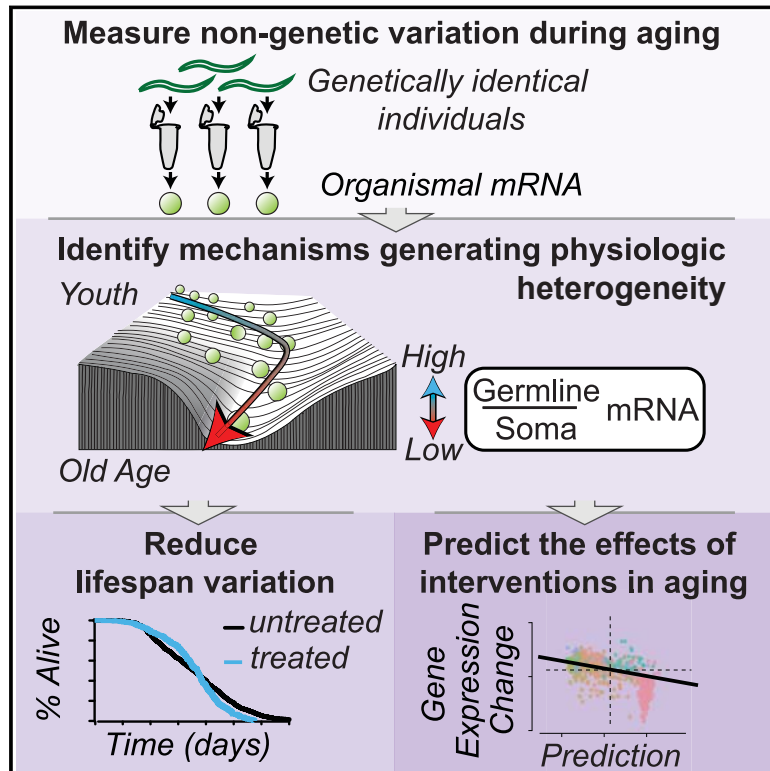Article

Systematic mapping of organism-scale gene-regulatory networks in aging using population asynchrony

Graphical abstract

Authors

Matthias Eder, Olivier M.F. Martin,
Natasha Oswal, ..., Arnau Sebé-Pedrós,
Holger Heyn, Nicholas Stroustrup

Correspondence

nicholas.stroustrup@crg.eu

In brief

A multi-scale measurement and perturbation approach reveals intrinsic biological sources of inter-individual heterogeneity in gene expression, health, and lifespan.

Highlights

- An atlas of 2,700 individual *C. elegans* transcriptomes quantifies heterogeneity in aging

- A perturbation screen identifies intrinsic drivers of this heterogeneity

- Heterogeneity in aging is driven by pleiotropic genes expressed in diverse cell types

- Suppressing intrinsic sources of heterogeneity reduces variation in health and lifespan

CellPress

Cell

CellPress

Article

Systematic mapping of organism-scale gene-regulatory networks in aging using population asynchrony

Matthias Eder,[1,5] Olivier M.F. Martin,[1,2,5] Natasha Oswal,[1,2,5] Lucia Sedlackova,[1] Cátia Moutinho,[3] Andrea Del Carmen-Fabregat,[1,2] Simon Menendez Bravo,[1] Arnau Sebé-Pedrós,[1,2,4] Holger Heyn,[3,4] and Nicholas Stroustrup[1,2,6,*]

[1]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
[2]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona, Spain
[4]ICREA, Pg. Lluis Companys 23, Barcelona 08010, Spain
[5]These authors contributed equally
[6]Lead contact
*Correspondence: nicholas.stroustrup@crg.eu
https://doi.org/10.1016/j.cell.2024.05.050

SUMMARY

In aging, physiologic networks decline in function at rates that differ between individuals, producing a wide distribution of lifespan. Though 70% of human lifespan variance remains unexplained by heritable factors, little is known about the intrinsic sources of physiologic heterogeneity in aging. To understand how complex physiologic networks generate lifespan variation, new methods are needed.
Here, we present Asynch-seq, an approach that uses gene-expression heterogeneity within isogenic populations to study the processes generating lifespan variation. By collecting thousands of single-individual transcriptomes, we capture the *Caenorhabditis elegans* "pan-transcriptome"—a highly resolved atlas of non-genetic variation. We use our atlas to guide a large-scale perturbation screen that identifies the decoupling of total mRNA content between germline and soma as the largest source of physiologic heterogeneity in aging, driven by pleiotropic genes whose knockdown dramatically reduces lifespan variance. Our work demonstrates how systematic mapping of physiologic heterogeneity can be applied to reduce inter-individual disparities in aging.

INTRODUCTION

In most species, including humans,[1,2] mice,[3] and nematodes,[4–7] aging individuals exhibit a remarkable degree of physiologic heterogeneity that ultimately leads to a wide distribution of lifespan. Twin and familial studies suggest that less than thirty percent of inter-individual differences in lifespan are caused by heritable factors,[8,9] and laboratory studies demonstrate that controlling for both genetic and environmental variation does little to reduce lifespan variation.[3–5] Diverse cellular and organismal mechanisms have been found to influence a population's average lifespan, including endocrine signaling, metabolism, protein homeostasis, and many others,[10–16] and yet it remains unclear how such mechanisms influence the physiologic heterogeneity driving lifespan variance.

A fundamental barrier to building mechanistic models of physiologic heterogeneity in aging multicellular systems is the multiscale functional interdependence of the mechanisms involved. Though a wide variety of mid-life physiologic measurements have been found to predict the lifespan of individuals within genetically homogeneous populations,[1–7,17–22] it remains challenging to identify the specific processes converting physiologic differences between individuals mid-life into lifespan differences. The mechanisms driving aging act not in isolation but embedded as components of complex homeostatic networks[23]—a barrier for many conventional experimental approaches that struggle in the presence of feedback loops and one-to-many interactions. Despite recent progress in unicellular organisms,[22,24] new methods are needed to build mechanistic, dynamical models of multicellular aging.

Here, we describe Asynch-seq, an interdisciplinary combination of experimental and modeling approaches that lets us systematically identify the molecular mechanisms driving non-genetic physiologic heterogeneity and lifespan variation. Taking advantage of the speed and scalability of modern RNA sequencing (RNA-seq) technologies, we generate an atlas of the non-genetic heterogeneity in gene expression that arises *de novo* during adulthood. We then model these gene-expression data and identify the mechanisms acting during adulthood to generate one-third of all gene-expression variance. Finally,

in a series of perturbation experiments, we demonstrate that direct manipulation of the mechanisms generating physiologic heterogeneity in aging can dramatically reduce that heterogeneity to produce better and more equal outcomes in aging across a population.

RESULTS

A high-resolution atlas of non-genetic, inter-individual variation in gene expression

To measure non-genetic variation, we first needed a high-throughput method that would let us systematically quantify gene-expression differences within isogenic populations. To collect transcriptomic data at the scale and quality required, we optimized the Smart-Seq2 single-cell sequencing protocol[25,26] to accurately estimate inter-individual gene-expression variance (STAR Methods).

We applied this method to compile an atlas of non-genetic gene-expression variation in aging that includes not only wild-type populations but also a set of mutations and environmental conditions known to alter lifespan. Our atlas contains 2,764 single-individual transcriptomes (Data S1), including (1) a longitudinal time series containing wild-type individuals collected every second day from day 1 until day 12 and (2) a very high statistical resolution comparison of wild-type young (day 1) and aged (day 8) individuals across multiple replicates. On days 1 and 8, we also observed (3) long-lived insulin/IGF receptor *daf-2(e1368)* mutants, (4) germline-ablated *glp-1(e2141)* mutants, (5) wild-type individuals whose lifespan was extended by a diet of ultraviolet (UV)-light-inactivated *E. coli*, and, finally, (6) individuals whose lifespan was shortened by exposure to elevated environmental temperature (20°C vs. 25°C) from hatching until death.

Our atlas reveals a striking degree of inter-individual heterogeneity present in the transcriptomes of wild-type populations. We find that transcriptomes are not uniformly variable and that at any age some genes exhibit 100-fold higher variance compared with the least variable genes at equivalent mean expression (Figures 1A–1C; Data S2). We find that 78% of all transcripts show significant increases in variance between days 1 and 8 (Figures 1D and S1S; Data S3; Wald $p < 0.01$; studentized bootstrap), highlighting a systematic age-associated increase in non-genetic variation across the transcriptome. Crucially, we find that 47% of the transcriptome shows age-associated increases in variance without large changes in the mean (Figure 1D inset; mean fold-change $\geq 2^{1.5}$ and Wald $p < 0.01$), signifying a type of gene-expression heterogeneity not captured by bulk or standard single-cell sequencing approaches. During aging, our results show that wild-type, age-synchronous individuals become physiologically heterogeneous in respect to almost half of their transcriptome.

Inter-individual variation aligns with population-average changes only during the second half of adulthood

A common assumption made in aging studies is that non-genetic heterogeneity arises as the result of individuals following the same stereotyped aging trajectory but at different rates.[5,18,27,28] To evaluate this assumption, we performed a principal-compo-

nent analysis (PCA) of our aging time series data and inspected the trajectory that individuals take along the first two principal components (PCs) that together explain 58.7% of all variance (Figure 1F). We observe more complex dynamics than anticipated—during the first 6 days of adulthood, cohorts progress on average along PC2, which explains only 9.1% of variance, but on day 6 they turn to progress along a diagonal in respect to both PC1 and PC2 (Figure 1F inset, blue arrow). Within populations on each day, we find that inter-individual variations do not consistently fall along the same path that populations follow on average over time. Instead, at all ages, individuals throughout life appear to vary along a diagonal of PC1 and PC2 nearly orthogonal to the population-average trajectory until day 6 (Figure 1F, red arrows). We conclude that the non-genetic variation in aging does not arise primarily from individuals following a stereotyped aging trajectory at different rates. Therefore, our data motivate a rethinking of a large class of "pseudo-time" models[18,27,29,30] conventionally used to estimate an individual's "biological age" as distinct from their "chronological" age, as these methods miss potentially crucial physiologic differences between age-synchronous individuals. Instead, alternate approaches for understanding non-genetic variation in aging are needed.

Inter-individual differences in aging involve a decoupling of germline and somatic gene expression

To identify the mechanisms generating non-genetic variability in our atlas, we first focused on the genes with the highest variance in old (day 8) populations, which we find are enriched[31] for genes involved in stress and pathogen responses,[19,32] sperm physiology, and vitellogenin yolk production during oogenesis.[33] Because the reproductive system plays a fundamental, evolutionarily conserved role in aging,[34–44] we then set out to understand the role of germline gene expression in non-genetic variation.

To distinguish between changes in germline and somatic gene expression, we considered 658 genes previously identified[45] as being expressed predominantly in single tissue types: the hypodermis, muscle, neurons, the alimentary system, including the pharynx and intestine, and the germline (STAR Methods). Aging in many species involves large changes in cell-type composition,[46,47] and to study this in individual *C. elegans* we grouped genes according to their tissue of origin and on day 8 observed a strong positive correlation among genes expressed in the same tissue type (Figure 2A). We further observed strong negative correlations between germline-specific and soma-specific genes, demonstrating that in aged individuals, somatic genes co-vary together as a group, germline genes co-vary together as a group, and these two groups exhibit partial independence in their variation (Methods S1).

