Supplementary note 1

Reference interpolation

For reference interpolation, we can take advantage of the redundancy of gene expression. We use either Principal or Independent Component Analysis (PCA, ICA) to decompose the expression matrix into components (sample scores, or eigengenes) that summarize the gene expression dynamics and gene loadings, that are genes contributions to each dynamic [46,47]. We interpolate the components with respect to time (Fig 1e), and reconstruct the full interpolated gene expression by the matrix product of interpolated components and the gene loadings (Fig 1f, methods). In this way, we simplify model building and validation across thousands of genes to a few components. We select a number of components by a threshold on cumulative variance explained (see methods), however, staging results are robust to the variation number of components used (Sup. Fig 1a, 1b, see below).

In order to validate an interpolated reference for staging with RAPToR, we first stage the original reference data on its interpolated version. We fit a linear model predicting RAPToR age estimates with chronological age expect a near-perfect match, with adjusted $R^2 > 0.99$, and non-significant intercepts. This was the case for all references presented throughout the study, aside from “Cel_YA_2” which is built from old data [22] (2004) and only had an R^2 of 0.901. Then, when possible we stage independent time course experiments and expect very good fits of linear models ($R^2 \geq 0.9$, as shown in Fig. 2).

Interpolation errors or bias are more frequent at the edges of a time series (e.g. splines are known to behave erratically at edges). In addition, when the real age of a sample is outside the reference range, it will likely match a time point close to the edge, making estimates close to the edges less reliable. Therefore, when age estimates are found near the edges, we suggest using another overlapping reference to confirm the estimates.

Evaluating RAPToR performance

Reference interpolation effectively increases estimates temporal resolution

The worm[26], fly[27], and mouse[29] time-series we staged (Fig. 2a, 2d, 2c) had 3, 8, and over 18 times the resolution of the reference data prior to interpolation respectively (see methods). This clearly shows that

interpolation allowed us to effectively increase the resolution of our reference time-series and consequently the resolution and accuracy and of the age estimates.

To further test this, we staged another zebrafish time-course consisting of 180 embryo samples spread in a very short developmental window around gastrulation [28]. Despite having over 40 times the initial resolution of the reference, the data is accurately staged by RAPToR, not only matching the established ranking [28] (ρ=0.99, Sup. Fig 3a) and expression dynamics (Sup. Fig 3b), but with the advantage of having absolute times that are comparable to any estimate obtained with the same reference.

Effect of gene set size and interpolation parameters on age estimates

We tested RAPToR robustness to changes in the gene set size by staging an independent time series[26] of C. elegans larval development on the reference[20] we built (as in Fig. 2a). With random gene sets of 4000 genes, estimates have a median absolute deviation under 10 minutes from full gene set estimates (18718 genes); with sets of 1000, under 20 minutes (Sup. Fig 4a, 4b). With smaller sets (<1000), estimates are unreliable likely because repeated expression patterns – such as oscillations or pulses – create multiple correlation peaks with the reference (Sup Fig 4c). For example, the repeated molts of *C. elegans* larval development generate oscillations in the correlation profiles of staged samples (Sup Fig 5). To address this, we implemented the possibility to include a prior for estimates (see methods). With the prior, no estimates get misplaced to other correlation peaks, even for sets of a few hundred genes (Sup. Fig 4c, 4d).

The default confidence intervals of RAPToR estimates for the *C. elegans* time series, built from bootstrap estimates with 6239 genes (see methods), are extremely small – between 5 and 20 minutes (Sup. Fig 6a). Staging zebrafish embryos[27] on a reference[23] with lower gene overlap (8662 genes) results in larger confidence intervals built from bootstrap estimates with 2887 genes – on average, slightly over 2h across the full time series, and under 50 minutes for samples under 30h of development (Sup. Fig 6b).

The combination of interpolation accuracy and estimate precision for small gene sets means that even time-series transcriptomic data with low sampling rate or gene coverage can be exploited to build interpolated references.

We also tested whether reference-building is robust to parameter changes. As expected, selecting more components leads to a decrease in prediction error of the model (Sup. Fig 7a, 7b), and a slight increase in the

correlation between staged samples and reference at estimate (Sup. Fig 7c, 7d). However, the age estimates were mostly unperturbed by the changes in number of components and were also robust to choosing PCA or ICA for interpolation (Sup. Table 2).

Inferring developmental speed factors

Beside the expected 1.5 fold increase in developmental speed due to temperature we observed in *C. elegans* (Fig. 2a), we also observe a difference between chronological and estimated times in the independent zebrafish developmental time series[27] we staged on the zebrafish reference[23] determining a developmental speed factor of 0.7. We suspect this speed factor is also due a temperature difference with the reference, as growth speed scales with temperature also in zebrafish. While the reference data embryos developed at 28.5°C[23], we were unable to confirm at which temperature the staged data embryos developed but we presume a lower one.

Quantitative Trait Loci analysis on soma-germline heterochrony

In our QTL analysis of soma-germline heterochrony in *C. elegans* Recombinant Inbred Lines[11], Random Forest prediction of the trait was poor and non-significantly correlated with the trait ($r = 0.12$, $p = 0.09$) and we found no significant marker, at FDR of 0.5. Removing the batch covariate from the analysis results in even poorer predictions ($r = 0.08$, $p = 0.2$), further pointing in the direction of uncontrolled environmental factors mainly driving heterochrony.

Staging samples on references of a different specie

When we first staged a *C. elegans* embryo development time course [27] on the *Drosophila* reference [25], we noted two clean breaks in the age estimates, possibly due to heterochrony of developmental processes between the 2 species ($R^2 = 0.938$, Sup. Fig. 13a). Staging was notably improved after rebuilding the reference with only 2 components which have broad dynamics ($R^2 = 0.958$, Fig 4c), further suggesting that sharper expression dynamics (like pulses or oscillations) diminished staging performance.

To explore the reason behind successful staging, we analyzed the 319 genes (10%) with highest correlation between *C. elegans* and *D. melanogaster* during embryogenesis (see methods), which suffice to stage the embryos well despite their small number ($R^2 = 0.87$, $\rho = 0.971$, Sup. Fig. 13b). We found this gene set clusters into an ascending gene expression signature of muscle development (cluster 1, Sup. Fig 13c, Sup.

Table 4), and two decreasing signatures of cell proliferation split between DNA replication (cluster 2, Sup. Fig 13c, Sup. Table 5) and splicing (cluster 3, Sup. Fig 13c, Sup. Table 6) respectively. Expression dynamics match well between fly and worm embryogenesis (Sup. Fig 13c), consistent with previous work showing widespread conservation of decrease in cell proliferation during embryogenesis [27].

Supplementary note 2

Exploiting inferred age in genome-wide expression studies

Inferring the impact of environmental or genetic perturbations on development

When staging C. elegans exposed to increasing concentrations of mefloquine, dichlorvos, and fenamiphos[35], we noted that beyond the germline developmental delay induced by all three drugs, dichlorvos also showed a significant and opposite effect of dose on somatic age. However, the scale of the effect is a fraction of the one observed on the germline (Sup. Fig 14b).

Exploiting developmental variation to increase power to detect differential expression

After comparing chronological age and RAPToR age estimates as predictors for *C. elegans* control and *pash-1* mutants profiled at 4 time points of late development[36] and finding better model fits (Sup. Fig 14a), we further tested whether random perturbations on age could induce a similar result. Thus, we generated age sets of similar deviation from chronological age than RAPToR estimates (see methods) and found that these consistently decreased model fits and DE gene detection, confirming that precise age estimates increase the power of the analysis (Sup. Fig 15b-d, methods).

We also found a curious batch effect on development: mutants are systematically older than controls in the first two replicates while it was the opposite in the third (Sup. Fig 15e).

Detecting and correcting expression changes caused by development using reference data

When samples are few and experimental groups have little or no overlap in development, the information available in the experiment is not sufficient to separate the effects of development from those of interest. To overcome this, we developed an approach using RAPToR interpolated references to quantify and correct for development in genome-wide expression data.

Quantifying developmental expression changes

To quantify the impact of development in DE analysis, we compare observed $\log_2$-fold changes (observed logFC) between the two groups (i.e. mutant and wt) with changes expected purely from developmental differences between groups (expected logFC) which we estimate comparing age-matched interpolated

reference profiles (Sup. Fig 17). We quantify development impact using Pearson correlation between observed and expected logFCs (or its square). We use Transcripts Per Million (TPMs) to compute the logFCs, as they are more comparable across samples and datasets.

Development can completely confound DE analysis leading to erroneous conclusions

Not accounting for confounding developmental variation in DE analysis can even lead to erroneous conclusions. To show this, we reanalyze a dataset comparing young adult *C. elegans* that developed through dauer state (post-dauer) to controls that did not[38].

The authors conclude that post-dauer animals have reduced spermatogenesis from a down-regulation of spermatogenesis-associated genes and an up-regulation of oogenesis-associated genes in their DE analysis. However, this could simply be explained by post-dauer samples being older than controls as *C. elegans* switch from spermatogenesis to oogenesis during development. Moreover increased brood size in post-dauer worms [37] would even suggest the opposite as sperm number limits brood size in *C. elegans*, this would suggest up- and not down-regulation of spermatogenesis genes in post-dauer animals.

To rigorously test if these expression changes are caused by development, we estimated the global and tissue-specific age of samples with our best-quality reference, and found post-dauer samples were systematically older. However, while germline age estimates were reliable, global and soma-specific staging put some samples at the edge of the reference, indicating development beyond reference bounds (Sup Fig. 16a). Thus, we validated the age estimates with an older and lower-quality reference spanning a few hours further and found the same divide in global, soma, and germline age between groups (Sup. Fig 16b).

