions as purely developmental expression changes are attributed to the perturbation of interest. To show an example of this, we reanalyze a dataset comparing young adult *C. elegans* that developed through dauer state (post-dauer) to controls that did not[37]. The authors found a down-regulation of spermatogenesis-associated genes and an up-regulation of oogenesis-associated genes from which they concluded that post-dauer animals have reduced spermatogenesis and increased oogenesis. However, as *C. elegans* switch from spermatogenesis to oogenesis during development, this pattern could simply be explained by post-dauer samples being older than controls. This is indeed what RAPToR found (Fig. 5d, Sup. Fig. 16, Sup Note 2). Furthermore, the strong correlation (r = 0.8) between the observed expression changes in germline genes and the expected developmental expression changes calculated from matching time points in the reference (Fig. 5e, Sup. Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16, 17, Sup Note 2) suggests that, despite synchronization efforts, most of the initially observed DE is due uncontrolled developmental variation.

Recovering specific effects even when the variable of interest is completely confounded with development

We reasoned that including reference data in differential expression analysis should provide enough data to extract perturbation-specific expression changes even when the variable of interest is completely confounded with development (Sup. Note 2, Sup. Fig. 18). We validated our approach using C. elegans larval development time series of xrn-2 mutant and relative relative wild-type (WT) control sampled every 1.5h[38]. We defined a gold standard of truly DE genes in the mutant and quantified the amount, intensity and the variance of expression changes due development as well as the decreasing performance of a standard linear model p-value in recovering truly DE genes at increasing age differences between mutant and WT (Fig. 5f-h, Sup. Note 2, Sup. Fig. 18). We found that the reference data integrated model effectively recovers truly DE genes for large age differences when mutant effect is completed confounded by development (Fig. 5i). At small age differences the detection of true DE is maximized by an age corrected classifier that combines the log fold change (logFC) from the reference integrated model with the p-value of a standard linear model weighted according to variance in observed expression changes explained by development (Fig. 5i, Sup. Fig. 18, Sup. Note 2).

In summary, we showed that using RAPToR and reference data it is possible to quantify developmental effects on gene expression and recover the specific effect of a perturbation even when completely confounded with development.

Discussion

We presented here RAPToR, a computational strategy to accurately stage samples from their genome-wide gene expression profile. Unlike trajectory-based methods, RAPToR exploits existing reference time-series data to stage each sample separately, providing several advantages: first, it eliminates the need for large datasets to infer developmental trajectories; second, it provides absolute developmental times that are comparable across data sets, conditions, genetic backgrounds, profiling technologies and other covariates; third, with RAPToR outliers have no impact on the staging of other samples.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

While RAPToR staging is limited by the existence of reference time-series data, reference interpolation allows precise staging well beyond the resolution of the original reference data, enabling the use of sparse time series as references. More importantly, we validated staging of one species on a close species reference, which dramatically expands the scope of RAPToR, including to non model organisms. Moreover, RAPToR works well on dissected tissue samples and can also infer tissues-specific age from whole-organism profiles.

We showed how RAPToR absolute estimates can be exploited in many ways: to detect the effect of a perturbation on developmental speed; as model covariates to increase statistical power to detect differential expression. Finally, we showed that even in the extreme scenario when the perturbation of interest is completely confounded with development, it is still possible to recover genuine perturbation-specific expression changes by integrating reference data in differential expression analysis.

We anticipate our RAPToR post-profiling staging strategy will be especially useful in large scale single organism profiling because it eliminates the need for synchronization or for a tedious and potentially difficult step of an accurate staging before profiling.

To conclude, we remark that our approach is not restricted to development but can in principle be applied to any process with robust underlying reference gene expression dynamics (e.g. cell differentiation, cell cycle, aging, disease progression, drug response) and its scope will only increase with the increasing availability of time series profiling data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Analyses were all performed using the R statistical software (v3.6.3)

Data accessibility

All the data used in this study were previously published and deposited in public databases or accessible by request to the authors. The data from Sémon et al. is, at the time of writing, awaiting publication[31]. The full list of datasets and accession numbers is given in Supplementary Table 7.

The code to download and (pre)process the data, perform the analyses and generate the figures of this paper can be found at https://gitbio.ens-lyon.fr/LBMC/qrg/raptor-analysis

Data pre-processing

Probe or gene IDs of datasets were converted to standard IDs (WBGene IDs for *C. elegans,* FBgn IDs for *D. melanogaster*, Ensembl IDs for *D. rerio* and *M. musculus*). When multiple probes or IDs matched a single standard ID, they were mean-aggregated for microarray, sum-aggregated for RNA-seq counts. IDs with no standard ID match were dropped.

For RNA-Seq datasets, TPM data was used when available, or computed from raw counts using the transcript lengths from the Ensembl biomart (v99). No remapping of the transcriptomes was done, aside from the *M. musculus tooth* data (see below). No background correction was applied to microarray data.

Samples were considered of poor quality and discarded when the 99th percentile of the distribution of their Spearman correlation coefficients with others samples fell below a threshold defined below for each dataset.

Expression values for all datasets were quantile-normalized using the *normalizeBetweenArrays* function from *limma*[39] (v3.42.0) on $log(X+1)$ transformed values unless otherwise specified.

RAPToR implementation

Our method is implemented in an R package : RAPToR (v1.1.4), which can be downloaded and installed from the following url. https://github.com/LBMC/RAPToR

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functions for staging samples, plotting results, interpolation and building references are included in the package. Detailed vignettes on general usage, reference building and showcases are also provided with the package.

Auxiliary R data-packages include references for *C. elegans* (embryonic, larval and young adult to adult development, https://github.com/LBMC/wormRef), *D. melanogaster* (embryonic development, https://github.com/LBMC/drosoRef), *D. rerio*, (embryonic and larval development, https://github.com/LBMC/zebraRef) and *M. musculus* (embryonic development, https://github.com/LBMC/mouseRef).

Reference interpolation

Let $\boldsymbol{X}$ ($\boldsymbol{m} \times \boldsymbol{n}$) be the gene expression matrix of $\boldsymbol{m}$ genes by $\boldsymbol{n}$ samples. The matrix is first gene-centered such that $\boldsymbol{X_0} = \boldsymbol{X} - \boldsymbol{rowMeans(X)}$. We then use ICA (*'ica'* function, '*icafast'* library v1.0.2) or PCA (*'prcomp'* base R function) to decompose the data into a component space of dimension $\boldsymbol{c}$ such that $\boldsymbol{X_0} = \boldsymbol{G}\,\boldsymbol{S^T}$, with $\boldsymbol{G}$ ($\boldsymbol{m} \times \boldsymbol{c}$) the gene loadings and $\boldsymbol{S}$ ($\boldsymbol{n} \times \boldsymbol{c}$) the sample scores. Columns of $\boldsymbol{S}$ are interpolated on with respect to time (and other potential variables of interest, *e.g.* batch), forming a new matrix $\boldsymbol{T}$ ($\boldsymbol{l} \times \boldsymbol{c}$) of $\boldsymbol{l}$ new time points in component space. The full interpolated expression matrix $\boldsymbol{Y}$ ($\boldsymbol{m} \times \boldsymbol{l}$) is then reconstructed by multiplying the gene loadings matrix by the transposed $\boldsymbol{T}$ and by adding the gene centers $\boldsymbol{Y} = \boldsymbol{G}\,\boldsymbol{T^T} + \boldsymbol{rowMeans(X)}$.