To quantify changes in germline and somatic mRNA in single individuals, we included synthetic RNA spike-ins in our sequencing approach that allowed us to perform absolute quantification of mRNA abundance (STAR Methods). We find that, at the start of adulthood (day 1), individuals show substantial differences in the overall amount of total mRNA in their bodies—some individuals have larger transcriptomes than others—but within each individual the amounts of germline and somatic mRNA
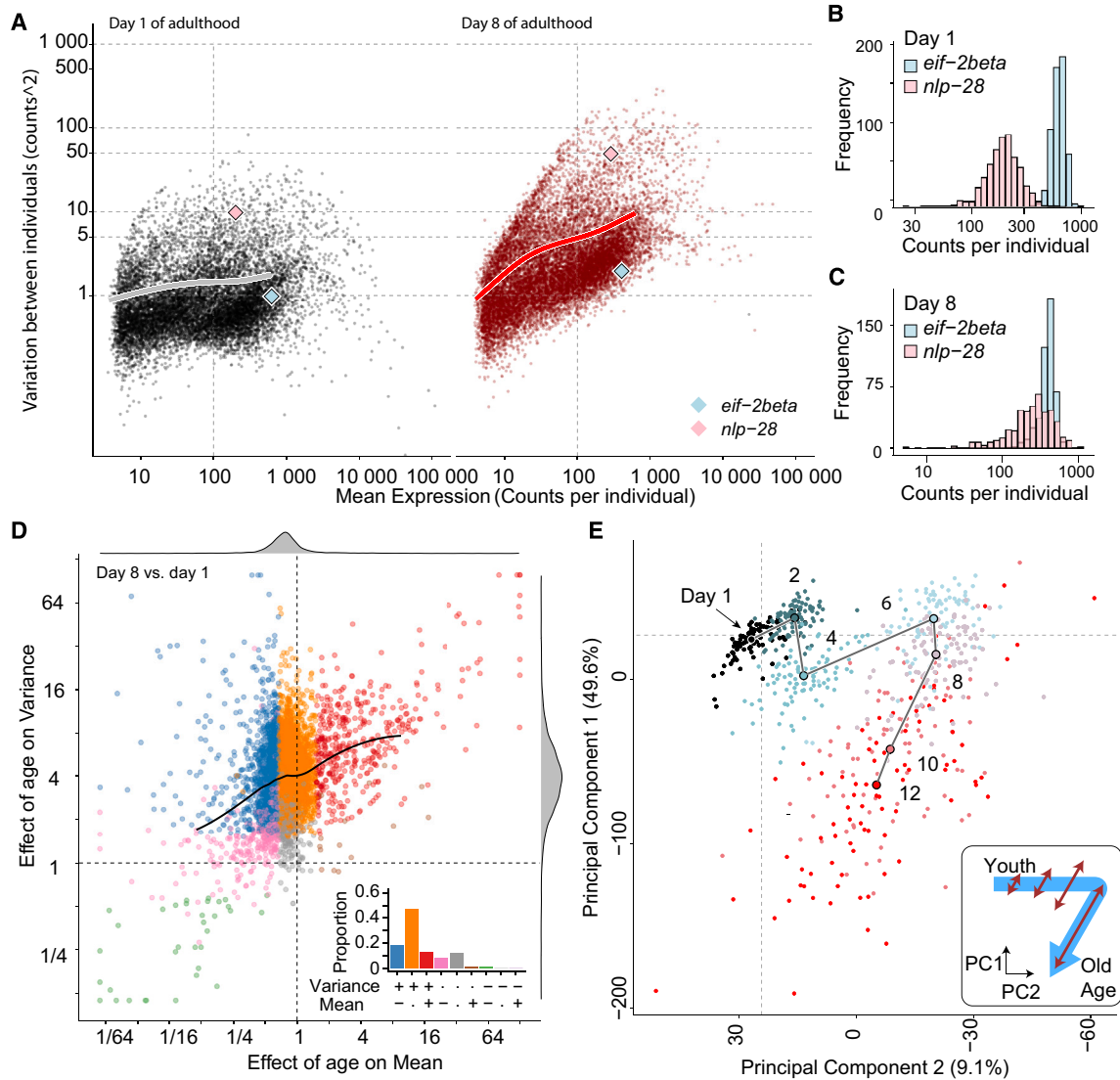
Cell
Article

CellPress

Figure 1. A population-scale atlas of non-genetic heterogeneity

(A) mRNA was separately obtained from 1,042 genetically identical individuals either in youth (black; day 1) or old age (red; day 8). Population mean and variance are shown for each transcript (points).

(B and C) Two examples, *nlp-28* (pink diamond) and *eif-2 beta* (blue diamond), are highlighted along with (B) the distribution of expression levels observed across individuals in youth (C) and old age.

(D) Age-associated changes in gene-expression mean and variance (day 8 vs. day 1) were estimated for each transcript (points). Genes are colored according to their significant increase (↑), decrease (↓) ($p < 0.001$, Wald test), or no change (Ø) in mean (M) and variance (V) with age: M↓V↑ (blue); MØV↑ (orange); M↑V↑ (red); M↓VØ (pink); MØVØ (gray); M↑VØ (brown); M↓V↓ (green). (D) (inset) The proportion of transcripts in each category.

(E) Individuals (points) from seven separate age cohorts are projected onto the first and second principal components (PCs) of gene expression. Outlined circles indicate cohort averages, connected by lines. (E) (inset) The non-genetic differences between individuals within each cohort (red arrows) do not align with the population-average aging trajectory (blue arrow) until old age.

See also Figure S1.

are highly correlated. We observe age-associated declines in this correlation, dropping from a Pearson correlation of $\rho = 0.94$ on day 1 to $\rho = 0.64$ on day 12 (Figures 2B and 2C). This drop in correlation is the result of disproportional decreases in somatic and germline mRNA content, starting on day 4 (Figures S2A and S2B). Our data highlight both the existence of some mechanism for coordinating the balance of germline and somatic mRNA content in young individual animals and the failure of this mechanism during aging.

To confirm and characterize the failure of coordination between germline and somatic mRNA content that occurs during aging, we attempted to induce the failure experimentally. Applying a tissue-specific RNA interference (RNAi) strategy,[48] we inhibited germline transcription via knockdown of RNA
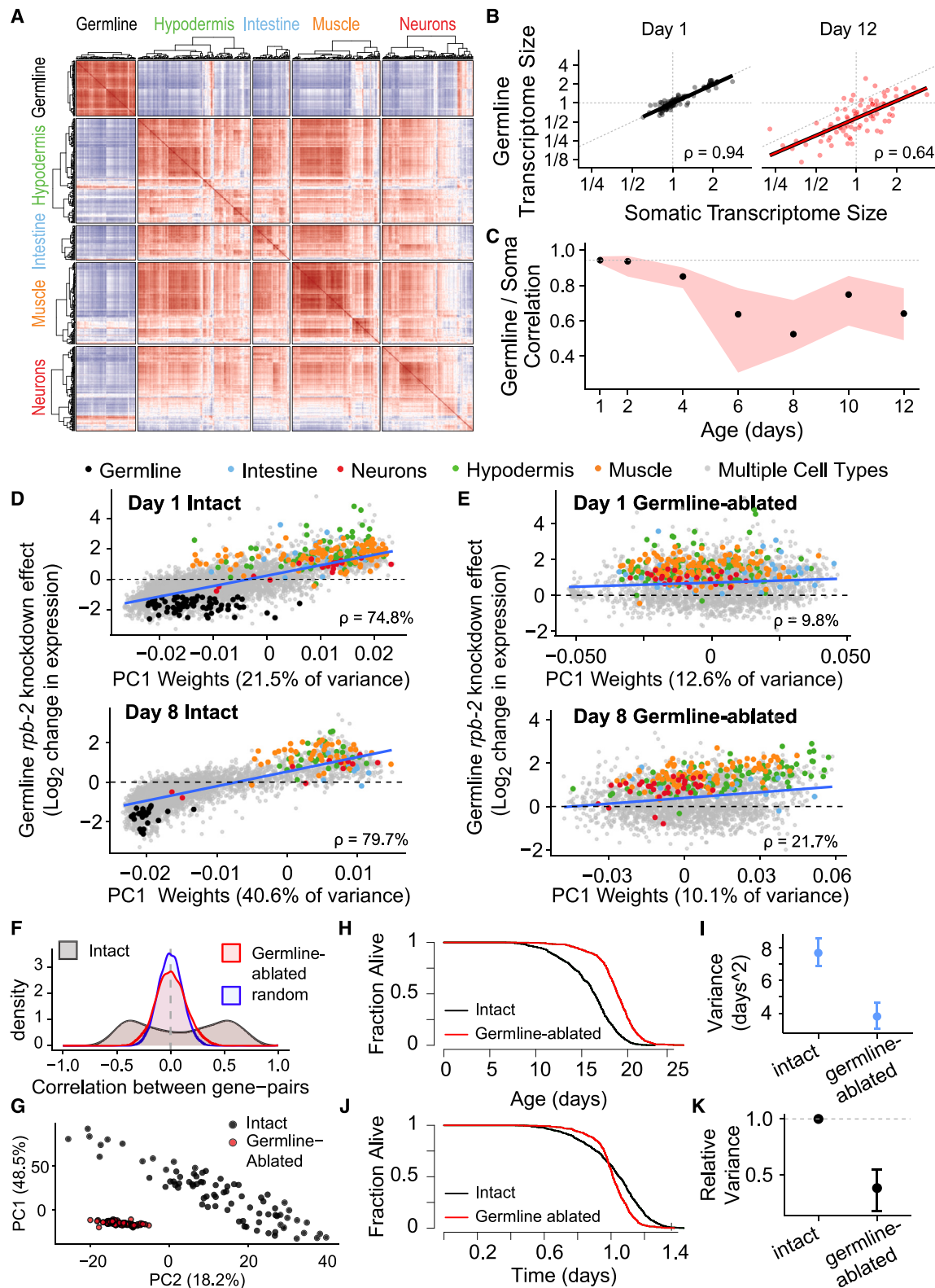
Figure 2. Aging involves a decoupling of germline and somatic mRNA content

(A) Transcripts expressed only in single tissue types were hierarchically clustered according to their co-expression across aged individuals (day 8). Pearson correlations (ρ) range from −1 (blue) to 0 (white) to 1 (red).

(legend continued on next page)

CellPress

polymerase II (RNA Pol II) subunits *rpb-2* or *ama-1*. By day 8, we find that experimental disruption of germline RNA Pol II produced global changes in organismal gene expression, decreasing germline mRNA in proportion to somatic mRNA (Figures 2D and S2C). The effects of germline *rpb-2* and *ama-1* knockdown recapitulate the naturally arising differences we observe between aged individuals: the magnitude of each gene's change after *rpb-2* knockdown is strongly correlated with that gene's weight in the first principal component (PC1) of the wild-type day 8 covariance matrix (Figures 2D and S2C). Importantly, each gene's weight in PC1 quantifies that gene's contribution to the largest axis of non-genetic variation in unperturbed populations. Therefore, we conclude that experimental inhibition of germline transcription can reproduce the differences in germline and somatic mRNA that exist between aged individuals.

We then used the gene-expression signature of *rpb-2* knockdown to quantify how the coordination of germline and somatic mRNA content naturally changes with age. We observe a high correlation between germline *rpb-2* knockdown and the PC1, both in the young population (day 1 ρ = 74.8%) and aged population (day 8 ρ = 79.7%). However, the fraction of all variation explained by PC1 doubles during aging, from 21.5% to 40.6% between days 1 and 8 (Figures 2D and S2D). This doubling demonstrates that the contribution to overall gene-expression variation made by individual differences in somatic and germline mRNA content increases with age. To confirm that such effects are not specific to any particular method of germline gene expression suppression, we also compared the gene-expression signatures of two temperature-sensitive mutations that disrupt germline gene expression by ablating the germline—*glp-1(e2141)* and *glp-4 (bn2)*[49,50]—and identified the same results as with *rpb-2*(RNAi) (Figures S2E–S2H).