RAPToR age estimates show the control samples are in late spermatogenesis, while the post-dauer samples are 5-10 hours older, fully switched to oogenesis (Fig 5c, Sup. Fig 16a). A DE analysis between the two conditions does recapitulate reported divide between spermatogenesis and oogenesis genes (Sup. Fig 16c). However, this is also fully recapitulated by the expected developmental changes (r = 0.82, Fig 5d, Sup. Fig 16d). Correlation also stands with all genes (r = 0.44, Sup. Fig 16e). Repeating the logFC comparison with the older reference yielded similar results (germline r = 0.74, all genes r = 0.41 Sup. Fig 16f, 16g).

This supports our hypothesis that the expression changes between groups, particularly in germline genes, is due to a difference in development between the samples, rather than a direct effect of the post-dauer condition.

Recovering direct perturbation effects using reference data

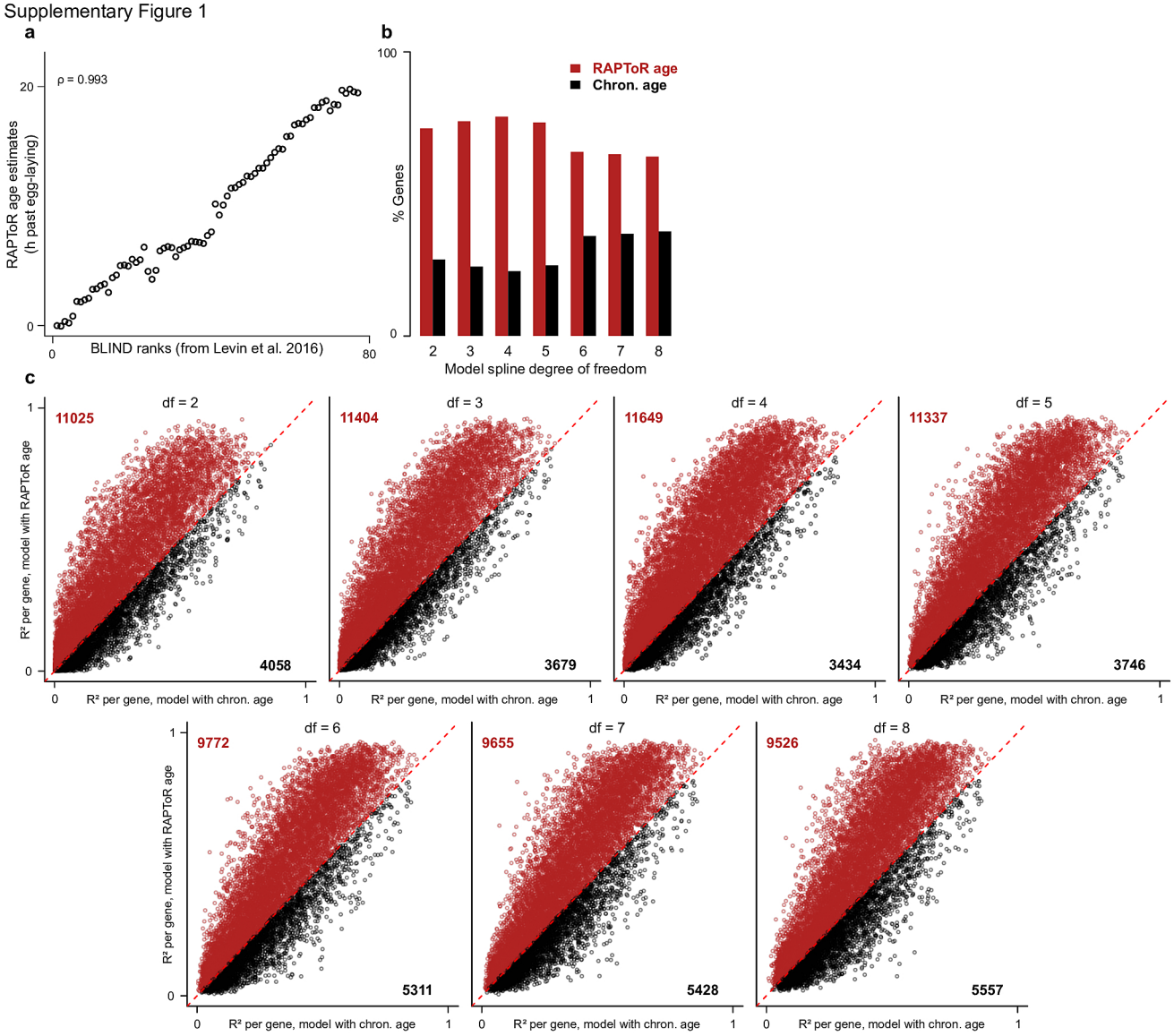
To recover the effect of a perturbation confounded by development we propose a model in which we include all reference expression profiles in a time window spanning the development of the selected samples (see methods) and model expression dynamics as a spline of developmental time, batch between reference and sample data, and the perturbation of interest. We first convert interpolated reference data from TPM to counts, assuming a fixed library size (see methods) to ensure compatibility with tools that require counts to calculate differential expression. Including reference data with artificially low dispersion invalidates models statistics such as p-values. We can however rely on the model coefficients (logFCs) estimated with the reference data to account for development. The perturbation (strain) logFC coefficients estimated by this reference integrated model are controlled for developmental changes shared by the samples and the reference (Sup. Fig 18a).

To evaluate how effectively our approach recovers really DE genes we exploit time series data of *C. elegans xrn-2* and WT late larval development sampled every hour at 25°C[38]. We first define a gold standard of truly DE genes by comparing three mutant and three WT samples with the best developmental match. Next we evaluate the effect of increasing developmental difference on the DE analysis by comparing the same three mutant samples with three increasingly mismatching WT samples (Fig. 5d, methods). We shifted WT samples back by 1, 2, 3, 5 and 7 time points (corresponding roughly to 1, 2, 3, 5, and 7 hours of development at 25°C). Correlation between expected and observed logFC quickly increases with increasing developmental shifts up to 0.9, meaning that around 80% of the variance in logFCs is explained by development at 7 hours of time shift. At the same time, the performance of a standard linear model p-value in identifying truly DE genes quickly drops (Fig. 5f) as more and more expression changes are due to development (Sup. Fig 18b).

Then we show that the strain logFC from the reference integrated model already performs better in detecting truly DE genes than the standard analysis p-value for developmental shifts of 3 or more hours (at 25°C, Sup Fig 18c). However, we propose an integrated predictor including the weighted mean between the standard analysis p-value and the strain logFC of the reference integrated where the optimal weight **w** is proportional to the variance of standard logFC explained by development (Sup. Fig 18d, 18e). At the optimal **w** (see methods, Sup. Fig 18e), our integrated predictor outperforms the standard analysis p-value for all time-shifts considered, with larger time-shifts showing the strongest improvements (Sup. Fig 18c, 18d). For the largest shift (WT -7), **w** = 1 meaning that no information from the standard DE p-value is used to get the best results.

As no gold-standard is usually available to guide the choice of **w**, we explored the relationship between the optimal **w** and the correlation of logFCs with the reference. Sampling more WT subsets including non-contiguous sets (see methods, Sup. Table 8) reveals a tight relationship between optimal ***w*** and the correlation between of observed and expected logFCs (Sup. Fig 18f, 18g) which can therefore suggests the appropriate value of ***w***.