To interpolate the components, we fit Generalized Additive Models (GAMs) to handle non-linear dynamics through splines with the '*gam'* function in the '*mgcv'* package (v 1.8.31) using a single model formula for all components selected by Cross-Validation (CV) as following: CV training sets are built with 80% of samples, with proportional representation of any covariate group (e.g. batch). The model is evaluated using the average relative error, mean squared error (MSE), and average root MSE [40]. We compared GAMs fitted with different splines (cubic, thin plate, duchon), and chose the model with minimal CV and prediction errors. Automatic spline parameter estimation from '*gam'* function was used. If the model was clearly performing poorly with automatic parameter estimation (overfitting, predictions not matching the component dynamics), we performed further CV on reasonable spline parameter spaces to tweak the model (defining a number of knots). We further verified that RAPToR age estimates match chronological age of the original reference data and of independent time series when staged on the interpolated reference, using the R^2 of linear models (Sup. Note 1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The number of components to fit was selected by setting a cutoff on cumulative explained variance (e.g. 99%). The cutoff was adjusted according to the number of components with intelligible dynamics with respect to time. Interpolation is robust to variation in the number of components used (Sup. Note 1).

We implemented reference interpolation with the '*ge_im'* function in the 'RAPToR' package. Model formulas and parameters for building all the references used in this study are displayed in Sup. Table 1.

Age estimation

To perform age estimation, we implemented the *'ae'* function that takes the gene expression matrix to stage (genes as rows, samples as columns), the reference matrix (genes as rows and time points as columns), and the reference times (time values associated with the columns of the reference matrix) as inputs. The *'ae'* function then finds common genes between sample and reference and computes the Spearman correlation between each sample and each reference time point. The age estimate for each sample is simply the reference time point with the highest correlation.

When an age estimate lands within 5% of the reference's edges, we implemented a warning suggesting to stage the samples on another appropriate reference if possible.

To compute confidence intervals on age estimates, staging is repeated on bootstrap gene samples of default size of one third of the total. Unless stated otherwise, the number of bootstraps is 30. A confidence interval is given by the median absolute deviation (MAD) of bootstrap estimates (est_{boot}) from the global estimate (est), and the resolution of the interpolation (res, time interval between 2 points of the interpolated reference) :

$$[est - (median(|est - est_{boot}|) + res/2) ; est + (median(|est - est_{boot}|) + res/2)]$$

Staging using a prior probability

We implemented the possibility of providing a prior probability in the form of parameters for a gaussian distribution per sample (mean, sd) which must be given in the time scale of the reference. A gaussian density function over the reference time is defined per sample from these parameters. During staging, all correlation peaks of the profile are determined and ranked by averaging their scaled correlation score (height of the peak in the correlation profile scaled to $[0 , 1]$) and prior score (value of the gaussian density function scaled to $[0 ,$

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1], at the peak time point). The first peak of the ranking is then kept as the estimate. Since the ranking is determined by averaging normalized priors and correlation scores, changing the prior standard deviation parameter results in scaling the importance of the prior with respect to the correlation information.

No priors were used for staging unless explicitly stated.

Evaluating RAPToR performance

Staging C. elegans larval development

We built the reference from a time series of WT larval development at 20°C sampled at 26 time points from L1 feeding to 48 hours[20] (see Sup. Table 1), we set the number of interpolated time points to 500.

Staged samples are WT *C. elegans* collected during mid to late larval development at 25°C from 22 to 37 hours after L1 feeding[26]. Only samples aged below 32 hours (corresponding to about 48 hours at 20°C) were staged, to stay within the reference boundaries.

Staging D. melanogaster embryonic development

We staged a *Drosophila* developmental time series [27] on an interpolated reference from another embryo developmental time series[25] (Dme_embryo reference of the drosoRef package, see Sup. Table 1). Samples were discarded when the 99th percentile of the distribution of their Spearman correlation coefficients with others samples fell below 0.6, leaving 90 samples to stage. The number of interpolated time points in the reference was set to 500.

We compared our rankings with the BLIND[19] rankings provided in the supplementary data [27] (restricting to 77 samples as the authors used a more stringent quality cutoff).

To test if our age estimates better capture physiological development than chronological age, we fit identical linear models using the '*lmFit'* function of '*limma'* with either chronological age or RAPToR estimates as the predictor. Age is modeled using a natural cubic spline with 2 to 8 degrees of freedom (built with the *ns* function of the *splines* package). For each gene, we use R^2 to compare the goodness of fit of the models with chronological age or RAPToR age estimates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Staging D. rerio embryonic development

We used the interpolated reference we built from embryo and larval development data [23] (Dre_emb_larv reference of the zebraRef package, see Sup. Table 1) to stage a zebrafish time course of embryonic development from fertilization to 72 hours post-fertilization[27]. Samples were discarded when the 99th percentile of the distribution of their Spearman correlation coefficients with others samples fell below 0.6, leaving 93 samples to stage. The number of interpolated time points in the reference was set to 1 000.

We then used the same reference, increasing the interpolation resolution between 0 and 15h to 800 time points (resulting in a reference time density of around 1 time point per minute instead of the previous 1 time point per hour) to stage an additional dense embryonic time series of 180 zebrafish embryos around gastrula[28]. We compare RAPToR staging to rankings (Sup. Fig. 3a) previously determined[28] as following: the 10 youngest and oldest embryos (determined through the morphological criterion of epiboly coverage) are used to select the genes with the largest decrease in expression from start to end of the time course. The average expression of these genes then determines the ranking.

Staging M. musculus embryonic development

We used the interpolated reference we built from mouse embryonic development time series data [24] (Mmu_embryo reference of the mouseRef package, see Sup. Table 1) to stage an independent mouse somite-staged developmental time course [29]. The number of interpolated time points was set to 500. We compare RAPToR staging with the provided embryos somite number as no chronological age is given [29].