Taken together, our analyses demonstrate that the single largest contributor to non-genetic variation in *C. elegans* aging is a decoupling of germline and somatic mRNA content.

Germline ablation decreases non-genetic gene-expression variation in aging

We then asked whether the germline itself might play a causal role in generating non-genetic gene-expression variation. Measuring old (day 8) germline-ablated *glp-1(e2141)* individuals,

we observe global decreases in gene-expression variance relative to wild type. In germline-ablated populations, genes show on average 25% (23%–28% bootstrap confidence interval [CI]) lower expression variance (Figure S2I). We find that this decrease involves both a cell-autonomous germline activity to eliminate high-variance germline-specific genes and also a cell-non-autonomous germline activity to decrease the variability of soma-specific genes (Figures S2J and S2K). These results highlight the causal role played by the germline in generating gene-expression variation in aging.

Repeating our PCA on the germline-ablated population, we identify a dramatic reduction in the correlation between PC1 and germline-specific *rpb-2(RNAi)* effects (Figure 2E), demonstrating that the germline is required specifically for the emergence of inter-individual variance along PC1 in wild-type populations. The total amount of variance explained by PC1 in germline-ablated populations is much lower than that explained by PC1 in wild-type populations: 10.1% in germline-ablated compared with 40.6% in wild type on day 8—a side-effect of the global 5-fold decrease in the covariance among all genes in germline-ablated populations relative to wild type (Figure 2F). We conclude that, following germline ablation, individuals not only differ less from each other but also differ in a less "structured" way, such that individuals differ across a larger number of independent dimensions of gene expression (e.g., PCs, Figure S2C). The germline generates 25% of all non-genetic gene-expression variance, while directing this variance along a single organismal gene-expression axis (Figures 2G and S2L).

Germline ablation halves inter-individual variation in healthspan and lifespan

Germline ablation via *glp-1(e2141)* is well known to extend mean lifespan[51] through partially characterized signaling pathways.[38] We therefore considered the possibility that the germline might contribute to lifespan variation. First, we confirmed previous findings that *glp-1(e2141)* germline ablation increases lifespan (Figure 2H). Then, remarkably, we find that our high-resolution lifespan data reveal a large decrease in lifespan variance 3.8 (3.1–4.6) days[2] in *glp-1(e2141)* populations compared with 7.6 (6.8–8.5) days[2] in wild type (Figure 2I). Such large, opposing effects on lifespan mean and variance are unprecedented in any lifespan-altering intervention, as mutations, diets, and environmental conditions

(B) Each individual's total germline and somatic mRNA contents are compared within young (day 1) and old (day 12) cohorts.

(C) The Pearson correlation between germline and somatic mRNA in successive age cohorts, with 95% confidence bands (pink).

(D) A separate population was exposed to germline-specific knockdown of RNA polymerase II (RNA Pol II), and the response of each transcript (points) to knockdown (y axis) is compared with that transcript's weight in the first principal component (PC1) (x axis) of wild-type individuals. PC1 weights quantify the contribution of each transcript to non-genetic variation along the PC1 of young (top) and aged (bottom) cohorts.

(E) The same RNA Pol II knockdown effects are compared with the first PC1 of germline-ablated *glp-1(e2141)* cohorts of young (top) and old (bottom) individuals.

(F) To identify global differences in transcriptome covariance between wild-type (black) and *glp-1(e2141)* (red) cohorts, the distributions of pairwise correlations between all transcript-pairs are shown for day 8 cohorts, and compared with the same distribution for a random matrix generated by permuting gene names in the wild-type cohort (blue).

(G) Wild-type (black) and *glp-1(e2141)* (red) individuals on day 8 are projected onto PCs 1 and 2 obtained from a PCA of both genotypes combined.

(H) The lifespans of *glp-1(e2141)* individuals are compared with those of wild-type individuals, shown as Kaplan-Meier survival estimates.

(I) The variance in lifespan for both populations estimated using Weibull parametric fits, with 95% confidence intervals.

(J) The AFT regression residuals for both populations highlight disproportional differences in lifespan variance relative to the mean.

(K) To quantify any disproportional effect of *glp-1(e2141)* on lifespan variance, we calculated the ratio of variances of the two populations' AFT regression residuals, with bootstrap 95% confidence intervals.
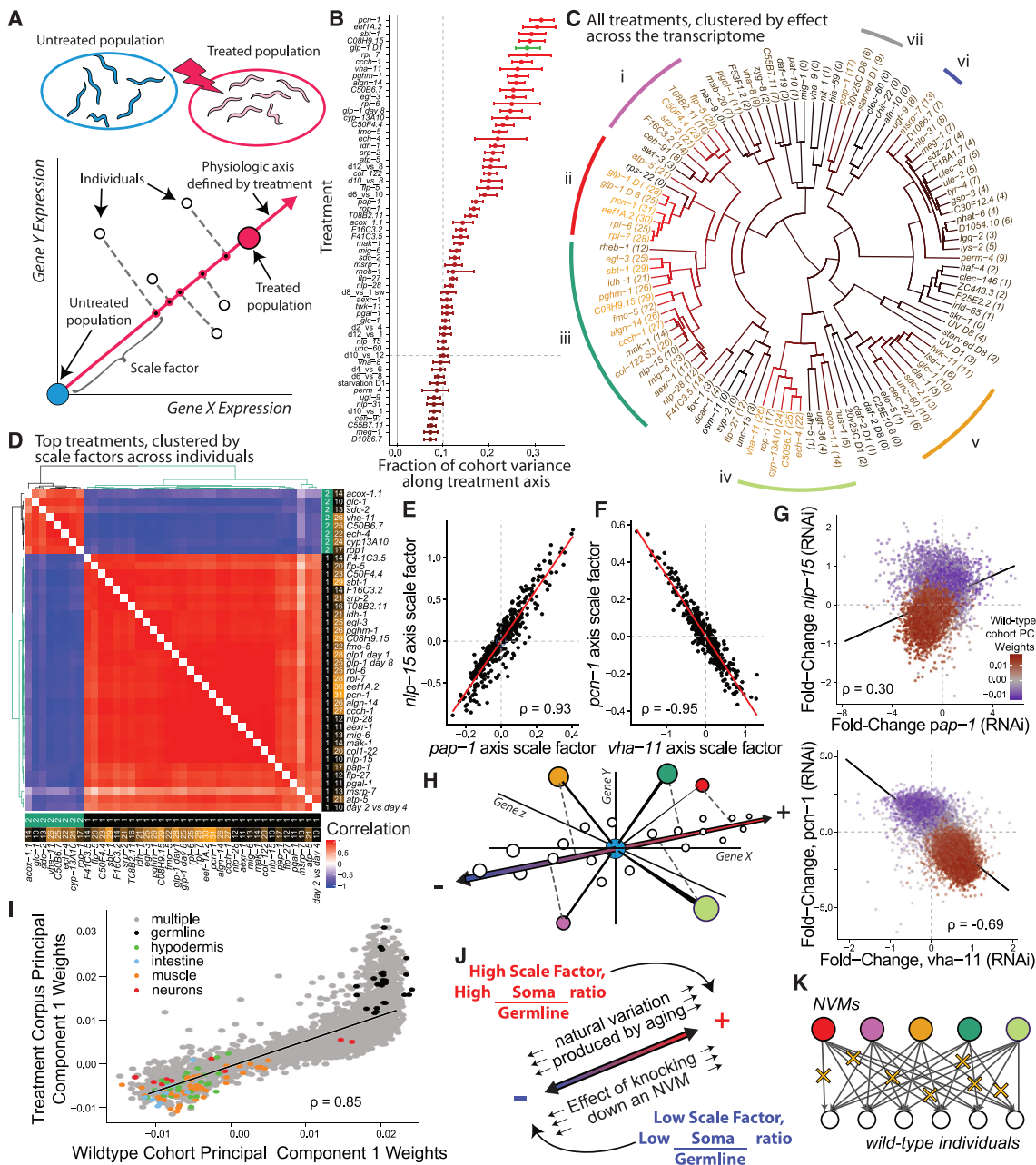
See also Figure S2.

Figure 3. A perturbation screen to identify molecular drivers of non-genetic heterogeneity

(A) We use transcriptome scaling models (TSMs) to quantify the extent to which non-genetic differences between aged individuals (white circles) fall along single, treatment-defined (pink circle), gene-expression axes (pink line).

(B) In a screen of 104 treatments, we identify 40 TSM hits—non-genetic variation mimetics (NVMs)—that define axes explaining more than 10% (gray line) of non-genetic variance within aged (day 8) wild-type cohorts. Additional treatments are shown in Figure S3F; bootstrap 95% CIs.

(C) All treatments were hierarchically clustered according to correlations in the fold-change effect of each treatment on gene expression. Dissimilar groups are highlighted (colored arches).

(D) The 40 hit NVM treatments were clustered according to correlations in the scale factors assigned by TSMs to individuals in the wild-type day 8 cohort along each NVM's axis.

(E and F) For example, (E) the scale factors assigned to individuals (black dots) along the *pap-1(RNAi)* axis are correlated to the scale factors assigned along the *nlp-15(RNAi)* axis. (F) The same, but highlighting an anti-correlation between *vha-11* and *pcn-1* scale factors.

(G) The fold-change effects of *pap-1* and *nlp-15* on gene sxpression (top), with genes (points) colored according to their weight in first principal component (PC1) of the wild-type day 8 cohort, with a linear regression line (black). (Bottom) The same analysis comparing *pcn-1* and *vha-11*.

(H) Our results suggest a model in which treatments produce pleiotropic effects on gene expression, having different overall effects (colored outer circles) relative to wild type (blue circle) but all sharing projections along a shared axis of non-genetic variation of aged wild-type individuals (white circles).

(legend continued on next page)

Cell
Article

CellPress

that alter lifespan usually produce exactly proportional changes in lifespan variance.[52] To quantify the disproportional effect of germline ablation on lifespan mean and variance, we fit an accelerated failure time (AFT) regression model (Figure 2J) to our data. We find that *glp-1(e2141)* populations show a 2.4-fold (1.9–3.1) lower variance in AFT residuals compared with wild type, demonstrating that the intact germline is required for 58% (47–68%) of all lifespan variation in wild-type populations (Figure 2K). Next we considered a widely used biomarker of organismal "health" in model organisms, vigorous movement,[7,20,53] and found that germline ablation extends vigorous movement from an average of 11.6 (11.4–11.8) days in wild-type to 14.0 (13.7–14.3) days in *glp-1(e2141)* populations (Figures S2M and S2N) while simultaneously decreasing variance from 7.1 (6.4–7.8) to 5.3 (4.7–6.0) days[2]—a disproportional decrease in variance of 2.0-fold (1.6–2.6) (Figures S2O–S2Q). Our results demonstrate that the germline plays a central role in generating age-associated heterogeneity, not only in gene expression but also in health and lifespan. By ablating the germline, we can alter the intrinsic biological generation of lifespan variation, yielding more equal outcomes in respect to multiple aging phenotypes.