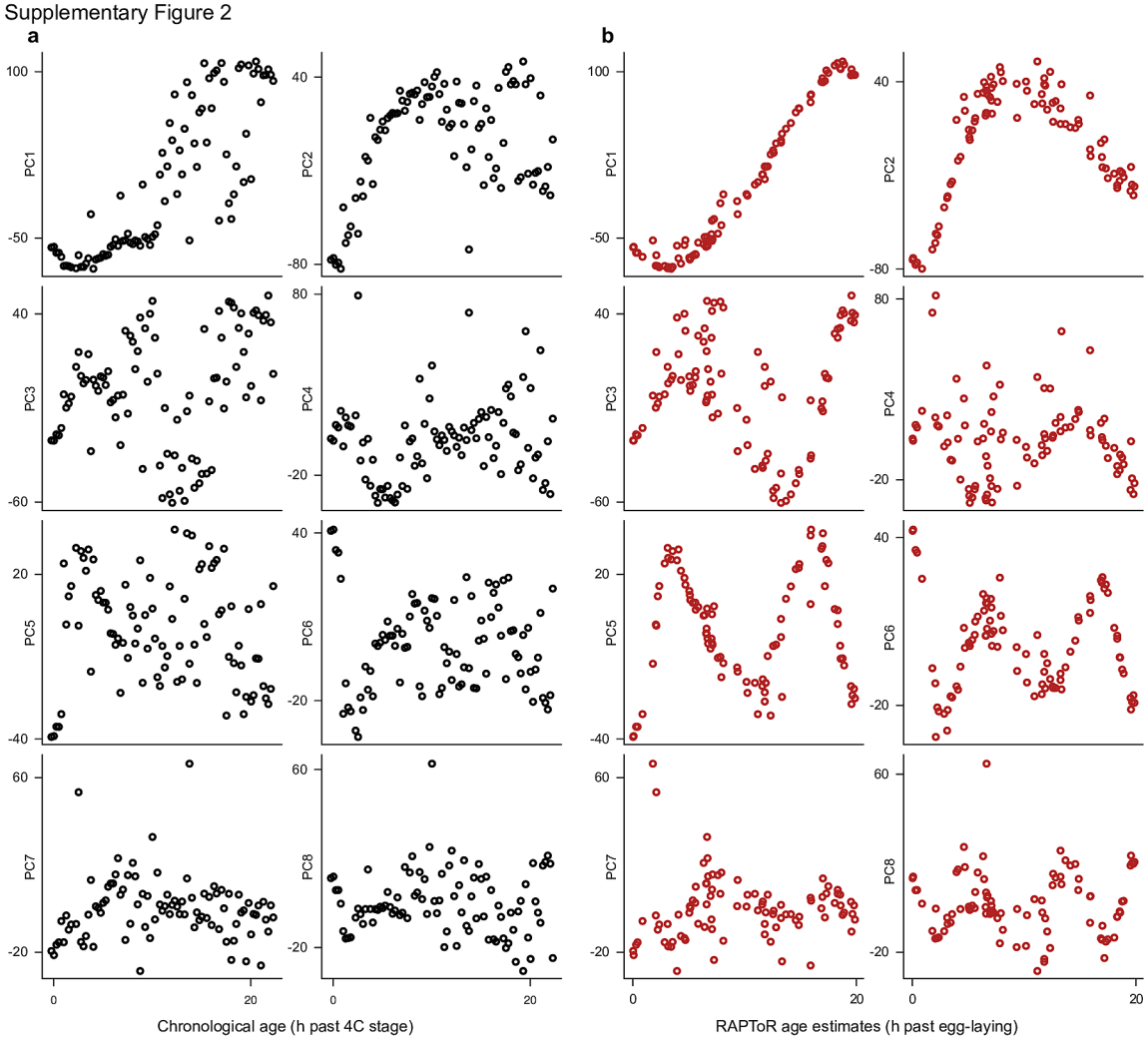
Supplementary Figure 1



Supplementary Figure 1 – RAPToR estimates fit gene expression data better than chronological age (related to Figure 2).

a, RAPToR estimates of *D. melanogaster* embryo samples [27] staged on a reference built from Graveley et al. data [25] plotted against established BLIND ranks[27]. Spearman's ρ is 0.993.
b, Percentage of genes better fitted by either RAPToR estimates or chronological age modeled with splines using from 2 to 8 degrees of freedom in otherwise identical models.
c, R^2 of models from (**b**) Gene count in each half of the plot is indicated in the top-left and bottom-right.

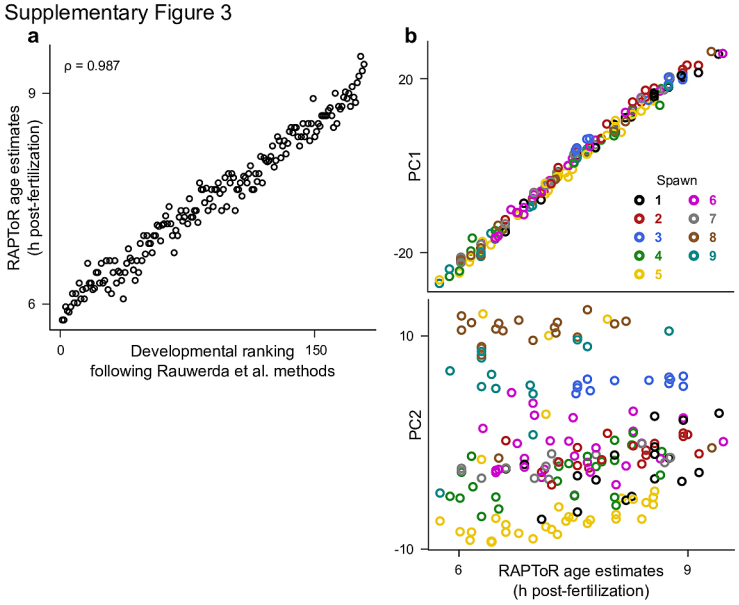
Supplementary Figure 2



Supplementary Figure 2 (related to Figure 2). RAPToR age estimates recover expression dynamics along development

a,**b**, Principal components of *D. melanogaster* embryo development [27] plotted along chronological age (**a**), and RAPToR age estimates (**b**) (as in Figure 2 d-f).

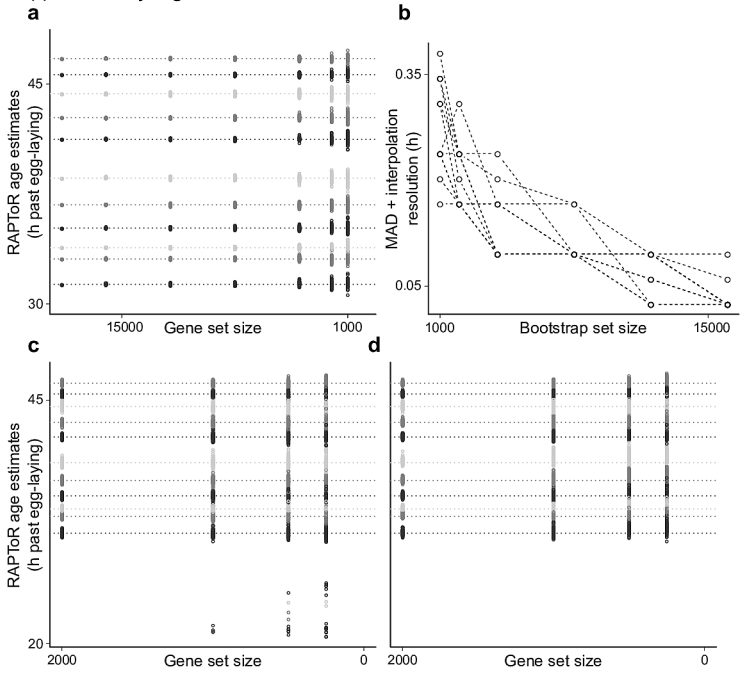
Supplementary Figure 3



Supplementary Figure 3 – Reference interpolation allows RAPToR estimates at high resolution

a, RAPToR estimates of a zebrafish embryonic time series from 9 spawns [28] staged on a reference built from Domazet et al. data[23] plotted against original developmental ranks[28] **b**, First 2 principal components of the zebrafish time series from plotted against RAPToR age estimates. Spawns are color-coded.

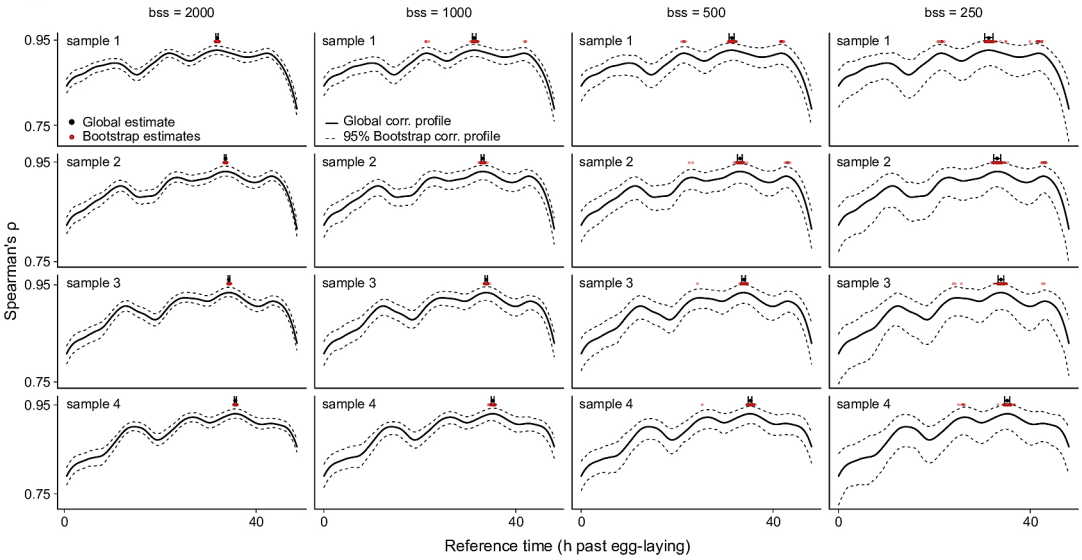
Supplementary Figure 4



Supplementary Figure 4 – Effect of gene set size on RAPToR estimates (related to Figure 2a).