Staging M. musculus first-molar embryonic development

First and second data replicates for mouse first molar embryonic development are from Pantalacci et al.[30], and Sémon et al.[31] respectively. Reads from both replicates were processed together, trimmed with *trimmomatic*[41] (v0.39) to remove adapters, and mapped using *salmon*[42] (v0.14.1) and the Ensembl 98 version of the mouse transcriptome to obtain TPM values.

Genes with a median expression of log(TPM+1) < 0.5 across all samples were filtered out, leaving 15362 genes. A reference was built from both replicates of the lower jaw samples (see Sup. Table 1) and used to stage all 32 samples.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Estimating developmental speed factors and resolution increase factors

Developmental speed factors and R^2 between chronological and estimated age of samples are estimated with linear models.

We call 'resolution increase factor' the factor between sampling frequencies of a reference prior to interpolation and of a successfully staged independent time series.

C. elegans larval development is sampled every 2 hours at 20°C (0.5/h) in the reference [20] and every hour at 25°C (1/h, 1.5 development speed factor) in the staged time series [26] resulting in a resolution increase factor $rf = (1.5 * 1)/0.5 = 3$.

Drosophila embryo development is sampled every 2 hours (0.5/h) in the reference [25] and every 15 min (4/h) in the staged time series [27], resulting in a resolution increase factor $rf = 4/0.5 = 8$.

Mouse embryo development is sampled every 1.5 days (0.66/day) in the reference [24] and somite-staged in the target time series [29]. Since the first 30 somites of *M. musculus* grow in ~2.5 days[43], the somite-staged times series has a resolution of 12 time points per day (12/day) determining a resolution increase factor $rf = 12/0.66 = 18.2$.

Zebrafish embryo development is sampled every hour (1/h) in the reference [23] and at a rate equivalent to 47 per hour (47/h) in the staged samples[28] (180 samples are roughly evenly staged between 5.7 and 9.5 hours post-fertilization: $180 / (9.5 - 5.7) = 47$/h), resulting in a resolution increase factor $rf = 47$.

Probing robustness of reference interpolation

Robustness of reference interpolation to the choice of dimensionality reduction method and number of components was evaluated using either the *C. elegans* time series by Kim et al.[20] (as above), or the one by Meeuse et al. [21] as references.

Robustness was evaluated computing Sum Squared (SSQ) of gene expression prediction error by reference models using PCA or ICA and 2 to 16 components with the Kim et al. time series, and 2 to 20 with the Meeuse et al. one. The model formula was fixed to the one defined in Sup. Table 1. The SSQ prediction error is defined as $SSQerror = \Sigma((X_{(n \times m)} - X_{pred})^2) / (n * m)$, with n samples, m genes.

For 6 conditions – ICA/PCA, each at 3 different numbers of component – we staged the reference samples as well as an independent *C. elegans* time series[26] on the interpolated reference (only samples within reference

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

boundaries were staged on the Kim et al. reference). We evaluated models built from 4, 9, and 14 PCA or ICA components for the Kim et al. reference and models built from 10, 20 and 25 PCA or ICA components for the Meeuse et al. reference.

We then reported the R^2 value of a linear fit of RAPToR estimates by the chronological age of the samples in each condition (Sup. Table 2), as well as the correlation score between the samples and the interpolated reference at the estimate (Sup. Fig. 7).

Estimating the impact of gene set size on staging

The impact of the gene set size on staging was evaluated by staging the *C. elegans* larval time series by Hendriks et al.[26] on the reference built from the Kim et al.[20] samples, as above.

We staged the samples using 50 random gene sets of sizes 16 000, 12 000, 8 000, 4 000, 2 000, and 1 000. The resulting estimates were used to compute confidence intervals for varying bootstrap set sizes. We reported the median absolute deviation of estimates to the full gene set estimate plus interpolation resolution (i.e. the size of half the confidence interval).

The same approach was repeated for smaller gene set sizes of 2 000, 1 000, 500, and 250, this time staging the samples with and without priors (defined as 1.5 times the chronological age of the samples to account for the developmental speed difference with the reference; prior standard deviation was set to 10).

Tissue-specific staging and quantification of soma-germline heterochrony

Microarray intensities of the Recombinant Inbred Lines (RILs) profiles [11] were first normalized within arrays with LOESS using the '*normalizeWithinArrays'* function of the '*limma*' library. Arrays corresponding to pooled mixed stage controls were then discarded. Samples were discarded when the 99th percentile of the distribution of their Spearman correlation coefficients with others samples fell below 0.95 , leaving 193 samples for analysis.

The reference used to stage the samples is the “Cel_larv_YA” reference[21] of the wormRef package (see Sup. Table 1). The number of interpolated time points in the reference was set to 1000.

Samples were first staged using the entire available gene set to obtain the global estimates, then with somatic and germline specific gene sets to obtain the corresponding tissue-specific estimates: the somatic gene set

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

corresponds to the oscillatory genes denoted "osc" in Hendriks et al. [26]. The "germline" gene set corresponds to the union of "germline_intrinsic", "spermatogenesis_enriched", and "oogenesis_enriched" gene sets defined in Reinke et al.[22]. Estimating somatic age, required the use of the global estimate as prior (due to gene expression oscillations generating multiple correlation peaks), with the prior standard deviation set to 10 for all samples. Germline age estimates required no prior.

To compare expression dynamics between reference and RILs, we kept the overlapping genes between the non-interpolated reference and the samples, quantile-normalized both datasets together, and performed an ICA ('*ica'* function of '*icafast'*) extracting 46 components, explaining 95% of the variance in the joined data. A two-sided hypergeometric test was used to evaluate the enrichment of the components in soma, oogenesis and spermatogenesis genes selecting genes above 1.96 of the absolute value of gene loadings (with the exception of IC1 which captured batch effect) and p-values were adjusted with the Benjamini-Holm method.

To test the existence of heterochrony among the RILs, we fit identical models on the RIL expression data using '*lmFit'* function in *limma* with global, soma, or germline age values as predictors. We used natural cubic splines ('*ns'* function in the '*splines'* library) on the age with 4, 6, or 8 degrees of freedom. Choice between models (at equal spline degrees of freedom) was done per gene based on highest R^2 value.

Quantitative Trait Loci (QTL) analysis on soma-germline heterochrony

The multivariate QTL analysis on soma-germline heterochrony among RILs defined as (soma age) - (germline age) was performed by Random Forest (RF) regression [44] with or without batch as a covariate. Each RIL was genotyped at 1455 SNP markers[11]. Redundant markers were filtered out from the selected 193 RILs, missing values for the remaining 1105 markers are imputed with the '*rfImpute'* function and random forest regression was fit with 5000 trees using the '*randomForest'* function; both functions are from the 'randomForest' package (v4.6.14). The RF Selection Frequency (RFSF) was used as importance measure, adjusted for selection bias[44] which was estimated by fitting 500 forests of 10 trees to gaussian noise.