Transcriptome scaling analysis quantifies non-genetic variation along experimentally defined axes

We then set out to systematically identify additional mechanisms that generate physiologic heterogeneity during aging. To apply our high-throughput RNA-sequencing strategy to this task, we developed a computational approach to predict any gene's contribution to inter-individual physiologic heterogenity. Our analysis of *rpb-2(RNAi)* suggests that some genes can contribute to inter-individual heterogenity by scaling each individual's transcriptome in a manner similar to what is observed in response to those genes' knockdown. Therefore, we developed transcriptome scaling models (TSMs) that take a set of single-individual transcriptomic data, such as our atlas, and project it onto a single high-dimensional vector describing the effects of some experimental intervention—e.g., the RNAi knockdown of a candidate gene. Using TSMs, we can quantify the amount of gene-expression variance present between individuals that can be explained by those individuals varying along a single gene-expression axis defined by the experimental intervention (Figures 3A and S3A). Formally, if C and T are the transcriptomes of control and treated populations, respectively (Figure 3A red and blue circles), a TSM decomposes the transcriptome of each individual, y_i, into two parts: a scale factor λ_i that measures the individual's position along the $C \rightarrow T$ axis and a residual ϵ_i that quantifies deviations from scaling (Figure 3A dotted lines; STAR Methods). TSMs are related to existing methods proposed to decompose organismal gene-expression signatures into constituent parts,[27,54] but crucially they allow us to rapidly screen interventions using bulk transcriptomic methods and identify genes that contribute to inter-individual heterogeneity in aging.

To test the method, we applied our TSM approach to confirm results obtained previously by PCA. Using TSMs to study inter-individual variation on day 8 along the *glp-1(e2141)* axis, we find that 21% (95% CI 18%–25%) of all gene-expression variance can be explained by differences between individuals' somatic and germline mRNA content (Figures S3B–S3D). Furthermore, using TSMs, we can confirm pseudo-time approaches assigning individuals a single biological age capture of less than 10% of all gene-expression variance in the first half of life (Figure S3E).

A systematic screen for genetic and environmental drivers of non-genetic variation in aging

We then performed a large-scale perturbation screen to systematically identify mechanisms that contribute to non-genetic variation in gene expression, health, and lifespan. We explored two classic sources of phenotypic heterogeneity in biological systems—extrinsic and intrinsic[55]—by exposing individuals to different diets, temperatures, mutations, and an RNAi knockdown screen of 94 genes selected based on these genes' high-connectivity in the co-expression graph generated from our non-genetic variation atlas (STAR Methods; statistical methods).

We identify 40 "hit" treatments that define physiologic axes along which inter-individual variation explains at least 10% of total population gene-expression variation (Figures 3B and S3F). We name these treatments "non-genetic variation mimetics" (NVMs), based on their ability to reproduce the gene-expression differences that we observe to arise naturally between old, genetically identical individuals. We see a large degree of molecular diversity among these 40 NVMs, including tissue-specific genes localized not only to the germline but also to the muscle, hypodermis, pharynx/intestine, and neurons. These NVMs exhibit diverse molecular functions, including cell-cycle regulation, endocrine signaling, lysosomal function, mitochondrial ATP synthesis, metabolism, and ribosomal function. Taken together, our results show that the knockdown of genes with diverse molecular functions located in multiple tissue types can recapitulate the differences we observe arising naturally between genetically identical individuals during aging.

Surprisingly, neither the changes in temperature nor diet we considered acted as NVMs. We conclude that variation in individual responses to these environmental factors are not a direct source of non-genetic variation in aging. Instead, our screen demonstrates that the intrinsic mechanisms represented by our 40 NVM genes are the largest source of physiologic heterogeneity in aging.

Screen hits have dissimilar effects on organismal gene expression but all influence germline and somatic mRNA content

We then sought to identify any functional similarity between the 40 NVMs we identified. Hierarchical clustering reveals seven distinct

---

(I and J) (I) PC1 of the wild-type day 8 cohort (x axis) compared to PC1 of the corpus of NVM knockdown effects (y axis), demonstrating that (J) NVM knockdown shifts gene expression, on average, along the same axis that old individuals vary naturally in aging.

(K) Aged wild-type individuals (white circles) vary as if random subsets of NVMs (colored circles) had been knocked down (yellow Xs).
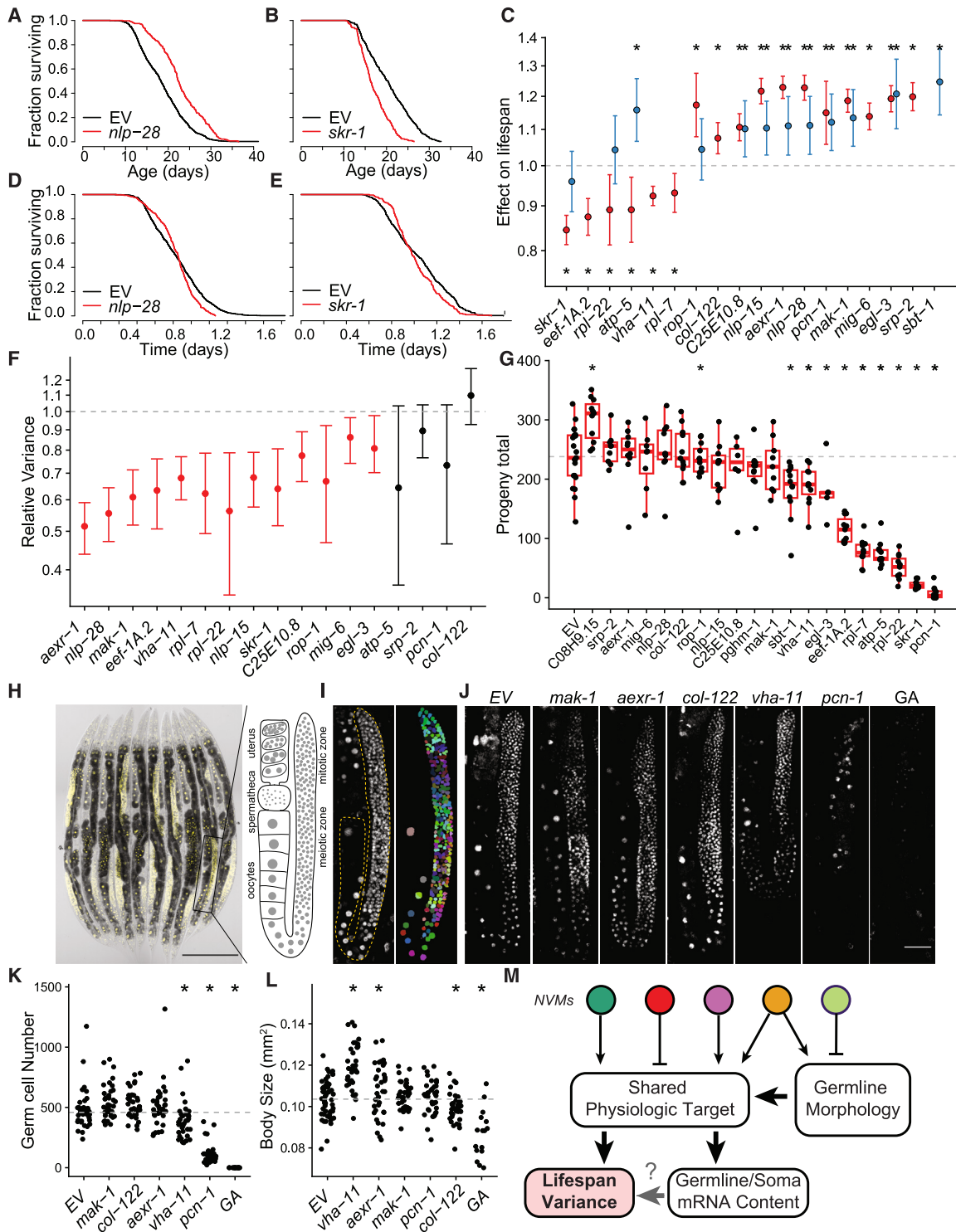
See also Figure S3.

Figure 4. A new class of molecular drivers of non-genetic variation in lifespan

(A and B) Several non-genetic variation mimetics (NVMs) were chosen for additional phenotyping. (A) The lifespan of populations exposed to *nlp-28(RNAi)* or (B) *skr-1(RNAi)* started at the beginning of adulthood *(late L4)*, plotted as Kaplan-Meier survival estimates.

(C) Estimates for the changes in lifespan produced by each NVM when knocked down, either at the beginning (red) or on the 4th day (blue) of adulthood. Estimates were obtained using accelerated failure time (AFT) regression; stars indicate Wald significance at $p < 0.05$.

(D and E) (D) AFT regression residuals are plotted for the *nlp-28* and (E) *skr-1* knockdowns shown in (A) and (B), highlighting disproportional effects on lifespan variance relative to the mean.

(legend continued on next page)

effects on gene expression produced by NVM knockdown (Figures 3C, S3G, and S3H), suggesting that these genes do not form a single gene-regulatory pathway.[56] We next asked whether NVMs interact by regulating each other's expression. Although knockdown of several NVMs significantly alters the expression of several others (Figures S3J and S3K), the NVMs are not all co-expressed during wild-type aging and instead form several loosely co-expressed groups (Figure S3L). We therefore conclude that our NVMs do not form a single gene-regulatory group.

Because the NVMs appeared dissimilar in many respects, we hypothesized that their knockdowns' mimicry of non-genetic variation might involve different aspects of aging—each NVM reflecting a qualitatively different way in which old individuals differ. Surprisingly, when we fit scaling models using each NVM to the wild-type day 8 cohort, we observe a striking similarity across all NVMs. The scale factors assigned by all NVMs, including glp-1(e2141), are strongly correlated, demonstrating that all NVMs define a single shared axis of inter-individual variation (Figure 3D). For example, we find that individuals with high scale factors along the pap-1(RNAi) axis—a conserved poly(A) RNA polymerase— also have high scale factors along the nlp-15(RNAi) axis—an endocrine signaling peptide involved in movement and body morphology (Figure 3E). We therefore conclude that, despite the apparent dissimilarity in the functions of nlp-15 and pap-1, their knockdowns shift gene expression along a shared gene-expression axis. Intriguingly, scaling models for vha-11(RNAi) and pcn-1(RNAi) assign anti-correlated scale factors across the day 8 population (Figure 3F), suggesting that these knockdowns produce shifts in opposite directions along the same axis. Despite an apparent dissimilarity in function and regulation, the 40 NVMs define a single axis of inter-individual variation in aging.

Though NVM knockdowns produce poorly correlated effects on the whole transcriptome, they nevertheless may produce correlated effects on a subset of the transcriptome. Comparing NVM knockdowns to PC1 of the wild-type day 8 cohort, we find that genes with high PC1 weights tend to be strongly up- or downregulated by NVMs (Figure 3G, red and blue points). Because PC1 captures inter-individual differences in total germline and somatic mRNA content, we conclude that NVM knockdowns all alter the balance of total germline and somatic mRNA content (Figure S3), despite these genes having overall dissimilar, pleiotropic influences on organismal gene expression (Figure 3H; Data S4 and S5).

In summary, we find that the balance of germline and somatic total mRNA is a polygenic trait, central to non-genetic variation in aging, and the shared target of all hits from our perturbation screen.

Intrinsic variation in NVM expression is sufficient to generate the non-genetic variation of germline and somatic mRNA in old cohorts

Our results reveal a striking equivalence between the consequences of extrinsic experimental modulation of physiologic networks and the intrinsic non-genetic variation in organismal gene expression in aging. We find that PC1 of the NVM knockdown corpus is highly correlated with PC1 of the wild-type day 8 population (Figures 3I and S3S), meaning that aged individuals vary as if each individual had been exposed to a different, randomly selected NVM knockdown (Figures 3J and 3K). Our data suggest that the natural variation we observe in the expression levels of any single NVM in young adulthood (Figure S3I) is qualitatively sufficient to generate the naturally arising differences in organismal gene expression produced by aging.

NVMs mediate up to half of wild-type lifespan variance

We then investigated whether any NVMs influence lifespan. Out of 18 tested, all NVM knockdowns significantly altered lifespan, with six decreasing and the remainder increasing lifespan (Figures 4A–4C and S4A). For many of the NVM genes, knockdown starting on day 4 was sufficient to extend lifespan, demonstrating that these NVMs act late in life, after the prime reproductive period, to determine wild-type lifespan.