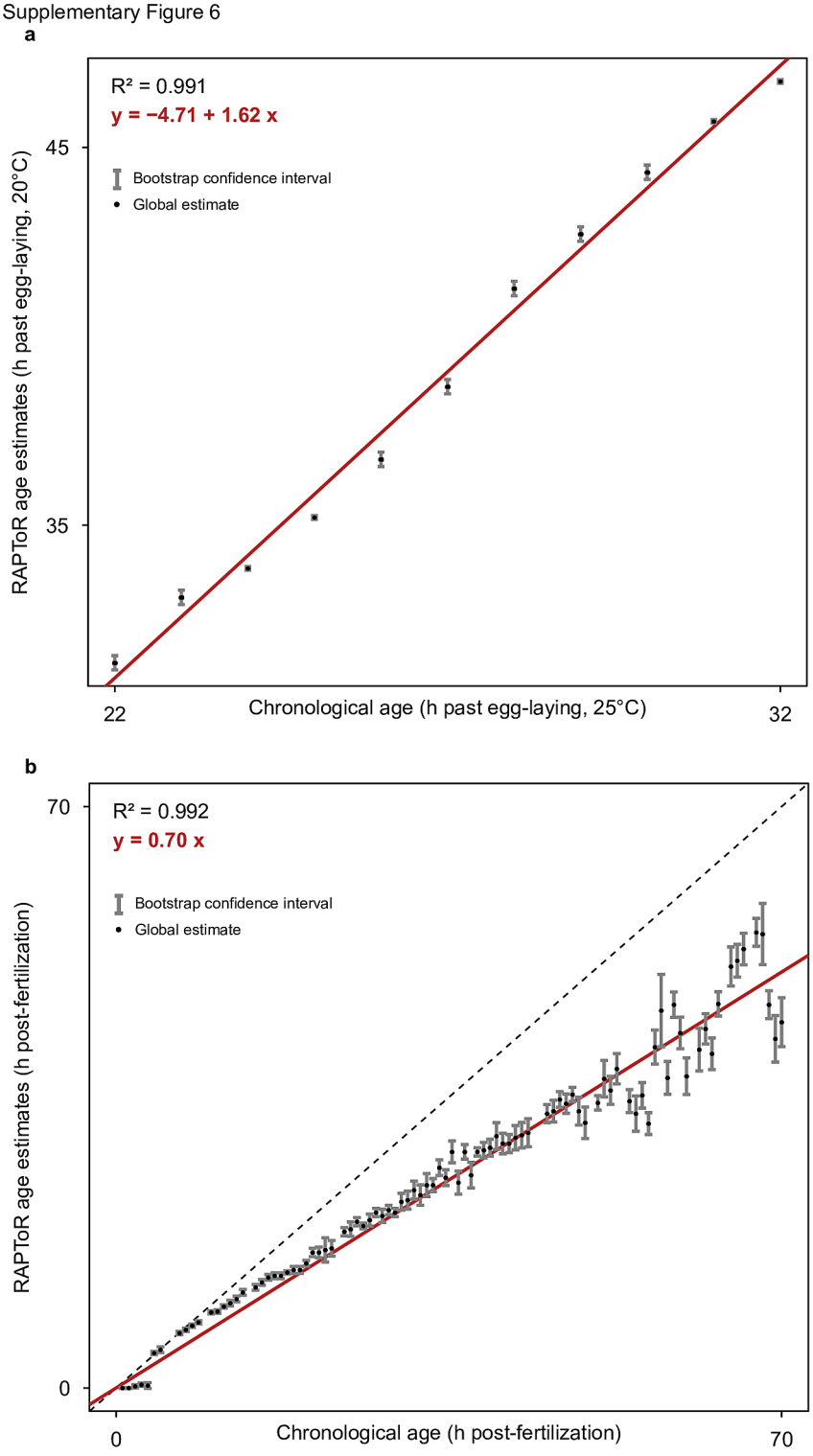
a, Effect of gene set size on RAPToR estimate of a *C. elegans* larval development time series [26] staged on a reference built from Kim et al.[20] data. Leftmost points and dashed lines indicate estimates on all genes (18718). **b,** Median Absolute Deviation of bootstrap estimates from global estimate + interpolation resolution (i.e. half a confidence interval) by bootstrap set size for 50 bootstrap estimates.
c,d, Effect of small gene sets on RAPToR estimates without prior (**c**), or with prior (**d**). Some samples are staged to a previous molt (see also Sup. Fig 5) when the gene set is small, an issue solved including a prior. Horizontal dashed lines indicate estimates of the samples using the full available gene set (18718).

Supplementary Figure 5



Supplementary Figure 5 – Effect of bootstrap (or gene) set size on RAPToR correlation profiles

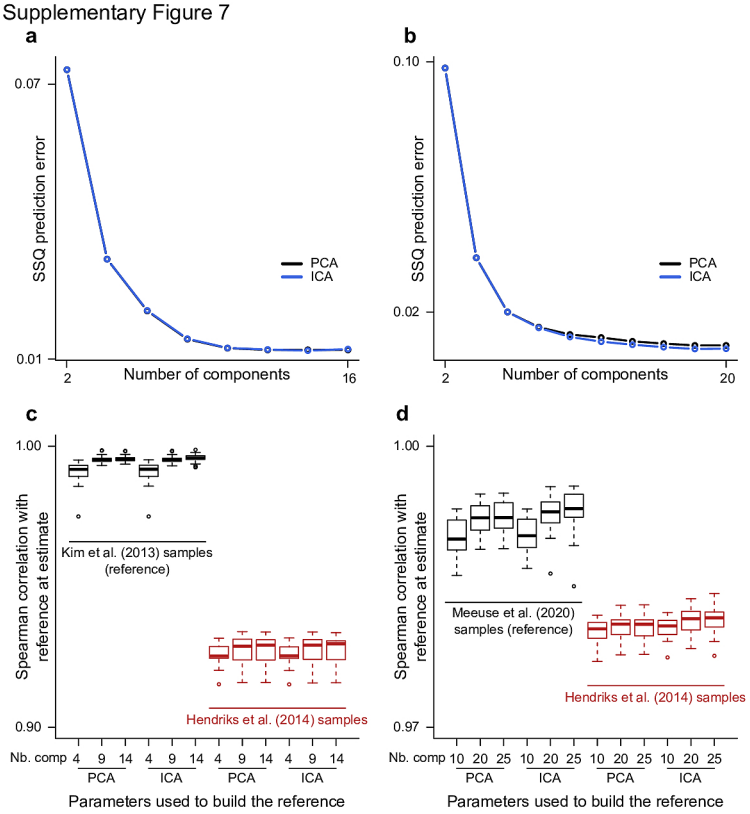
Correlation profiles of the first 4 samples of the Hendriks et al. [26] time course staged on the reference built from Kim et al. samples [20]. Bootstrap gene set size (bss) was set to 2 000, 1 000, 500 or 250, with 100 bootstraps. With small gene sets, samples can be staged to other points in the transcriptomic landscape with similar expression (e.g, the four maxima in the profiles shown here correspond to the successive larval molts of *C. elegans*).
Global and bootstrap estimates, as well as the confidence interval, are shown above each profile. Dotted lines around the global correlation profiles correspond to 0.025 and 0.975 quantiles of the bootstrap correlation profiles. As expected, this interval gets larger for smaller bootstrap gene set sizes.



Supplementary Figure 6 – Minute-scale precision staging of development

a,b, Chronological age vs. RAPToR age estimates and their confidence intervals of *C. elegans* late larval development[26] (**a**), and *D. rerio* embryo development[27] (**b**) staged on appropriate references[20,23] (as in Fig 2a, 2b).

Supplementary Figure 7

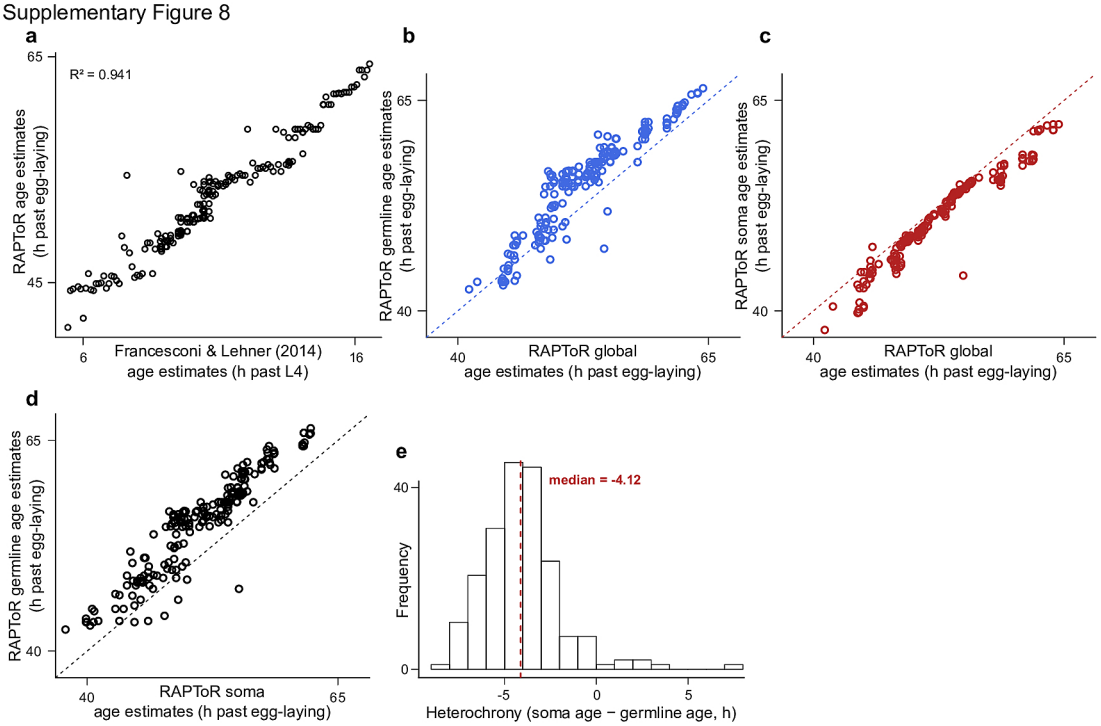


Supplementary Figure 7 – Robustness of reference-building to parameter change

a,b, Sum Squared prediction error of gene expression interpolation at known time points by number of components and dimension-reduction method used for interpolation in the Kim et al. [20] (**a**), and Meeuse et al. [21] (**b**) references of *C. elegans* larval development.
c,d, Spearman correlation between interpolated reference and (non-interpolated) reference and independent time series[26] samples at their age estimates when varying dimension reduction method and component number used for reference-building, using Kim et al. samples (**c**) or Meeuse et al. (**d**) samples to build a reference (see also Sup. Table 2).