We estimated the null probability distribution of RFSF through 100 trait permutations, calculated empirical p-values and adjusted them for FDR.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cross-species staging

Staging non-model Drosophila on D. melanogaster

We used the interpolated reference we built from the *D. melanogaster* embryo development [25] (Dme_embryo reference of the drosoRef package, see Sup. Table 1) to stage time courses of development of 6 *Drosophila* species [33] : *Drosophila melanogaster*, *simulans*, *ananassae*, *pseudoobscura*, *permisilis* and *virilis* profiled by microarrays. We used orthologs provided by the authors[33]. The number of interpolated time points in the reference was set to 500.

Developmental speed difference from *D. melanogaster* was determined with a linear model without intercept predicting RAPToR estimates with the chronological age of samples, with species as covariate and including interaction. Comparison with the original scaling factors[33] is shown in Sup. Table 3.

To compare the RAPToR estimates and the linearly-scaled age from the study[33] as developmental predictors, we fit identical linear models on gene expression (*lmFit* function of *limma*) with either the linearly-scaled age or RAPToR estimates as predictor, and species as covariate. Age is modeled using a natural cubic spline with 2 to 8 degrees of freedom (*ns* function of *splines*). For each gene, we use R^2 to compare the goodness of fit of either model. No interaction between age and species coefficients was considered as temporal scaling of development between species is already applied.

We evaluated the effect of species distance on staging through the maximal correlation coefficient between the samples and the reference (i.e. at their age estimate).

Staging C. elegans on Drosophila

We staged a *C. elegans* embryo time series[27] on the interpolated reference we built from the *D. melanogaster* embryo development time series[25] ("Dme_embryo" reference, drosoRef package). First, poor quality *C. elegans* samples were discarded when the 99th percentile of the distribution of their Spearman correlation coefficients with others samples fell below 0.67. Additionally, a sample (GSM1487346, or "sample_0029") was also excluded as it clearly appeared as an outlier on multiple ICA components (Sup. Fig. 13). 4 samples (GSM1487318, GSM1487319, GSM1487320, GSM1487321, or "sample_0001" through "_0004") were further removed due to erroneous chronological age (Sup. Fig. 13), leaving 127 samples.

We then performed the staging using a restricted fly-worm ortholog set [34].

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We also did staging on a second reference interpolated as above but using the first 2 instead of 8 components.

For both interpolated references, the number of interpolated time points was set to 500.

Further analysis is restricted to the overlapping set of orthologs between worm and fly datasets (3194 genes). We ranked genes by Spearman correlation between the *C. elegans* embryo time series and their matching timepoints in the second *D. melanogaster* reference. We then selected the 10% genes with highest correlation (319 genes) and staged the *C. elegans* samples once more on the second *D. melanogaster* reference, evaluating staging performance with Spearman correlation and the R^2 of a linear model between chronological age and estimated age.

Hierarchical clustering the top 10% genes in the original *D. melanogaster* reference data [25] (‘*hclust’* function on the euclidean distance matrix of gene-centered log(TPM+1)), resulted in 3 clusters with over 10 genes. We then evaluated gene ontology enrichment in each cluster with gProfiler [45] using the 3194 overlapping set of worm-fly orthologs as background (Sup. Table 4, 5, 6).

Exploiting RAPToR age estimates

Drug dose response on developmental delay in C. elegans.

Expression profiles of young *C. elegans* adults exposed to drugs [35] were staged on the “Cel_larv_YA” reference[21] from the wormRef package (Sup. Table 1), with 500 interpolated time points in the reference. We estimated global, soma-specific, and germline-specific ages (see Tissue-specific staging). For each age type, we then subtracted the age of the control sample within each replicate of each drug assay to compute the developmental difference by treatment group. We fit a linear model with drug, dose, and interaction on the age differences to assess the significance of the effects.

Increasing statistical power in differential expression analyses

WT and *pash-1*ts *C. elegans* samples [36] were staged on the “Cel_YA_2” reference[22] from the wormRef package (Sup. Table 1), with 500 interpolated time points in the reference. The second replicate of the first wild-type time point (wt_h0.2) was omitted from further analysis due to its extreme developmental displacement and lack of comparable mutant sample.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We fit identical linear models with the '*lmFit'* function in the '*limma*' library to test for differential expression, including either chronological or estimated age modeled with a natural cubic spline ('*ns*' function in '*splines'*, df = 2), strain and their interaction.

Effect of strain and development was then assessed by considering the significance of appropriate model coefficients (interaction and strain coefficients for strain effect, spline and interaction coefficients for development effect), with the '*topTable'* function in the '*limma*' library. Differential expression was considered significant at 0.05 Benjamini-Hochberg False Discovery Rate (FDR).

To test the effect of similar random age differences from chronological age, we generated 100 "random age" sets by sampling age differences from the distribution of (chronological age) - (estimated age) values, estimated with the '*density'* function in R. Sampled age differences were then added to the chronological age, and the same model and analysis as above was applied. The goodness of fit per gene is assessed using R^2 .

Quantifying developmentally driven gene expression changes

Given any two groups of expression profiling samples 'A' and 'B', we first stage them, then fit a linear model per gene on $\log_2$(TPM+1) (or $\log_2$(Intensity+1) for microarray expression data) to compute the observed $\log_2$-fold changes of 'A' vs. 'B' samples. Then we fit the same model on reference profiles at matching time points to compute log2-fold changes expected from development only (Sup Fig. 17) and we use squared Pearson correlation between observed and expected logFCs to quantify the variance explained by development in the observed logFC.

Control and post-dauer *C. elegans* samples[37] were germline-staged (see *Tissue-specific staging*) on the "Cel_larv_YA" reference[21], and on the "Cel_YA_2" reference[22] of the wormRef package for confirmation, as they landed near the edges of the first reference. The number of interpolated time points in the Cel_larv_YA and Cel_YA_2 references were set to 1000 and 500 respectively. Using the method described above, we quantified the differential expression explained only by difference in developmental stages between the control and post-dauer samples.

We could not compare our results to the original results as we were unable to exactly reproduce the distribution of DE and p-values of the original t-test based analysis. We therefore recalculated DE gene expression using linear models (function '*lmFit*' in '*limma*' library in R).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Recovering direct perturbation effects using reference data.

WT and *xrn-2* time series of *C. elegans* late larval development[38] were staged on the "Cel_larv_YA" reference[21] from the wormRef package (Sup. Table 1), with 500 interpolated time points. We restricted further analysis to the genes with both ≥5 raw counts for at least one sample, and overlapping with the reference gene set (17656 genes).