We then asked how NVM genes contribute to lifespan variance. Remarkably, we identify 14 NVMs whose knockdown at the start of adulthood produces large decreases in lifespan variance (Figures 4D–4F and S4B). Notably, the knockdown of aexr-1, nlp-28, and mak-1 reduced lifespan variance by almost half—equivalent to the effect of germline ablation via glp-1(e2141). Knockdown of four targets, including mak-1, was sufficient to reduce lifespan variation started on day 4 of adulthood (Figure S4C). Furthermore, we find that the NVMs whose knockdown produces the largest reductions in lifespan variance also reduce variance in a common measure of organismal health— vigorous movement span (Figures S4D and S4E). Our data highlight a central role for NVMs in generating physiologic heterogeneity during adulthood that leads to inter-individual differences in healthspan and lifespan.

Disruption to germline structure and function is not an inherent part of NVM activity

We then asked whether all our NVMs act like glp-1(e2141) to ablate the germline and eliminate progeny production. Ten out of twenty NVM genes tested significantly decreased (p < 0.05, Wald test) total fecundity when knocked down either from the

(F) Disproportional changes in lifespan variance and mean are quantified using the ratio of variances in the two populations' AFT regression residuals, with bootstrap 95% confidence intervals.

(G) The total number of progeny produced per individual (points) after life-long NVM knockdown.

(H) RPB-2::GFP (yellow) allows rapid in vivo quantification of germline morphology; scale bars indicate 250 μM.

(I and J) (I) We performed automated counting of mitotic and meiotic nuclei to measure (J) the effects of NVM knockdown from the beginning of adulthood on germline morphology, as imaged on day 8; scale bars, 40 μM.

(K) Quantification of the total number of mitotic/early mitotic nuclei present per gonad arm in each individual (points) exposed to NVM knockdown or glp-1(e2141) germline ablation (GA). Stars indicate significance at p < 0.05 in log-linear regression models.

(L) The effect of the same treatments on body size.

(M) Taken together, our data suggest that knockdown of different NVMs (colored circles) similarly alters a physiologic target required for generating half of lifespan variance, as well as the balance of germline and somatic mRNA content, without necessarily altering germline morphology.
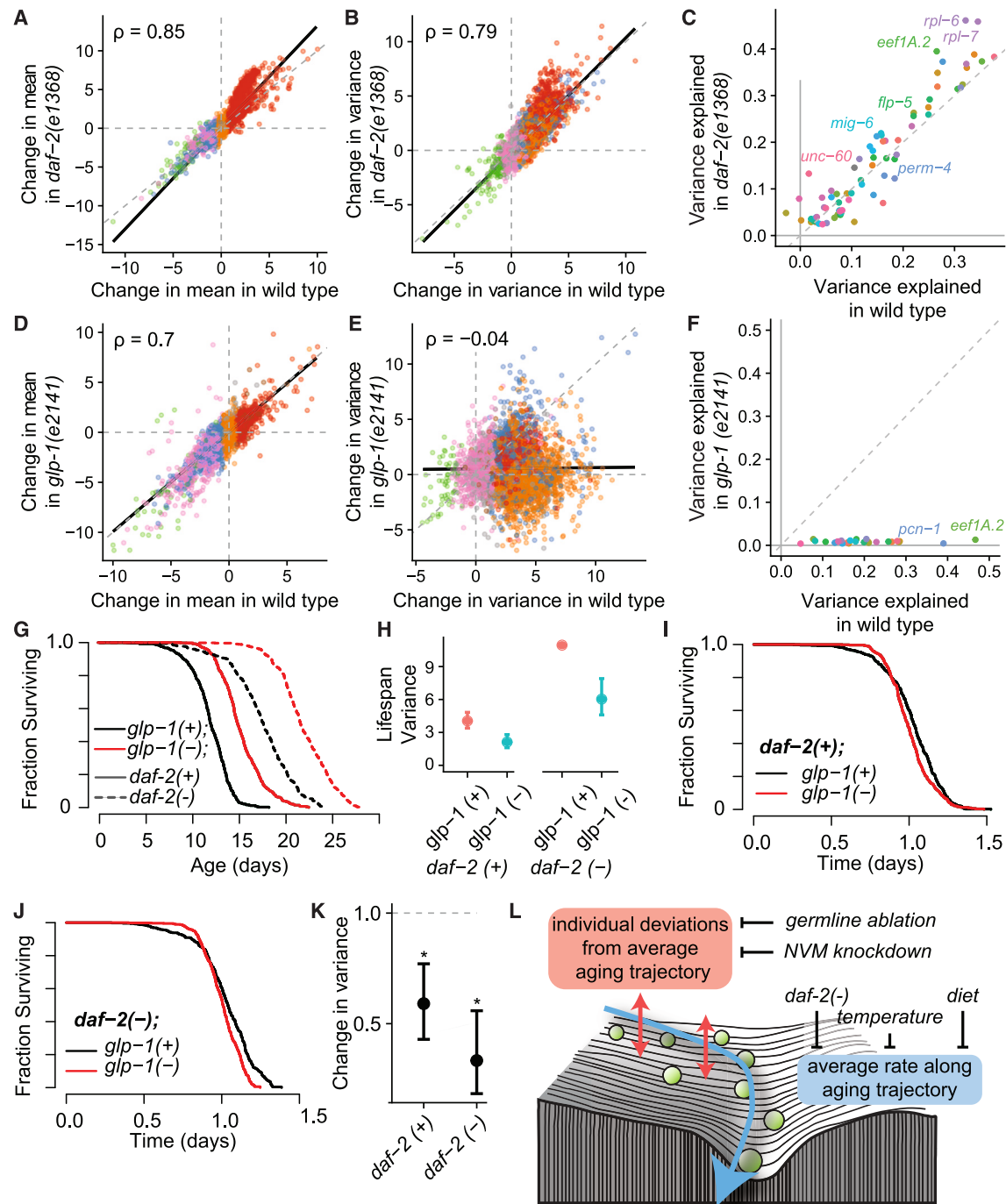
See also Figure S4.

Figure 5. A partial decoupling of lifespan mean and variance

(A and B) (A) The fold-change in gene-expression mean and (B) variance across the transcriptome between days 1 and 8 compared between wild type and insulin/IGF signaling *daf-2(1368)* mutants. Genes (points) are colored according to statistically significant changes with age ($p < 0.001$, Wald test), using the same colors as in Figure 1D.

(C) A transcriptome scaling model (TSM) was estimated comparing each non-genetic variation mimetic (NVMs) to the set of wild-type (x axis) or *daf-2(e1368)* (y axis) single-individual transcriptomes, measured on day 8. The fraction of variance explained by each TSM, compared for *daf-2(e1368)* and wild-type populations.

(D–F) The same as (A–C), but comparing wild-type to *glp-1(e2141)* populations.

(G) Kaplan-Meier estimates for the lifespan of germline-ablated *glp-1(e2141)* (red) and intact *glp-1(+)* (black) individuals either with wild-type DAF-2(+) (solid) or low DAF-2(−) (dashed) insulin/IGF signaling (IIS) achieved by auxin-inducible degradation of DAF-2::AID at 25°C.

(legend continued on next page)

start of adulthood (Figure 4G) or at hatching (Figure S4D). However, nine NVMs had no effect and one, *C08H9.15*, increased brood size. We conclude that, though many NVMs contribute to fertility, this action is not inherent to their contribution to physiologic heterogeneity in aging.

During aging, changes in stem cell activity and number produce dramatic changes in germline morphology.[57] We find that the knockdown of some, but not all, NVMs alters germline morphology. By day 8, knockdown of *pcn-1* almost entirely eliminates meiotic and mitotic cells in the germline, much like *glp-1(e2141)* (Figures 4H–4K). However, the knockdown of other NVMs, including *mak-1*, *col-122*, and *aexr-1*, produced no significant effects on germline cell number. We find that knockdown of *aexr-1*, *vha-11*, and *col-122* produces organismal changes in somatic morphology, increasing individuals' overall body size on day 8 (Figure 4L). Taken together, our data clarify that the influence of many NVMs on non-genetic variation is not mediated by their influence on reproduction (Figure 4M). Rather, the knockdown of several NVMs can dramatically reduce lifespan variance while maintaining a fertile and morphologically intact germline.

Conventional lifespan-extending treatments do not alter inter-individual variation in somatic and germline mRNA content

We then sought to connect the activity of NVMs to a variety of mechanisms previously shown to determine the rate of aging in *C. elegans*. Previous work has suggested that reductions in gene-expression variation might be an important activity of many lifespan-extending treatments,[58] and, to measure this directly, we investigated the effect of four lifespan-altering interventions on gene-expression variation: a lifespan-extending mutation in the insulin/IGF receptor *daf-2(e1368)*,[16,17,59] a lifespan-extending ablation of the germline via *glp-1(e2141)*, a lifespan-shortening increase in body temperature from 20°C to 25°C,[60] and a dietary change: UV irradiation of the bacterial food source.[32]

We find that *daf-2(e1368)* extends lifespan by 32% (95% CI 24%–40%), while producing only small changes to gene-expression mean or variability on day 8 (Figures 5A, 5B, and S5A–S5C). Age-associated changes in per-gene-expression variance in *daf-2(e1368)* showed a 79% correlation with those in wild type (Data S2) and, on day 8, *daf-2(e1368)* produced a significant differential expression in only 6% of genes' mean and 7% of genes' variance (Figure S5D; $p < 0.001$; studentized bootstrap; Data S3). Remarkably, the NVM genes knocked down in wild-type populations explained similar fractions of total non-genetic variation in *daf-2(e1368)* populations, as in wild type (Figures 5C and S5E). Vice versa, NVM knockdown performed in a *daf-2(e1368)* background explained similar fractions of total non-genetic variation of wild-type populations, as in *daf-2(e1368)*

(Figures S5F and S5G). Therefore, we conclude that *daf-2(e1368)* extends lifespan without modifying the influence of NVMs on the age-associated decoupling of germline and somatic mRNA content.

Increases in body temperature and changes in diet produced more substantial changes in gene-expression variance (Figures S5I–S5L and S5N–S5Q) than *daf-2(e1368)* but, again, did not substantially disrupt the relationship between NVM genes and non-genetic variation (Figures S5M and S5R). We therefore conclude that diverse lifespan-extending interventions act without substantially changing non-genetic variation in germline and somatic mRNA content. In stark contrast, germline ablation via *glp-1(e2141)* produced a 20% lifespan extension while dramatically altering variation in the transcriptome. By day 8, germline ablation mostly preserved the age-associated changes in mean expression seen in wild type (Figure 5D) but substantially lowered transcriptome variation (Figures 2F, 5E, and S5H). The NVM genes identified in wild-type populations explained essentially none of the gene-expression variation present in germline-ablated populations (Figure 5F).

Germline ablation makes the aging process less noisy, but *daf-2* knockdown does not

Because *daf-2(e1368)* and *glp-1(e2141)* mutations produced dramatically different effects on gene-expression variation, we then set out to understand how the two mutations interact to determine lifespan variation. To characterize any epistatic interactions between insulin/IGF signaling (IIS) and the germline on lifespan variance, we measured the effects of auxin-mediated DAF-2 degradation[61] in populations with or without a germline. We find that *glp-1(e2141)* extends lifespan and decreases lifespan variance in both IIS conditions (Figures 5G and 5H). Germline ablation disproportionally reduced variance 1.76-fold (95% CI 1.35–2.26) and 2.52-fold (95% CI 1.98–3.32) in wild-type and low IIS backgrounds, respectively (Figures 5I–5K), as quantified by an AFT regression model (STAR Methods). From this, we conclude that the germline functions to generate lifespan variance independently from the context of a population's IIS state.