In **c**, each box is n=26 for Kim et al. samples and n=12 for Hendriks et al. [26] samples; in **d**, each box is n=44 for Meeuse et al. samples and n=16 for Hendriks et al. samples.

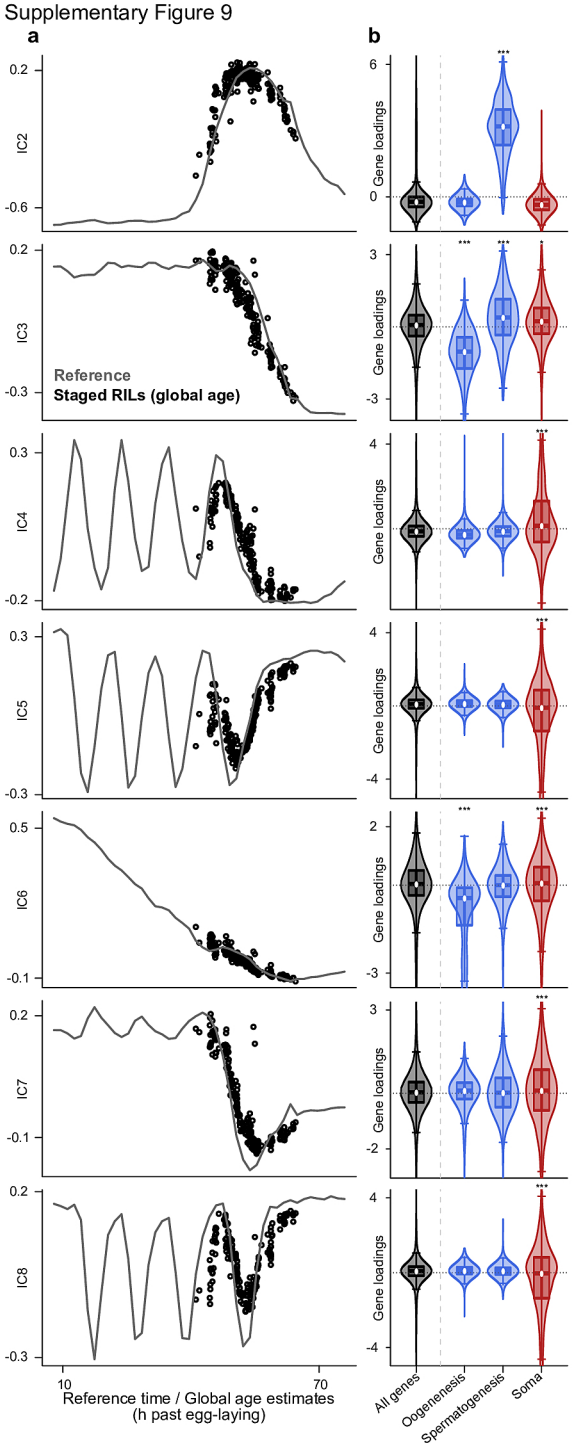
Supplementary Figure 8



Supplementary Figure 8 – Tissue-specific staging yields soma and germline ages

a, RAPToR estimates of *C. elegans* Recombinant Inbred Lines (RILs) [11] staged on the larval to young-adult reference built from Meeuse et al. [21] vs Francesconi & Lehner [13] estimates.
b-d, Comparison of RAPToR global age estimates vs. germline age estimates (**b**), global age estimates vs. soma age estimates (**c**), and soma age estimates vs. germline age estimates (**d**).
e, Distribution of soma-germline heterochrony.

Supplementary Figure 9

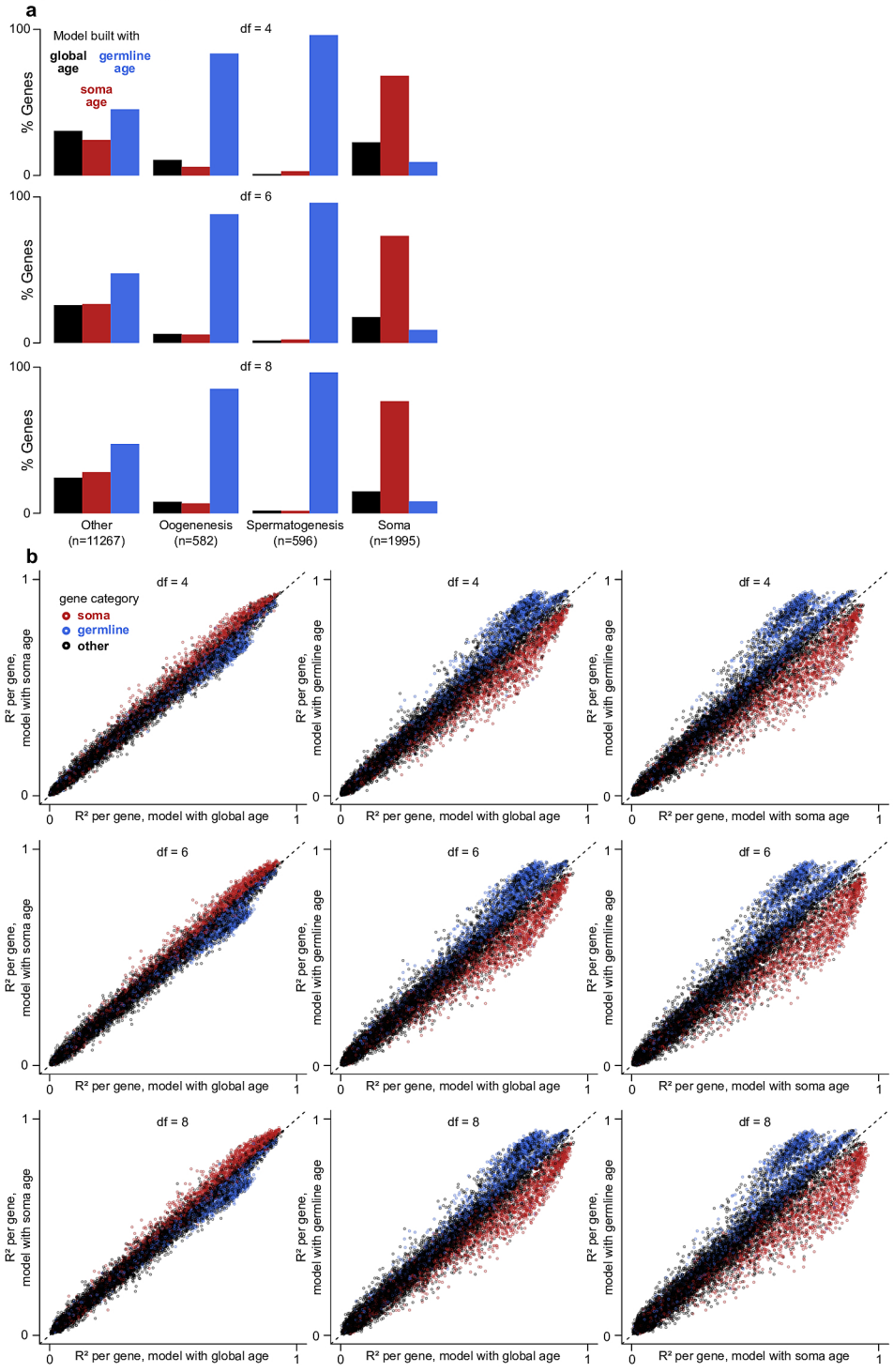


Supplementary Figure 9 – A delayed germline and an advanced soma

a, ICA components from ICA on *C. elegans* Recombinant Inbred Lines (RILs)[11] joined to the (non interpolated) reference data [21] plotted along chronological age and RAPToR global estimates for the reference (grey) and RILs (black) respectively.
b, Enrichment of gene loadings on ICA components for germline and soma categories.