Defining the differential expression gold standard

To establish the gold standard of DE genes, we selected time points 8 to 10 of *xrn-2* and WT, as they had the best (estimated) developmental match. We then calculate differential expression fitting a generalized linear model (GLM) on raw counts using the *glmFit* function of *egdeR* (v3.28.1), including only the strain variable (model 1), and considered genes DE with Bonferroni-Holm adjusted p-values < 0.05 of a likelihood ratio test (*glmLRT* function of '*edgeR*') on the strain coefficient.

Evaluating gold-standard gene detection decrease with age gap

To test how increasing mismatch in developmental time between *xrn-2* and WT impacts DE analysis we apply the same GLM used for the gold standard (model 1) to calculate differential expression between the mutant and WT samples shifted by -1, -2, -3, -5, and -7 time points and we estimated expression changes explained by development as detailed above (*Quantifying developmentally driven gene expression changes).* We then evaluated how well model 1 p-values detect gold standard DE genes at increasing age gaps by Precision-Recall Curves (PRC) and area under PRC using the '*prediction'* function of the '*ROCR'* package (v1.0.11).

Correcting expression changes from development

To accurately account for developmental changes we combine the samples of interest with the interpolated reference.

For each set of samples (including WT and mutant samples), we define the window of reference to include as the range of age estimates widened by a 1 hour margin on either side. For example, in the 'WT-1' set, the youngest sample (WT_05h) is 51.7h old, and the oldest (xrn.2xe31_09h) 58.3h old. Thus, we include the interpolated reference from 50.7h to 59.3h of development.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We transform the interpolated reference data to artificial counts assuming a fixed library size of 25*10^6 counts per sample and a fixed number of reads "per gene length" defined by the median of available gene lengths :

ArtificialCounts = (interpolatedTPM/(10^6))* ((25*10^6)/median(geneLengths)) * geneLengths

The artificial count matrix is then joined to the sample count matrix, and a GLM is fit ('*glmFit'* in '*edgeR"),* including batch (between reference and sample data), the variable of interest (strain) where reference data is grouped together with the control, and developmental time modeled with splines ('*ns*' function in '*splines*'). To select the optimal spline degree of freedom for each window, we minimized the residual sum of squares of a linear model fit on the reference window only (Sup. Fig. 18g). Only model coefficients of the variable of interest (strain logFCs) are considered.

We first evaluated how well strain logFCs detects DE genes from the gold standard using PRC and AUPRC ('*prediction'* function in '*ROCR'*). We then defined an Age-Corrected Classifier (ACC) as the weighted mean of the model 1 p-value and strain logFC of the model including the reference :

$$ACC = w * strainLogFC + (1-w)*(-\log_{10}(model1Pval))$$

with ***w***, the weight ratio of either classifier. We defined the optimal ***w*** as the value for which the area under the precision recall curve (AUPRC) is maximal, and estimated it for each set of WT shifts. At optimal ***w***, we then reported the AUPRC of our age-corrected classifier and compared it to the standard model.

As the optimal ***w*** cannot usually be estimated in this way, we explored the relationship between optimal ***w*** and the correlation between observed and expected logFC (as defined in *Quantifying developmentally driven gene expression changes*) calculated for a larger amount of WT 3-sample sets (Sup. Table 8).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We are grateful to Sarah E. Hall, Marie Sémon, and Sophie Pantalacci for providing data from their profiling experiments. We are also grateful to Gael Yvert, Daniel Jost, Marie Sémon, Sophie Pantalacci, and Ben Lehner for their critical reading of the manuscript.

Fundings

M.F. is supported by INSERM. Work in M.F.'s lab is supported by a grant from the Agence Nationale pour la Recherche (ANR-19-CE12-0009 "InterPhero"), Université de Lyon (IDEX IMPULSION G19002CC) and ENS-Lyon (Projet emergent 2019). R.B. PhD fellowship is funded by the french ministry of research.

Author Contributions

MF and RB conceived the method, RB developed the tool and performed the analyses, MF and RB wrote the manuscript.

Competing interests

The authors report no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1

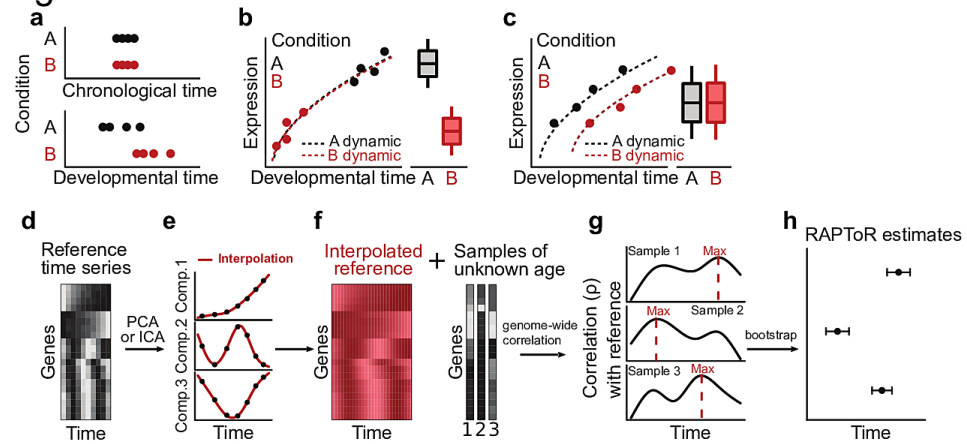

Figure 1 – Estimate age from transcriptome using RAPToR

a-c Cartoons showing **a**, a condition that impacts developmental speed; the effect of a perturbation of interest confounded (**b**) or masked (**c**) by hidden variation in developmental time.
d-f, RAPToR staging exploits existing reference time-series expression data (**d**). This data is first decomposed into principal / independent components which are interpolated with respect to time (**e**). Interpolated reference is then reconstructed by matrix product of interpolated components and gene loadings (**f**).
g-h, For each sample, a correlation profile is built by computing genome-wide Spearman correlation with every time point of the reference (**g**). The reference time with maximal correlation becomes the estimate, and bootstrapping on random gene subsets defines a confidence interval with median absolute deviation of bootstrap estimates to the estimate acquired on all genes (**h**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2