These lifespan data reveal a striking genetic decoupling of the determinants of lifespan mean and variance. In the demographic literature, a decoupling of lifespan mean and variance is often described as a "frailty" phenomenon[62] in which some random variable Z generates individual deviations from the population hazard function $h(t)$, such that an individual's hazard function becomes $h_i(t) = Zh(t)$. The context-independence of *glp-1(e2141)* effects relative to *daf-2* activity suggests that the germline acts to increase the variance of some Z in contrast to disruption to IIS, which affects a scaling transform[52] of $h(t)$. Although demographic models usually attribute changes in frailty to environmental heterogeneity, our data suggest that the germline acts as an intrinsic biological source of frailty.

(H) Quantification of the variance of each population using Weibull parametric fits.
(I and J) The AFT regression residuals for the same populations as (G), highlighting disproportional changes in lifespan variance.
(K) Estimates of any disproportional effects of *glp-1(e2141)* on lifespan variance relative to mean in both IIS conditions, with bootstrap 95% confidence intervals.
(L) These results suggest a model in which aging individuals move on an aging landscape shaped such that directions of population-average drift (blue) over time and individuals' noisy motion in respect to that drift (red) are influenced by distinct mechanisms.
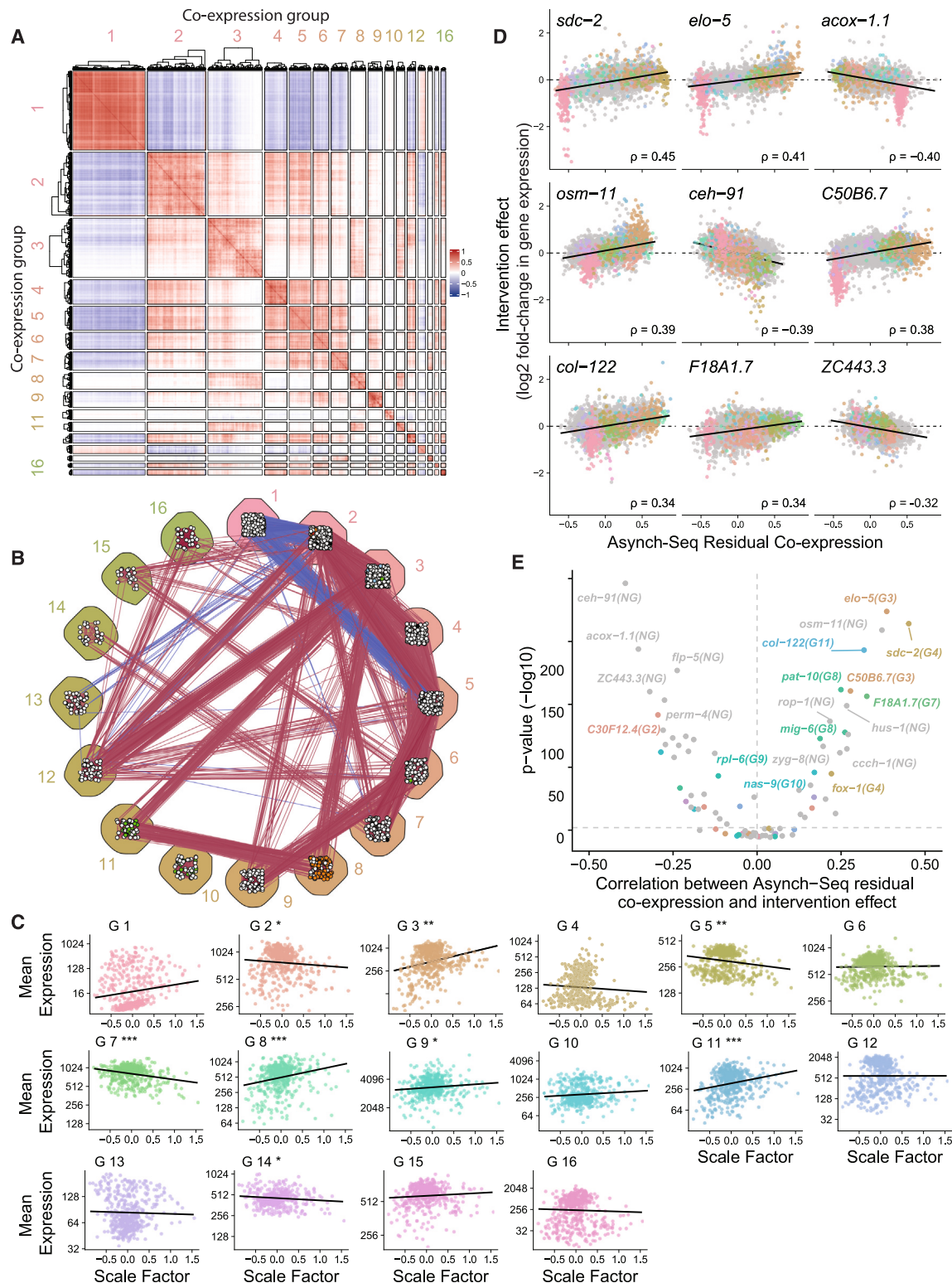See also Figure S5.

Figure 6. Inter-individual heterogeneity within cohorts can be used to predict the effect of interventions on aging

(A) Community analysis identifies 16 robust co-expression groups of genes in the wild-type day 8 cohort, visible after controlling for individual variation in germline and somatic mRNA content. The residual correlations among 1,893 genes are shown, grouped according to genes' co-expression group and hierarchically clustered.

Cell
Article

🌀 CellPress

Our results further allow us to continue previous efforts to describe organismal aging dynamics in respect to a single state variable r that determines lifespan according to $\frac{dr}{dt} = -k_r F(r)$, where $k_r F(r)$ is an unknown function of r that does not depend on the aging rate k_r.[52] The activity of the germline suggests that we add a noise term into this equation: $\frac{dr}{dt} = -k_r F(r, Z)$, where Z is a random variable influenced not only by germline ablation but also by all of the NVMs we identify—in contrast to *daf-2* knockdown, which alters only k_r (Figure 5L). This model places a constraint on any stochastic process proposed to determine lifespan, as its parameters must change in a coordinated fashion in response to changes in k_r and Z. For example, even absent knowledge of the exact form of $F(r, Z)$, we can conclude that any biomarker intended to measure r and to predict remaining lifespan must have its variance in age-synchronous cohorts reduced by any intervention that reduces the variance of Z.

Sixteen additional axes of inter-individual variation

Though our perturbation screen identified many NVMs, we find that they all contribute to a single aspect of non-genetic variation in aging—the balance of somatic and germline mRNA content. To identify any additional axes of non-genetic variation present in aged (day 8) wild-type populations, we pursued a separate analysis strategy, first controlling for individual variation in somatic and germline mRNA content (Methods S1) and then identifying any residual co-expression between genes across the transcriptome (STAR Methods). We identify 16 robust co-expression groups (Figures 6A, 6B, S6A, and S6B; Data S7) in the residual co-expression matrix that contain between 27 (G16) and 403 (G1) genes each. Each group includes a distinct and mostly non-overlapping set of genes that co-vary together between individuals in the aged (day 8) cohort—in effect, a sixteen-dimensional state-space in respect to which old individuals differ (Figure S6C).

Eleven of these 16 co-expression groups emerge progressively throughout adulthood (Figures S6D and S6E)—their cohesion (STAR Methods) begins low and increases with age. The other five groups (G2, G6, G9, and G15) do not exhibit age-associated changes. Eleven groups contain tissue-specific genes from several different tissues, revealing a cross-tissue regulation of gene expression. Most groups contain a large fraction of genes conserved between nematodes and humans—30% of genes across all groups are conserved. Taken together, these analyses highlight the phenomenological richness of aging captured our atlas of non-genetic variation. Furthermore, these

analyes identify, super-imposed on top of global changes in somatic and germline mRNA content, 16 additional axes of non-genetic variation.

Seven out of 16 co-expression groups showed a correlation in their expression between individuals, with the scale factors assigned by NVM models (Figures 6C, S6G, and S6H), suggesting that these co-expression groups have some degree of functional interaction with the mechanisms coupling somatic and germline mRNA content. However, the remaining 9 groups show weak or no correlation with NVM scale factors, highlighting the existence of independent axes of non-genetic variation produced by aging. These co-expression groups can be used as a reference with which to investigate the contribution of specific molecular mechanisms to the non-genetic variation in aging. For example, the enrichment in G3 of targets differentially regulated by *atfs-1*(et18)[63] suggests that differential activation of systemic mitochondrial unfolded protein response signal[64] might contribute to non-genetic organismal heterogeneity in aging, and a similar enrichment in G4 of MPK-1 and PMK-1 targets[63] suggests a variable activation of the ubiquitin-proteasome system[65] among aged individuals.

Membership in expression groups predicts the effect of interventions on aging

To understand the physiologic relevance of the co-expression groups we identified, we returned to perturbation screen data collected earlier. Remarkably, we find that even after normalizing for differences in individuals' germline and somatic mRNA content, co-expression patterns across aged wild-type individuals reflect the gene-expression changes produced by RNAi knockdown of many different targets. Specifically, when considering the expression of two genes, A and B, we find that an RNAi knockdown of gene A produces a larger change in the expression of gene B when A and B are strongly correlated in the aged cohort's residual co-expression matrix (Figure 6D). This correspondence between knockdown effects and co-expression within cohorts was significant for 55 out of the 85 knockdown targets (Figure 6E; STAR Methods). We therefore conclude that gene co-expression data within cohorts reflect physiologically relevant axes of individual variation that mediate, in part, the effects of experimental perturbations to aging.

We note that the correspondences we identify between naturally arising inter-individual differences within aged cohorts and the effect of interventions on aging may be useful for future efforts toward rational intervention design. In aging research,

(B) The same co-expression groups shown as a graph, with genes (small circles) connected by lines indicating positive (red) or negative (blue) residual correlations in expression. Genes are colored according to their tissue-specific expression following the same scheme as in Figure 2D.

(C) In aged (day 8) wild-type individuals (points), we plot the relationship between each co-expression group's (G1–G16) expression and an individuals' balance of somatic and germline mRNA content. This balance is estimated using a transcriptome scaling model (TSM) along the *glp-1(e2141)* axis (STAR Methods). Statistical significance of the log-linear relationship is indicated for each group as *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(D) The residual covariance of gene expression in the day 8 wild-type cohort (x axis) is used to predict (Methods S1) the fold-change in expression (y axis) measured after knockdown of one gene (black label in upper right corner) on all other genes (points). Each gene is colored according to its co-expression group membership in (C) The nine best-predicted RNAi treatments are shown.

(E) For 55 out of 85 RNAi treatments, there is a significant relationship ($p < 0.01$; horizontal line) between Asynch-seq predictions and fold-changes in gene expression measured after knockdown (Bonferroni-corrected p values; Wald test).

See also Figure S6.

interventional experiments are usually time-consuming and costly and any method that can predict in advance the effect of interventions in a systematic way could efficiently direct experimental efforts. The non-genetic variation atlas collected using our approach captures important information that can support such predictions.

DISCUSSION

In animals, lifespan appears to be a quantitative trait influenced by many molecular mechanisms.[16,66,67] In this study, we demonstrate how such mechanisms can be identified using naturally arising stochastic variations in gene expression within isogenic cohorts. We then demonstrate that it is possible to produce dramatic reductions in age-associated physiologic heterogeneity, not by controlling extrinsic factors like diet or environment but rather by manipulating the intrinsic sources of inter-individual differences in aging.