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

Supplementary Figure 10

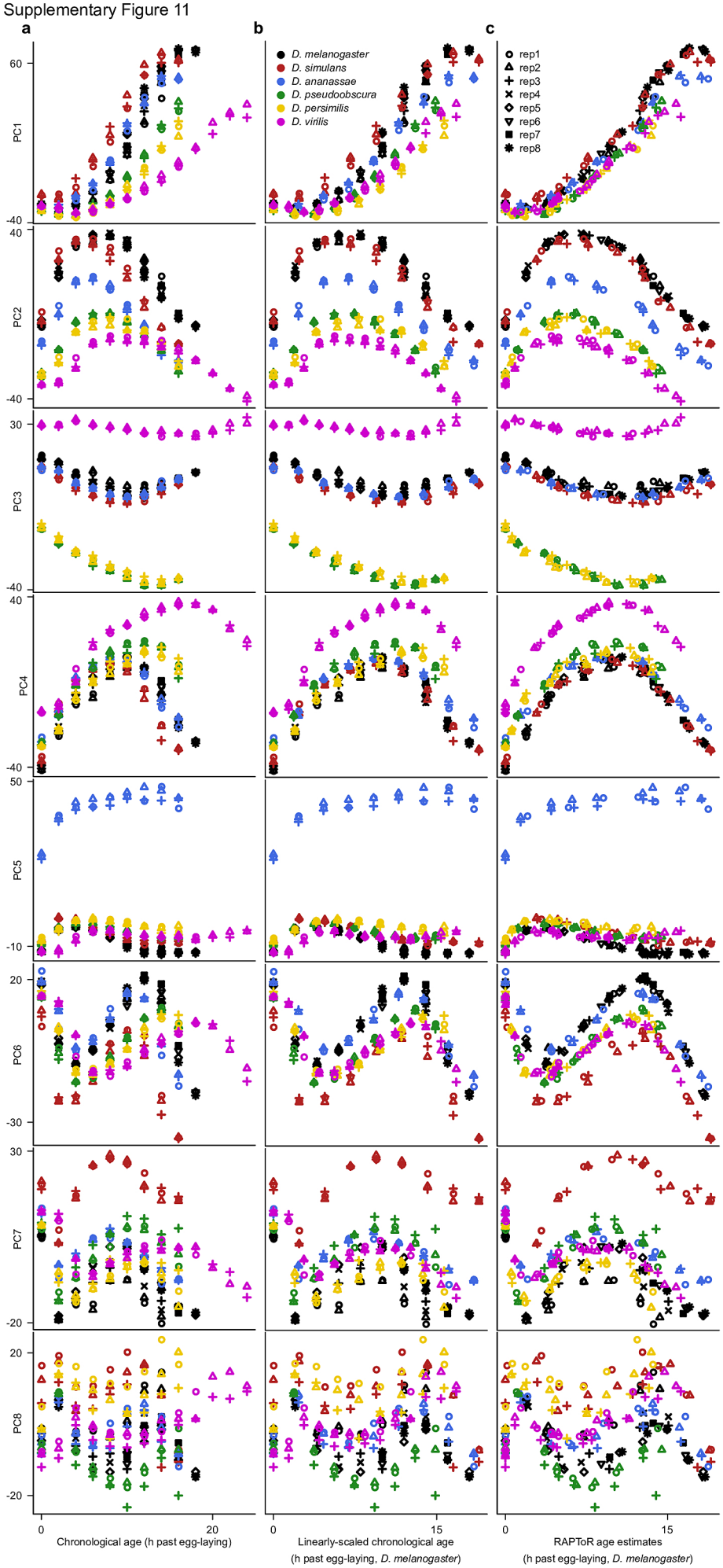


Supplementary Figure 10 – Soma-germline heterochrony among *C. elegans* recombinant lines.

Lines [11] are staged on the larval to young-adult reference built from Meeuse et al. samples[21].

a, Percentage of genes better fitted by either RAPToR global, soma, or germline age estimates, modeled with splines with 4, 6 or 8 degrees of freedom in otherwise identical models. Genes are classified into spermatogenesis oogenesis, somatic or other (see methods).
b, R^2 per gene of models with global, soma or germline age estimates as predictors, for 4,6,and 8 spline degree of freedom values.

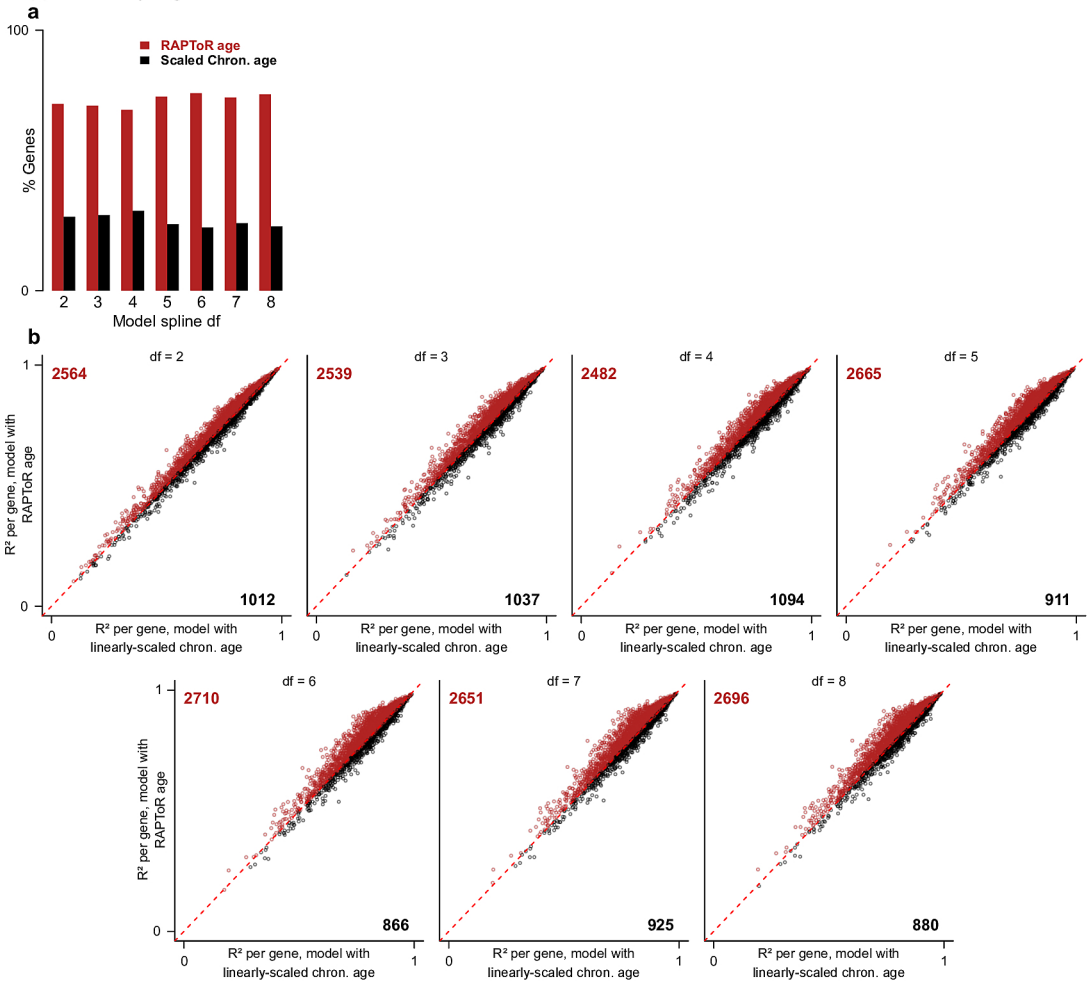
Supplementary Figure 11



Supplementary Figure 11 – RAPToR age estimates synchronize expression dynamics across species

a-c, Principal components of *Drosophila* embryogenesis in 6 species [33] plotted along chronological age (**a**), linearly-scaled chronological age[33] (**b**), and RAPToR age estimates on a *D. melanogaster* reference [25](**c**).

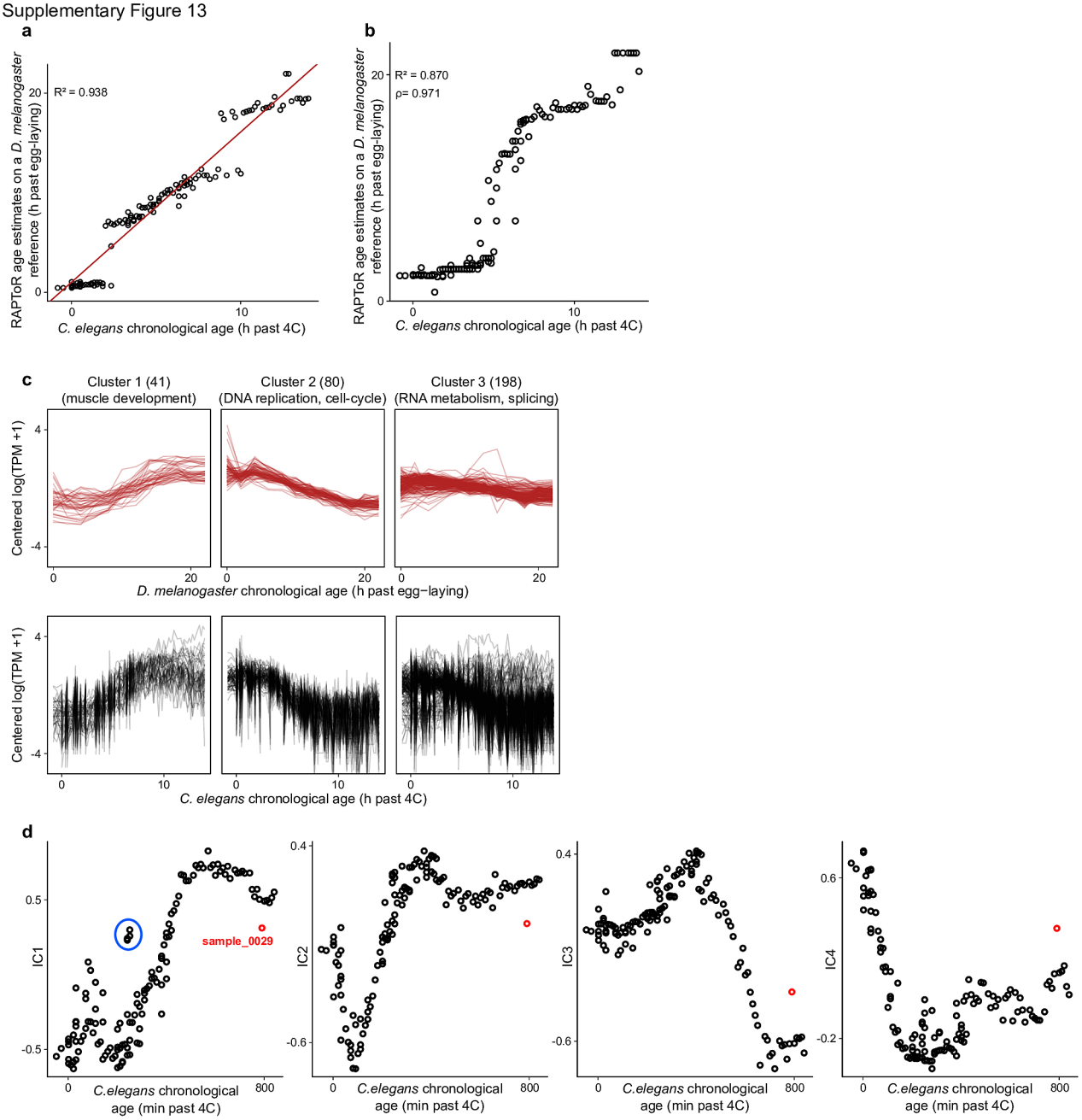
Supplementary Figure 12



Supplementary Figure 12 – RAPToR age estimates improve model fits over linear age scaling
*Drosophila e*mbryo samples from Kalinka et al. [33] of 6 species are staged on a *D. melanogaster* reference built from Graveley et al. [25] samples (as in Fig 3a).

a, Choice between identical models fit on gene expression with linearly-scaled chronological age or RAPToR estimates as predictors.
b, R² per gene of models with chronological age as predictor vs. the same model using RAPToR estimates, across 2-8 of spline degrees of freedom.