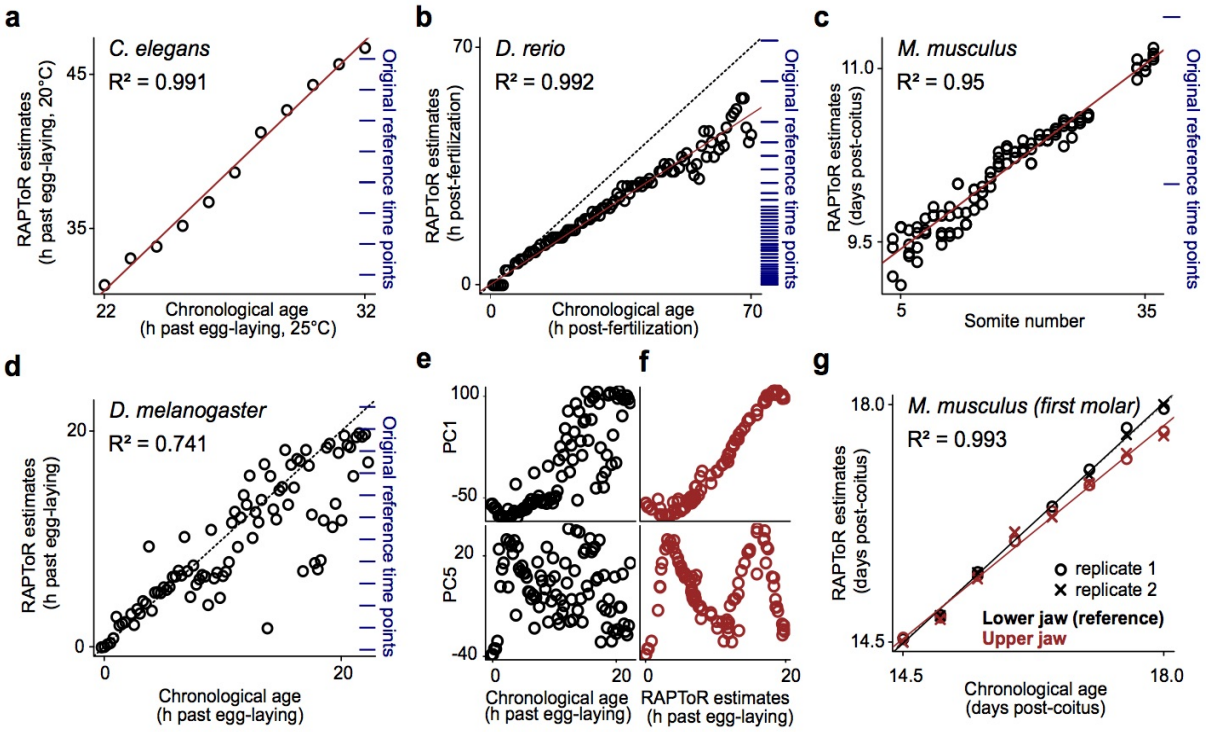

Figure 2 – Evaluating RAPToR's performance by staging time series of model organisms

a,b, Chronological age vs. RAPToR estimates of *C. elegans* late-larval samples[26] (linear model is y = -4.7 + 1.6x) (**a**), and *D. rerio* embryo samples[27] (linear model is y = 0.7x) (**b**)
c, Somite-number vs. RAPToR estimates of *M. musculus* embryo samples[29] (linear model is y = 9.2 + 0.05x)
d, Chronological age vs. RAPToR estimates of *D. melanogaster* embryo samples[27].
e-f, Selected Principal Components of the data staged in (**d**), plotted in black along chronological age (**e**) and in red along RAPToR estimates (**f**) .
g, Chronological age vs. RAPToR estimates of dissected samples of upper jaw first molars from *M. musculus* embryos staged using the lower jaw samples as reference [30,31].
a-d, Original time points of the reference are shown to the right of plots in blue.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3