In yeast, age-associated changes in cell physiology result in individuals being canalized into two discrete pathological states.[24] In *C. elegans*, we find that aging works differently: changes across a diverse set of genes result in individuals being canalized onto a single continuous axis of organismal gene expression. Though we do not know yet how physiologic heterogeneity is canalized during human aging, our study suggests a path forward via the systematic measurement and modeling of physiologic heterogeneity. The physiologic axes of inter-individual variability may differ between species—for example, for mice in which VEGF overexpression increases lifespan,[68] it stands to reason that non-genetic variation in any mechanism involved in capillary maintenance may contribute to lifespan variance through the same physiologic interactions altered by VEGF over-expression. However, because a functional interdependence among organs, tissues, and cells is a fundamental architectural feature of all multicellular life, the canalization phenomenon identified here is likely widespread.

The resolving power of Asynch-seq is limited mostly by the population sizes considered and not by the rate of aging. Therefore, Asynch-seq side-steps many of the practical challenges inherent in longitudinal aging studies by recognizing that the effects of aging, slowly accumulated over a lifetime, can be measured and modeled using the physiologic heterogeneity present within a single cross-sectional study. Our work here focuses on gene expression, but the approach can be applied, in principle, to any quantitative high-dimensional measurement, including proteomic and metabolomic methods. We envision that, just as statistical genetics techniques like genome-wide association studies (GWAS) benefit from large-scale population sequencing efforts, Asynch-seq can be scaled up to obtain increasingly resolved mechanistic models of non-genetic variation in aging.

Limitations of the study

Our study relies on gene-expression measurements as a proxy for the physiologic state of an individual. Aspects of physiology that differ between old individuals without measurably influencing organismal gene expression are not considered. Furthermore, our study identifies the intrinsic mechanisms generating

lifespan variance only for populations housed in standard culturing conditions and a few common alternatives, including different temperatures and food sources. We do not know how other environments might influence the relative contributions of genes to lifespan, healthspan, and gene-expression variance, motivating future studies. Finally, our study identifies the mechanistic drivers of only half of wild-type lifespan variance within isogenic populations and leaves the remaining 50% of lifespan variance unexplained, for now.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Caenorhabditis elegans
- METHOD DETAILS
  - mRNA sequencing
  - MARS-Seq
  - Tissue-specific RNA Polymerase II knockdown
  - Perturbation screen
  - Lifespan Assays
  - Fecundity Assay
  - Fluorescence Microscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Image Quantification
  - Quantification of RNA-seq counts
  - Specification of tissue-specific transcripts
  - Normalization of RNA-seq counts and estimation of size factors
  - PCA and Functional Enrichment Analysis
  - Relating single and pooled nematode RNAseq counts
  - Transcriptome Scaling Models (TSMs)
  - Consensus scale factor estimates
  - Estimating residual variance in gene expression
  - Statistical tests for per-gene changes in mean and cohort variance in Gene Expression
  - Weighted correlation network analysis and the identification of a consensus network
  - Statistical tests involving communities
  - Community bootstrap
  - Co-expression group cohesion
  - Analyzing the relationship between residual co-expression and the effect of interventions
  - Analysis of fecundity and lifespan
  - Quantifying lifespan variance

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.05.050.

Cell
Article

CellPress

AUTHOR CONTRIBUTIONS

Conceptualization and experimental design, O.M.F.M., M.E., N.O., and N.S.; statistical methods development and analyses, O.M.F.M., A.S.-P., and N.S.; analytic and simulation studies, O.M.F.M. and N.S.; investigation, M.E., N.O., L.S., S.M.B., A.D.C.-F., and N.S.; training for Smart-Seq2 protocols, H.H. and C.M.; funding acquisition, A.S.-P., H.H., and N.S.; supervision, A.S.-P., H.H., and N.S.; writing, M.E., N.O., L.S., and N.S.

DECLARATION OF INTERESTS

H.H. is co-founder of Omniscope and scientific advisory board member of MiRXES. C.M. is scientific consultant member of GLG.

REFERENCES

1. Belsky, D.W., Caspi, A., Houts, R., Cohen, H.J., Corcoran, D.L., Danese, A., Harrington, H., Israel, S., Levine, M.E., Schaefer, J.D., et al. (2015). Quantification of biological aging in young adults. Proc. Natl. Acad. Sci. USA *112*, E4104–E4110. https://doi.org/10.1073/pnas.1506264112.

2. Nie, C., Li, Y., Li, R., Yan, Y., Zhang, D., Li, T., Li, Z., Sun, Y., Zhen, H., Ding, J., et al. (2022). Distinct biological ages of organs and systems identified from a multi-omics study. Cell Rep. *38*, 110459. https://doi.org/10.1016/j.celrep.2022.110459.

3. Chen, Z., Raj, A., Prateek, G.V., Francesco, A.D., Liu, J., Keyes, B.E., Kolumam, G., Jojic, V., and Freund, A. (2021). Automated, high-dimensional evaluation of physiological aging and resilience in outbred mice. Elife *11*, e72664. https://doi.org/10.1101/2021.08.02.454830.

4. Martineau, C.N., Brown, A.E.X., and Laurent, P. (2020). Multidimensional phenotyping predicts lifespan and quantifies health in Caenorhabditis elegans. PoOS Comput. Biol. *16*, e1008002. https://doi.org/10.1371/journal.pcbi.1008002.

5. Zhang, W.B., Sinha, D.B., Pittman, W.E., Hvatum, E., Stroustrup, N., and Pincus, Z. (2016). Extended twilight among isogenic C. elegans causes a disproportionate scaling between lifespan and health. Cell Syst. *3*, 333–345.e4. https://doi.org/10.1016/j.cels.2016.09.003.

6. Golden, T.R., and Melov, S. (2004). Microarray analysis of gene expression with age in individual nematodes. Aging Cell *3*, 111–124. https://doi.org/10.1111/j.1474-9728.2004.00095.x.

7. Herndon, L.A., Schmeissner, P.J., Dudaronek, J.M., Brown, P.A., Listner, K.M., Sakano, Y., Paupard, M.C., Hall, D.H., and Driscoll, M. (2002). Stochastic and genetic factors influence tissue-specific decline in ageing Caenorhabditis elegans. Nature *419*, 808–814. https://doi.org/10.1038/nature01135.

8. Ruby, J.G., Wright, K.M., Rand, K.A., Kermany, A., Noto, K., Curtis, D., Varner, N., Garrigan, D., Slinkov, D., Dorfman, I., et al. (2018). Estimates of the heritability of human longevity are substantially inflated due to assortative mating. Genetics *210*, 1109–1124. https://doi.org/10.1534/genetics.118.301613.

9. Herskind, A.M., McGue, M., Holm, N.V., Sørensen, T.I., Harvald, B., and Vaupel, J.W. (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. Hum. Genet. *97*, 319–323. https://doi.org/10.1007/BF02185763.

10. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

11. Kawamura, K., and Maruyama, I.N. (2019). Forward genetic screen for *Caenorhabditis elegans* mutants with a shortened locomotor healthspan. G3 (Bethesda) *9*, 2415–2423. https://doi.org/10.1534/g3.119.400241.

12. Hamilton, B., Dong, Y., Shindo, M., Liu, W., Odell, I., Ruvkun, G., and Lee, S.S. (2005). A systematic RNAi screen for longevity genes in *C. elegans*. Genes Dev. *19*, 1544–1555. https://doi.org/10.1101/gad.1308205.

13. Curran, S.P., and Ruvkun, G. (2007). Lifespan regulation by evolutionarily conserved genes essential for viability. PLoS Genet. *3*, e56. https://doi.org/10.1371/journal.pgen.0030056.

14. Derisbourg, M.J., Wester, L.E., Baddi, R., and Denzel, M.S. (2021). Mutagenesis screen uncovers lifespan extension through integrated stress response inhibition without reduced mRNA translation. Nat. Commun. *12*, 1678. https://doi.org/10.1038/s41467-021-21743-x.

15. Nadon, N.L., Strong, R., Miller, R.A., and Harrison, D.E. (2017). NIA Interventions Testing Program: investigating putative aging intervention agents in a genetically heterogeneous mouse model. EBioMedicine *21*, 3–4. https://doi.org/10.1016/j.ebiom.2016.11.038.

16. Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., and Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of \emphCaenorhabditis elegans. Nature *424*, 277–283. https://doi.org/10.1038/nature01789.

17. Podshivalova, K., Kerr, R.A., and Kenyon, C. (2017). How a mutation that slows aging can also disproportionately extend End-of-Life Decrepitude. Cell Rep. *19*, 441–450. https://doi.org/10.1016/j.celrep.2017.03.062.

18. Meyer, D.H., and Schumacher, B. (2021). BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. Aging Cell *20*, e13320. https://doi.org/10.1111/acel.13320.

19. Mendenhall, A.R., Tedesco, P.M., Taylor, L.D., Lowe, A., Cypser, J.R., and Johnson, T.E. (2012). Expression of a Single-Copy hsp-16.2 Reporter Predicts life span. J. Gerontol. A Biol. Sci. Med. Sci. *67*, 726–733. https://doi.org/10.1093/gerona/glr225.

20. Huang, C., Xiong, C., and Kornfeld, K. (2004). Measurements of age-related changes of physiological processes that predict lifespan of Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA *101*, 8084–8089. https://doi.org/10.1073/pnas.0400848101.

21. Kinser, H.E., Mosley, M.C., Plutzer, I.B., and Pincus, Z. (2021). Global, cell non-autonomous gene regulation drives individual lifespan among isogenic C. elegans. eLife *10*, e65026. https://doi.org/10.7554/eLife.65026.

22. Yang, Y., Karin, O., Mayo, A., Song, X., Chen, P., Santos, A.L., Lindner, A.B., and Alon, U. (2023). Damage dynamics and the role of chance in the timing of E. coli cell death. Nat. Commun. *14*, 2209. https://doi.org/10.1038/s41467-023-37930-x.

23. Cohen, A.A., Ferrucci, L., Fülöp, T., Gravel, D., Hao, N., Kriete, A., Levine, M.E., Lipsitz, L.A., Olde Rikkert, M.G.M., Rutenberg, A., et al. (2022). A

complex systems approach to aging biology. Nat. Aging 2, 580–591. https://doi.org/10.1038/s43587-022-00252-6.

24. Li, Y., Jiang, Y., Paxman, J., O'Laughlin, R., Klepin, S., Zhu, Y., Pillus, L., Tsimring, L.S., Hasty, J., and Hao, N. (2020). A programmable fate decision landscape underlies single-cell aging in yeast. Science 369, 325–329. https://doi.org/10.1126/science.aax9552.

25. Serra, L., Chang, D.Z., Macchietto, M., Williams, K., Murad, R., Lu, D., Dillman, A.R., and Mortazavi, A. (2018). Adapting the smart-seq2 protocol for robust single worm RNA-seq. Bio Protoc. 8, e2729. https://doi.org/10.21769/BioProtoc.2729.

26. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smartseq2. Nat. Protoc. 9, 171–181. https://doi.org/10.1038/nprot.2014.006.

27. Bulteau, R., and Francesconi, M. (2022). Real age prediction from the transcriptome with RAPToR. Nat. Methods 19, 969–975. https://doi.org/10.1038/s41592-022-01540-0.

28. Oh, H.S.-H., Rutledge, J., Nachun, D., Pálovics, R., Abiose, O., Moran-Losada, P., Channappa, D., Urey, D.Y., Kim, K., Sung, Y.J., et al. (2023). Organ aging signatures in the plasma proteome track health and disease. Nature 624, 164–172. https://doi.org/10.1038/s41586-023-06802-1.