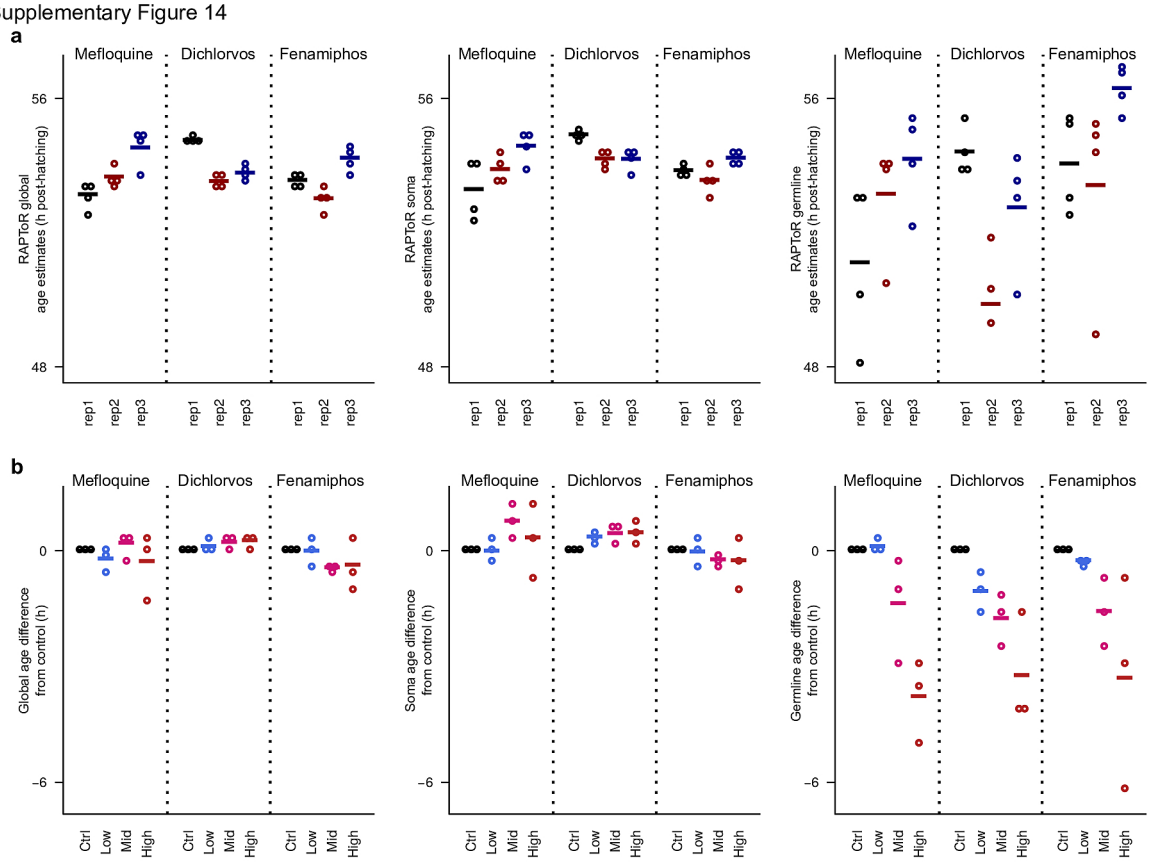
Supplementary Figure 13



Supplementary Figure 13 – Staging *C. elegans* embryogenesis with *D. melanogaster*

a, *C. elegans* embryo samples from Levin et al. [27] staged on the *D. melanogaster* reference built from Graveley et al. [25] samples. Gaps appear in the estimates, likely at points where fly expression dynamics are incompatible with those of worms.
b, As in (**a**), staging on the adjusted fly reference and using only highest-correlated genes between fly and worm embryogenesis (see methods).
c, *D. melanogaster* (red) and *C. elegans* (black) clustered gene expression profiles of highest-correlated genes between both species (see methods).
d, ICA components of the *C. elegans* embryo time course plotted along sampling time. Both the red highlighted outlier and 4 samples with erroneous chronological age (circled in IC1) are omitted from analysis (see methods).

Supplementary Figure 14



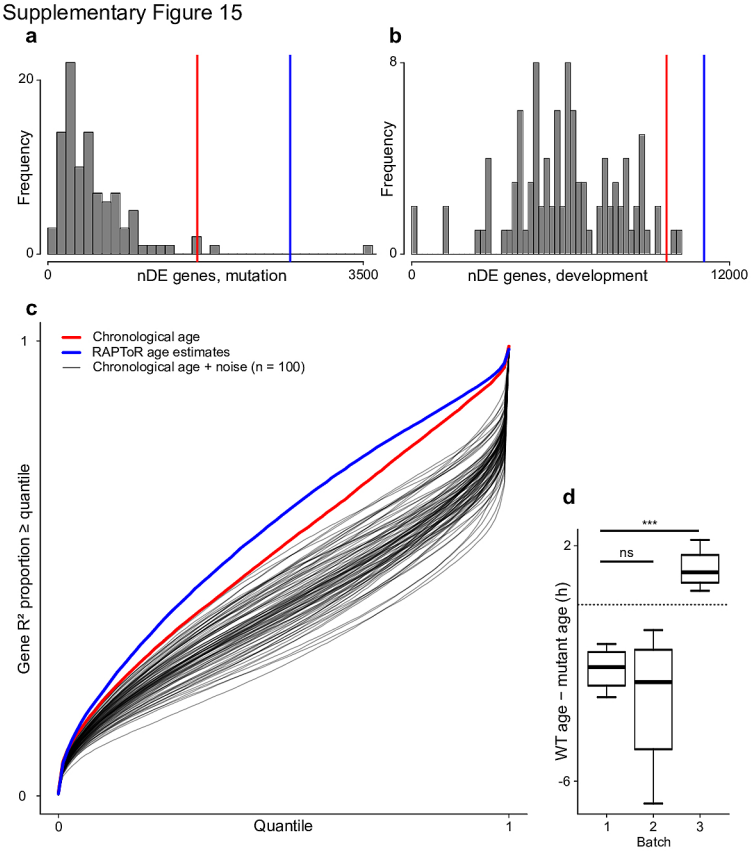
Supplementary Figure 14 – Impact of drugs on C. elegans germline development

C. elegans samples exposed to 3 doses of 3 different drugs profiled by Lewis et al. [35] are staged on the larval to young-adult reference built from Meeuse et al. [21] samples.

a, Impact of batch on global, soma and germline age.
b, Impact of drug dose on global, soma and germline age, normalized per batch. Age difference is computed by subtracting the age of the control sample within each batch.

In **a,b**, bars indicate group mean.