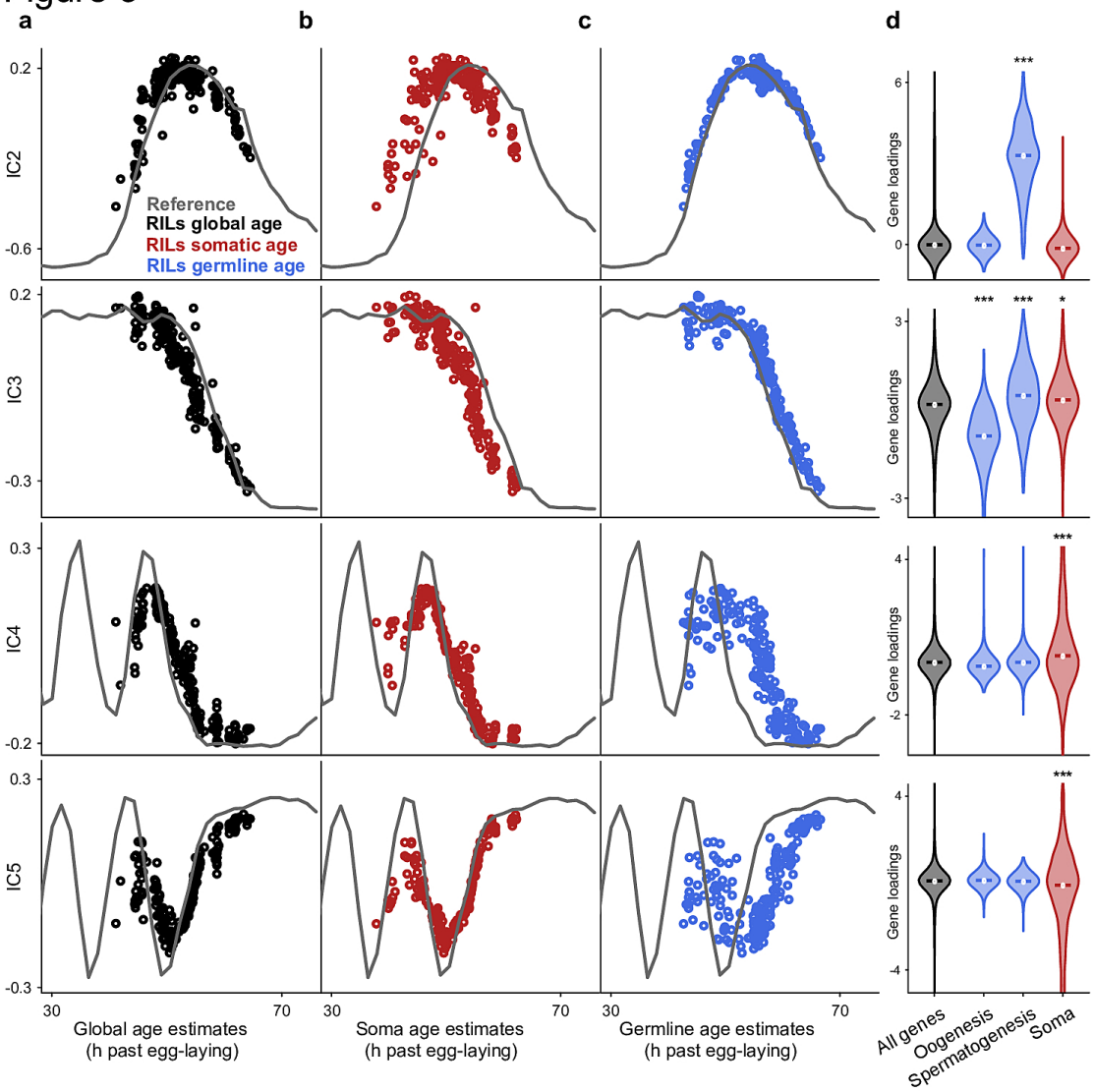

Figure 3 – Tissue-specific staging

a-c, Independent components 2-5 from ICA on joint *C. elegans* recombinant inbred lines[11]. (points) and reference data they were staged on (grey line). Samples are plotted (**a**) in black along "global age", (**b**) in red along "soma age", and (**c**) in blue along "germline age".
d, Gene loading enrichment of ICA components 2-5 for soma and germline categories.
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4

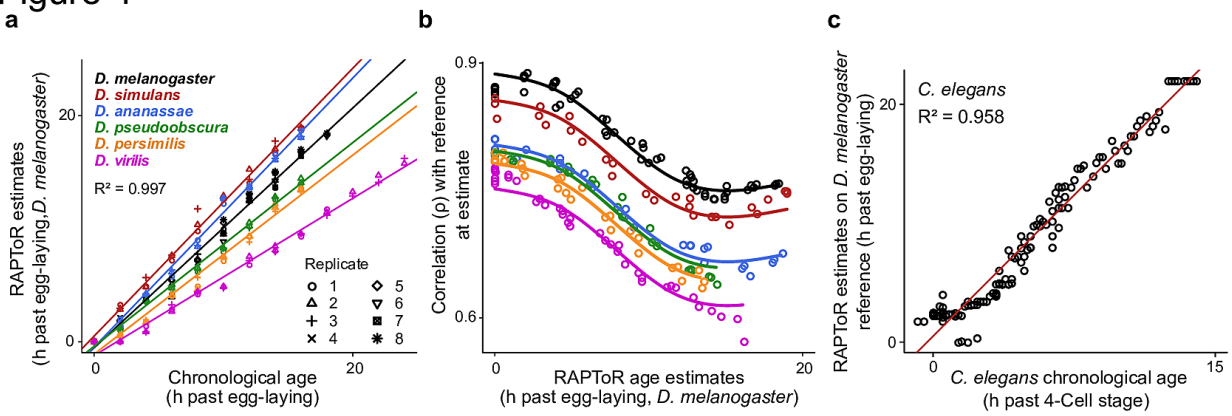

Figure 4 – Staging samples cross-species

a, Chronological age vs. RAPToR estimates for time series of embryo development of 6 *Drosophila* species[33] staged on a *D. melanogaster* reference (see also Sup. Fig. 11).
b, Spearman correlation between samples from (**a**) and the reference at age estimate, along RAPToR estimates.
c, Chronological age vs. RAPToR estimates for *C. elegans* embryo samples[27], staged on a *D. melanogaster* reference using orthologs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5

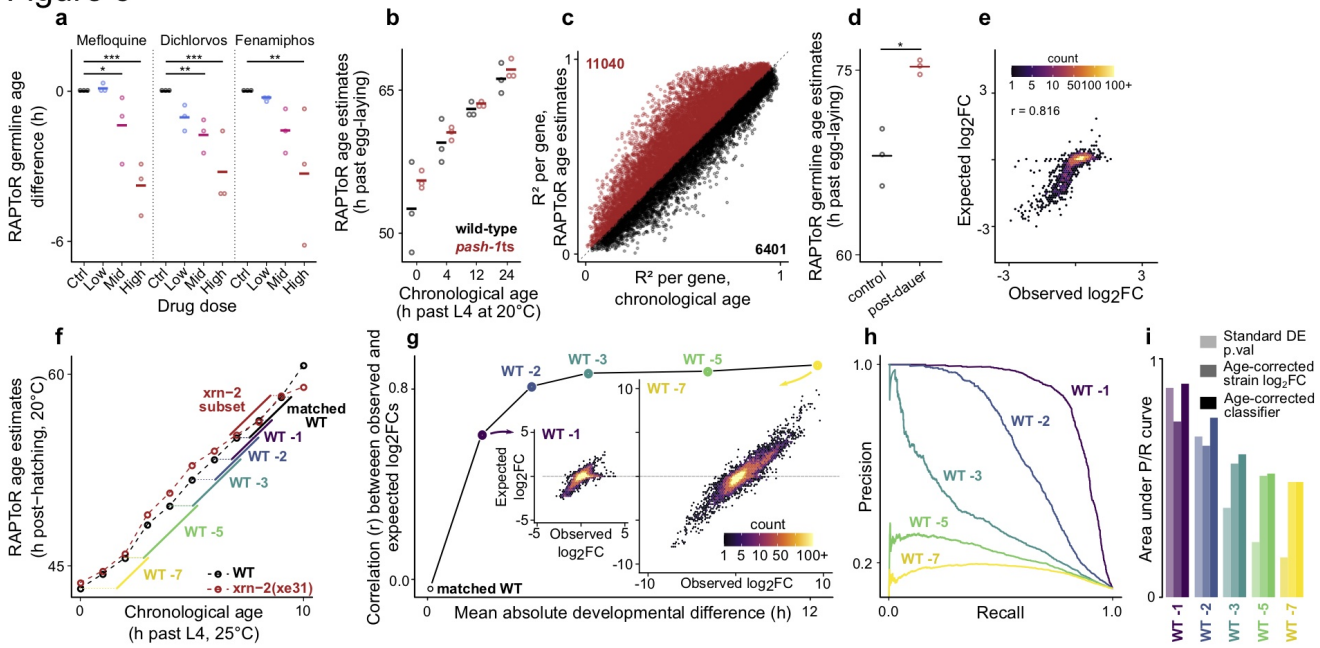

Figure 5 – Quantifying and correcting for developmental effects using RAPToR age estimates

a, Effect of increasing drugs dose exposure[35] on RAPToR estimates of *C. elegans* germline age (normalized by subtracting control age within groups, see Sup. Note 2, Sup. Fig. 14).

b, RAPToR age estimates vs. reported chronological age highlight large developmental spread within time points of *C. elegans* WT and *pash-1*ts time series[36] (see Sup. Note 2, Sup. Fig. 15).

c, R² per gene of identical models with chronological age, or RAPToR age estimates. Genes and gene counts above and below the dashed line (x=y) are indicated in red and black respectively.

d, Germline age estimates of control and post-dauer *C. elegans* adults[37].

e, Germline genes logFCs between control and post-dauer from (**d**) compared to logFCs expected from developmental time difference only (see Sup. Note 2, Sup. Fig. 18).

f, Chronological age vs. RAPToR estimates of a time-course of *C. elegans* WT and *xrn-2* late larval development[38]. Sample subsets defining a gold standard of truly DE genes and shifted WT sets used in subsequent panels are color-coded.

g, Correlation of observed logFCs and expected developmental logFCs computed from the interpolated reference between the *xrn-2* subset and increasingly shifted WT sets from (**f**). (see Sup. Note 2).

h, Precision-Recall curves showing the performance of a standard DE model p-value for each shifted WT subset in detecting gold-standard DE genes.

i, Area under precision-recall curves (AUPRC) of standard DE model p-value (**h**) or of the age-corrected classifier for each shifted WT subset in detecting gold-standard DE genes (see Sup. Note 2).

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

WT, wild-type. logFC, log2 fold-change. DE, Differentially Expressed or Differential Expression. FDR, false discovery rate.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

670

1. Francesconi, M. & Lehner, B. Reconstructing and analysing cellular states, space and time from gene expression profiles of many cells and single cells. *Mol. Biosyst.* **11**, 2690–2698 (2015).
2. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
3. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, e161 (2007).
4. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
5. Gómez-Orte, E. *et al.* Effect of the diet type and temperature on the C. elegans transcriptome. *Oncotarget* **9**, 9556 (2018).
6. MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J. & Walhout, A. J. Diet-induced developmental acceleration independent of TOR and insulin in C. elegans. *Cell* **153**, 240–252 (2013).
7. Ludewig, A. H. *et al.* Larval crowding accelerates C. elegans development and reduces lifespan. *PLoS Genet.* **13**, e1006717 (2017).
8. Kuntz, S. G. & Eisen, M. B. Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. *PLoS Genet.* **10**, (2014).
9. Perez, M. F., Francesconi, M., Hidalgo-Carcedo, C. & Lehner, B. Maternal age generates phenotypic variation in Caenorhabditis elegans. *Nature* **552**, 106 (2017).
10. Snoek, L. B. *et al.* A rapid and massive gene expression shift marking adolescent transition in C. elegans. *Sci. Rep.* **4**, 3912 (2014).
11. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at Linked Sites Shapes Heritable Phenotypic Variation in C. elegans. *Science* **330**, 372 (2010).
12. Hibbs, M. A. *et al.* Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* **23**, 2692–2699 (2007).
13. Francesconi, M. & Lehner, B. The effects of genetic variation on gene expression dynamics during development. *Nature* **505**, 208 (2014).
14. Lu, P., Nakorchevskiy, A. & Marcotte, E. M. Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations. *Proc. Natl. Acad. Sci.* **100**, 10370–10375 (2003).
15. O'Duibhir, E. *et al.* Cell cycle population effects in perturbation studies. *Mol. Syst. Biol.* **10**, 732 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

17. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845 (2016).

18. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381 (2014).

19. Anavy, L. *et al.* BLIND ordering of large-scale transcriptomic developmental timecourses. *Development* **141**, 1161–1166 (2014).

20. Kim, D. hyun, Grün, D. & van Oudenaarden, A. Dampening of expression oscillations by synchronous regulation of a microRNA and its target. *Nat. Genet.* **45**, 1337–1344 (2013).

21. Meeuse, M. W. *et al.* Developmental function and state transitions of a gene expression oscillator in Caenorhabditis elegans. *Mol. Syst. Biol.* **16**, e9498 (2020).

22. Reinke, V., San Gil, I., Ward, S. & Kazmer, K. Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. *Development* **131**, 311–323 (2004).

23. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).

24. Xue, L. *et al.* Global expression profiling reveals genetic programs underlying the developmental divergence between mouse and human embryogenesis. *BMC Genomics* **14**, 568 (2013).

25. Graveley, B. R. *et al.* The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473 (2011).

26. Hendriks, G.-J., Gaidatzis, D., Aeschimann, F. & Großhans, H. Extensive oscillatory gene expression during C. elegans larval development. *Mol. Cell* **53**, 380–392 (2014).

27. Levin, M. *et al.* The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637 (2016).

28. Rauwerda, H. *et al.* Transcriptome dynamics in early zebrafish embryogenesis determined by high-resolution time course analysis of 180 successive, individual zebrafish embryos. *BMC Genomics* **18**, 1–15 (2017).

29. Collins, J. E. *et al.* Common and distinct transcriptional signatures of mammalian embryonic lethality. *Nat. Commun.* **10**, 1–16 (2019).

30. Pantalacci, S. *et al.* Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology. *Genome Biol.* **18**, 1–25 (2017).

31. Sémon, M. *et al.* Comparison of developmental genome expression in rodent molars reveals extensive developmental system drift. *bioRxiv* (2020) doi:10.1101/2020.04.22.043422.

32. Perez, M. F. *et al.* Neuronal perception of the social environment generates an inherited memory that controls the development and generation time of C. elegans. *Curr. Biol.* **31**, (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.07.459270; this version posted September 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).

34. Li, J. J., Huang, H., Bickel, P. J. & Brenner, S. E. Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res.* **24**, 1086–1101 (2014).

35. Lewis, J. A., Szilagyi, M., Gehman, E., Dennis, W. E. & Jackson, D. A. Distinct patterns of gene and protein expression elicited by organophosphorus pesticides in Caenorhabditis elegans. *BMC Genomics* **10**, 202 (2009).

36. Lehrbach, N. J. *et al.* Post-developmental microRNA expression is required for normal physiology, and regulates aging in parallel to insulin/IGF-1 signaling in C. elegans. *Rna* **18**, 2220–2235 (2012).

37. Hall, S. E., Beverly, M., Russ, C., Nusbaum, C. & Sengupta, P. A cellular memory of developmental history generates phenotypic diversity in C. elegans. *Curr. Biol.* **20**, 149–155 (2010).

38. Miki, T. S., Carl, S. H. & Großhans, H. Two distinct transcription termination modes dictated by promoters. *Genes Dev.* **31**, 1870–1879 (2017).

References (Methods)

39. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

40. Borchani, H., Varando, G., Bielza, C. & Larrañaga, P. A survey on multi-output regression. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **5**, 216–233 (2015).

41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

42. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

43. Hill, M. A. Mouse Stages. *Embryology*.

44. Michaelson, J. J., Alberts, R., Schughart, K. & Beyer, A. Data-driven assessment of eQTL mapping methods. *BMC Genomics* **11**, 502 (2010).

45. Raudvere, U. *et al.* g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).