Supplementary Figure 15

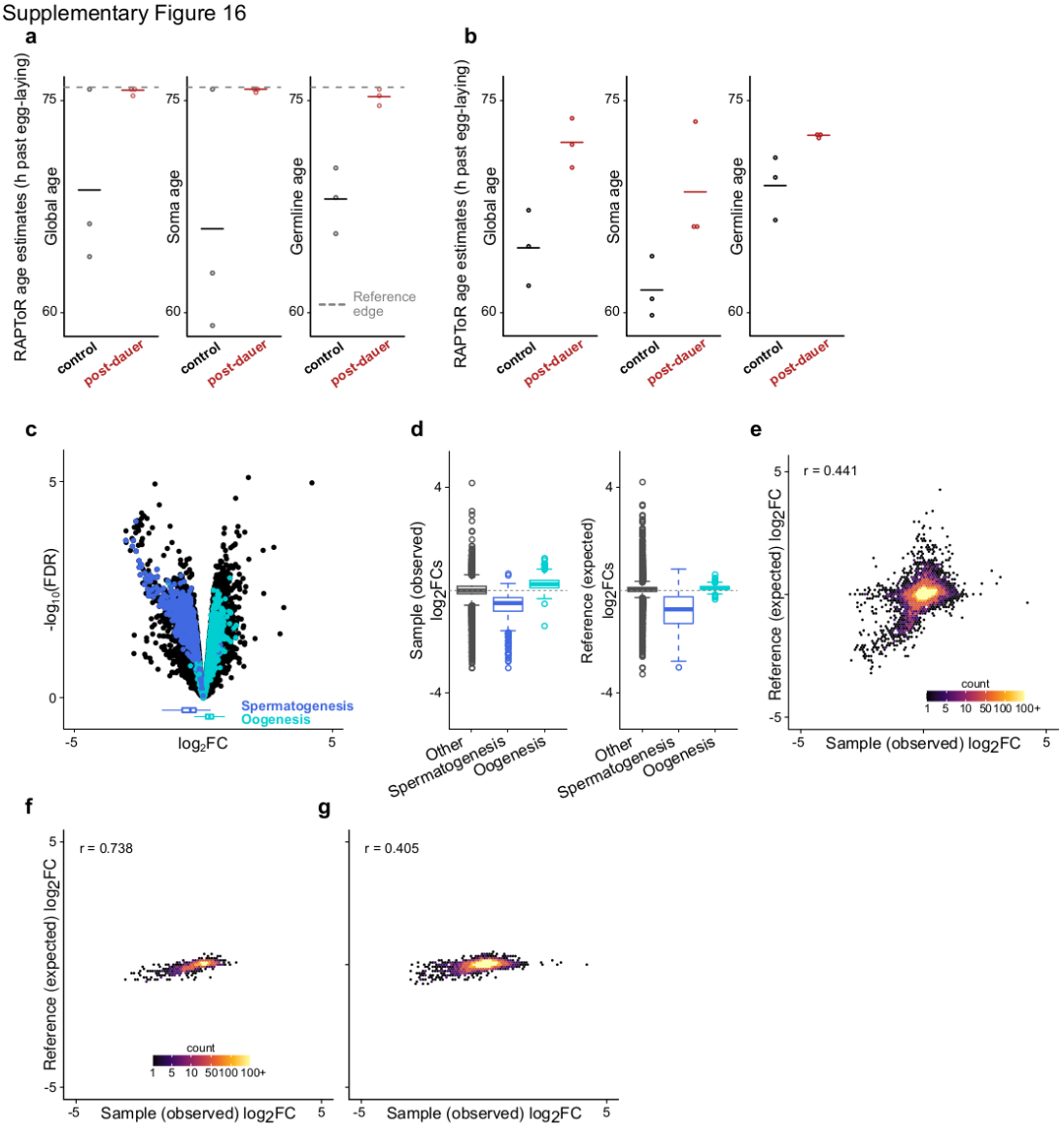


Supplementary Figure 15 – Increasing the power of DE gene detection with RAPToR estimates

*pash-1*ts or wild-type C. *elegans* samples profiled by Lehrbach et al. [36] are staged on the young-adult to adult reference built from Reinke et al. [22] samples.

a, R^2 per gene of identical models with chronological age, or RAPToR age estimates.
b-d, Random perturbations on chronological age (n=100 sets, see methods) produce overall fewer DE genes for strain (**b**), development (**c**), and have lower model fits than chronological age whereas RAPToR age estimates systematically outperform chronological age (**d**).
e, Batch effect on developmental difference between control and *pash-1*ts samples. Each box is n = 4.

Supplementary Figure 16

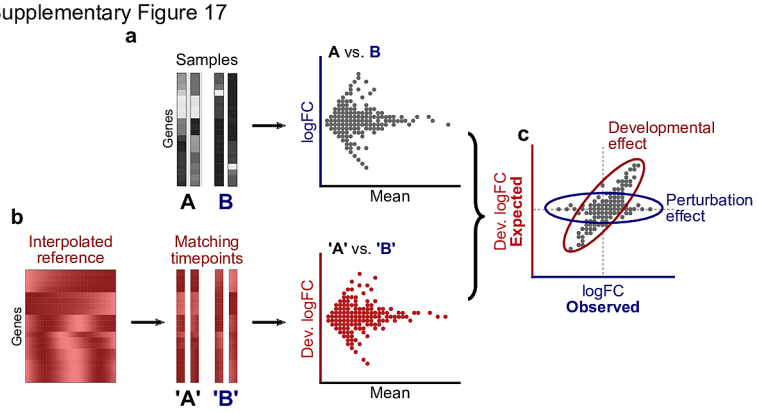


Supplementary Figure 16 – Germline expression changes recapitulated by a developmental shift

a,b, Global, soma-specific and germline-specific RAPToR age estimates of *C. elegans* control and post-dauer (PD) samples[37] staged on a reference built from Meeuse et al.[21] (**a**), and Reinke et al.[22] (**b**). Bars indicate mean age per group.
c, Volcanoplot of control vs. PD groups. Spermatogenesis and oogenesis genes are color-coded and logFCs of both categories are shown in boxplots at the bottom.
d, logFCs of control vs. PD for spermatogenesis and oogenesis genes observed in samples (left) and expected from development in the reference built from Meeuse et al., (right)
d, Observed expression logFCs between control and PD samples vs. expected developmental expression logFCs from the reference built from Meeuse et al. (as in Fig 5h, but for all genes)

e,f, Observed expression logFCs genes between control and PD samples vs. expected developmental expression logFCs from the reference built from Reinke et al., for germline genes (**e**), and all genes (**f**).

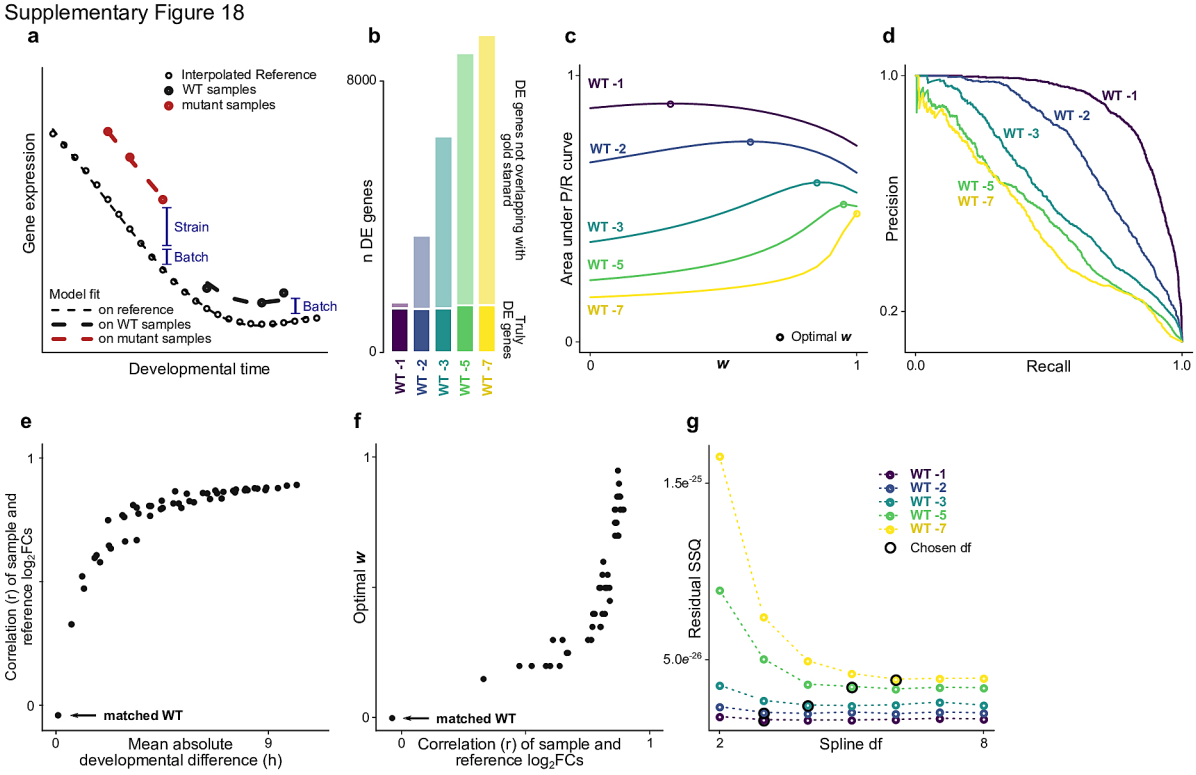
logFC, log2 fold-change. FDR, false discovery rate.



Supplementary Figure 17 – Estimating the impact of development by integrating reference data

a-c, Cartoon detailing how the logFCs of a differential expression analysis between two sample groups (**a**) and the logFCs of their matching time points in the RAPToR interpolated reference (**b**) can be compared to quantify the impact of development (**c**).

Supplementary Figure 18



Supplementary Figure 18 – Correcting the effect of development by integrating reference data

Samples from *C. elegans* time-course experiments of WT and *xrn-2* mutants, profiled by Miki et al. [38], and staged on the larval to young-adult reference built from Meeuse et al. samples[21], are used to validate developmental correction approach. See also Fig 5e-h.

a, Cartoon of a model integrating a window of reference data, with Strain and Batch coefficients shown in blue.

b, Number of DE genes found by a standard DE model (FDR < 0.05) increases with the age gaps between compared groups, with a quasi-constant fraction of truly DE genes.

c, w parameter optimization for shifted WT sets, by maximizing area under the Precision-Recall curves.

d, Precision-Recall curves of gold-standard gene detection by the age-corrected classifier for each shifted WT subset.

e, Correlation of expected development logFCs and observed logFCs between the *xrn-2* subset and combinations of 3-sample WT sets (note these are not the “WT -n” subsets, see Sup. Table 8).

f, Relationship between optimal w and sample-reference logFC correlation, as in (**f**).

g, Optimal spline degree-of-freedom (df) selection for the different WT shifted sets by reaching a residual Sum of Square (SSQ) plateau. The selected df increases with the shift, which is expected since the reference window to include gets larger and may thus contain more complex dynamics.

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

References (Supplementary)

46. Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci.* **97**, 10101–10106 (2000).

47. Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. & Davis, R. W. Significance analysis of time course microarray experiments. *Proc. Natl. Acad. Sci.* **102**, 12837–12842 (2005).