29. Buckley, M.T., Sun, E.D., George, B.M., Liu, L., Schaum, N., Xu, L., Reyes, J.M., Goodell, M.A., Weissman, I.L., Wyss-Coray, T., et al. (2023). Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. Nat. Aging 3, 121–137. https://doi.org/10.1038/s43587-022-00335-4.

30. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. 37, 547–554. https://doi.org/10.1038/s41587-019-0071-9.

31. Holdorf, A.D., Higgins, D.P., Hart, A.C., Boag, P.R., Pazour, G.J., Walhout, A.J.M., and Walker, A.K. (2020). WormCat: an online tool for annotation and visualization of Caenorhabditis elegans genome-scale data. Genetics 214, 279–294. https://doi.org/10.1534/genetics.119.302919.

32. Garigan, D., Hsu, A.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Kenyon, C. (2002). Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. Genetics 161, 1101–1112. https://doi.org/10.1093/genetics/161.3.1101.

33. Perez, M.F., and Lehner, B. (2019). Vitellogenins - yolk gene function and regulation in Caenorhabditis elegans. Front. Physiol. 10, 1067. https://doi.org/10.3389/fphys.2019.01067.

34. Jiang, N., Cheng, C.J., Gelfond, J., Strong, R., Diaz, V., and Nelson, J.F. (2023). Prepubertal castration eliminates sex differences in lifespan and growth trajectories in genetically heterogeneous mice. Aging Cell 22, e13891. https://doi.org/10.1111/acel.13891.

35. Angeles-Albores, D., Leighton, D.H.W., Tsou, T., Khaw, T.H., Antoshechkin, I., and Sternberg, P.W. (2017). The Caenorhabditis elegans female-like state: decoupling the transcriptomic effects of aging and sperm status. G3 (Bethesda) 7, 2969–2977. https://doi.org/10.1534/g3.117.300080.

36. Luo, S., Kleemann, G.A., Ashraf, J.M., Shaw, W.M., and Murphy, C.T. (2010). TGF-β and insulin signaling regulate reproductive aging via oocyte and germline quality maintenance. Cell 143, 299–312. https://doi.org/10.1016/j.cell.2010.09.013.

37. Perez, M.F., Francesconi, M., Hidalgo-Carcedo, C., and Lehner, B. (2017). Maternal age generates phenotypic variation in Caenorhabditis elegans. Nature 552, 106–109. https://doi.org/10.1038/nature25012.

38. Berman, J.R., and Kenyon, C. (2006). Germ-cell loss extends C. elegans life span through regulation of DAF-16 by kri-1 and lipophilic-hormone signaling. Cell 124, 1055–1068. https://doi.org/10.1016/j.cell.2006.01.039.

39. Flatt, T., Min, K.-J., D'Alterio, C., Villa-Cuesta, E., Cumbers, J., Lehmann, R., Jones, D.L., and Tatar, M. (2008). Drosophila germ-line modulation of insulin signaling and lifespan. Proc. Natl. Acad. Sci. USA 105, 6368–6373. https://doi.org/10.1073/pnas.0709128105.

40. Medawar, P.B. (1951). An Unsolved Problem of Biology: an Inaugural Lecture Delivered at University College London (HK Lewis and Co).

41. Shi, C., and Murphy, C.T. (2014). Mating induces shrinking and death in Caenorhabditis mothers. Science 343, 536–540. https://doi.org/10.1126/science.1242958.

42. Pickett, C.L., Dietrich, N., Chen, J., Xiong, C., and Kornfeld, K. (2013). Mated progeny production is a biomarker of aging in Caenorhabditis elegans. G3 (Bethesda) 3, 2219–2232. https://doi.org/10.1534/g3.113.008664.

43. Shi, C., Booth, L.N., and Murphy, C.T. (2019). Insulin-like peptides and the mTOR-TFEB pathway protect Caenorhabditis elegans hermaphrodites from mating-induced death. eLife 8, e46413. https://doi.org/10.7554/eLife.46413.

44. Moses, E., Atlan, T., Sun, X., Franěk, R., Siddiqui, A., Marinov, G.K., Shifman, S., Zucker, D.M., Oron-Gottesman, A., Greenleaf, W.J., et al. (2024). The killifish germline regulates longevity and somatic repair in a sex-specific manner. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.18.572041.

45. Kaletsky, R., Yao, V., Williams, A., Runnels, A.M., Tadych, A., Zhou, S., Troyanskaya, O.G., and Murphy, C.T. (2018). Transcriptome analysis of adult Caenorhabditis elegans cells reveals tissue-specific gene and isoform expression. PLoS Genet. 14, e1007559. https://doi.org/10.1371/journal.pgen.1007559.

46. Zhang, M.J., Pisco, A.O., Darmanis, S., and Zou, J. (2021). Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. eLife 10, e62293. https://doi.org/10.7554/eLife.62293.

47. Zhang, Y., Cuerdo, J., Halushka, M.K., and McCall, M.N. (2021). The effect of tissue composition on gene co-expression. Brief. Bioinform. 22, 127–139. https://doi.org/10.1093/BIB/BBZ135.

48. Zou, L., Wu, D., Zang, X., Wang, Z., Wu, Z., and Chen, D. (2019). Construction of a germline-specific RNAi tool in C. elegans. Sci. Rep. 9, 2354. https://www.nature.com/articles/s41598-019-38950-8.

49. Beanan, M.J., and Strome, S. (1992). Characterization of a germ-line proliferation mutation in C. elegans. Development 116, 755–766. https://doi.org/10.1242/dev.116.3.755.

50. Austin, J., and Kimble, J. (1987). glp-1 is required in the germ line for regulation of the decision between mitosis and meiosis in C. elegans. Cell 51, 589–599. https://doi.org/10.1016/0092-8674(87)90128-0.

51. Arantes-Oliveira, N., Apfeld, J., Dillin, A., and Kenyon, C. (2002). Regulation of life-span by germ-line stem cells in Caenorhabditis elegans. Science 295, 502–505. https://doi.org/10.1126/science.1065768.

52. Stroustrup, N., Anthony, W.E., Nash, Z.M., Gowda, V., Gomez, A., López-Moyado, I.F., Apfeld, J., and Fontana, W. (2016). The temporal scaling of Caenorhabditis elegans ageing. Nature 530, 103–107. https://doi.org/10.1038/nature16550.

53. Oswal, N., Martin, O.M.F., Stroustrup, S., Bruckner, M.A.M., and Stroustrup, N. (2021). A hierarchical process model links behavioral aging and lifespan in C. elegans. Preprint at bioRxiv. https://doi.org/10.1101/2021.03.31.437415.

54. Cary, M., Podshivalova, K., and Kenyon, C. (2020). Application of transcriptional gene modules to analysis of Caenorhabditis elegans' gene expression Data. G3 (Bethesda) 10, 3623–3638. https://doi.org/10.1534/g3.120.401270.

55. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science 297, 1183–1186. https://doi.org/10.1126/science.1070919.

56. Angeles-Albores, D., Puckett Robinson, C., Williams, B.A., Wold, B.J., and Sternberg, P.W. (2018). Reconstructing a metazoan genetic pathway with transcriptome-wide epistasis measurements. Proc. Natl. Acad. Sci. USA 115, E2930–E2939. https://doi.org/10.1073/pnas.1712387115.

57. Kocsisova, Z., Kornfeld, K., and Schedl, T. (2019). Rapid population-wide declines in stem cell number and activity during reproductive aging in C.

CellPress

elegans. Development *146*, dev173195. https://doi.org/10.1242/dev.173195.

58. Rangaraju, S., Solis, G.M., Thompson, R.C., Gomez-Amaro, R.L., Kurian, L., Encalada, S.E., Niculescu, A.B., Salomon, D.R., and Petrascheck, M. (2015). Suppression of transcriptional drift extends C. elegans lifespan by postponing the onset of mortality. eLife *4*, e08833. https://doi.org/10.7554/eLife.08833.

59. Gems, D., Sutton, A.J., Sundermeyer, M.L., Albert, P.S., King, K.V., Edgley, M.L., Larsen, P.L., and Riddle, D.L. (1998). Two pleiotropic classes of daf-2 mutation affect larval arrest, adult behavior, reproduction and longevity in Caenorhabditis elegans. Genetics *150*, 129–155. https://doi.org/10.1093/genetics/150.1.129.

60. Mair, W., Goymer, P., Pletcher, S.D., and Partridge, L. (2003). Demography of dietary restriction and death in Drosophila. Science *301*, 1731–1733. https://doi.org/10.1126/science.1086016.

61. Venz, R., Pekec, T., Katic, I., Ciosk, R., and Ewald, C.Y. (2021). End-of-life targeted degradation of DAF-2 insulin/IGF-1 receptor promotes longevity free from growth-related pathologies. eLife *10*, e71335. https://doi.org/10.7554/eLife.71335.

62. Aalen, O.O. (1994). Effects of frailty in survival analysis. Stat. Methods Med. Res. *3*, 227–243. https://doi.org/10.1177/096228029400300303.

63. Yang, W., Dierking, K., and Schulenburg, H. (2016). WormExp: a web-based application for a Caenorhabditis elegans-specific gene expression enrichment analysis. Bioinform. Oxf. Engl. *32*, 943–945. https://doi.org/10.1093/bioinformatics/btv667.

64. Calculli, G., Lee, H.J., Shen, K., Pham, U., Herholz, M., Trifunovic, A., Dillin, A., and Vilchez, D. (2021). Systemic regulation of mitochondria by germline proteostasis prevents protein aggregation in the soma of C. elegans. Sci. Adv. *7*, eabg3012. https://doi.org/10.1126/sciadv.abg3012.

65. Ermolaeva, M.A., Segref, A., Dakhovnik, A., Ou, H.-L., Schneider, J.I., Utermöhlen, O., Hoppe, T., and Schumacher, B. (2013). DNA damage in germ cells induces an innate immune response that triggers systemic stress resistance. Nature *501*, 416–420. https://doi.org/10.1038/nature12452.

66. Price, T., and Schluter, D. (1991). On the low heritability of life-history traits. Evolution *45*, 853–861. https://doi.org/10.1111/j.1558-5646.1991.tb04354.x.

67. Rose, M.R. (1994). Evolutionary Biology of Aging (Oxford University Press).

68. Grunewald, M., Kumar, S., Sharife, H., Volinsky, E., Gileles-Hillel, A., Licht, T., Permyakova, A., Hinden, L., Azar, S., Friedmann, Y., et al. (2021). Counteracting age-related VEGF signaling insufficiency promotes healthy aging and extends life span. Science *373*, eabc8479. https://doi.org/10.1126/science.abc8479.

69. Baeriswyl, S., Diard, M., Mosser, T., Leroy, M., Manière, X., Taddei, F., and Matic, I. (2010). Modulation of aging profiles in isogenic populations of Caenorhabditis elegans by bacteria causing different extrinsic mortality rates. Biogerontology *11*, 53–65. https://doi.org/10.1007/s10522-009-9228-0.

70. de Fernandes de Abreu, D.A.F., Caballero, A., Fardel, P., Stroustrup, N., Chen, Z., Lee, K., Keyes, W.D., Nash, Z.M., López-Moyado, I.F., Vaggi, F., et al. (2014). An insulin-to-insulin regulatory network orchestrates phenotypic specificity in development and physiology. PLoS Genet. *10*, e1004225. https://doi.org/10.1371/journal.pgen.1004225.

71. Keren-Shaul, H., Kenigsberg, E., Jaitin, D.A., David, E., Paul, F., Tanay, A., and Amit, I. (2019). MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. Nat. Protoc. *14*, 1841–1862. https://doi.org/10.1038/s41596-019-0164-4.

72. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinform. Oxf. Engl. *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

73. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

74. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

75. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

76. Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. USA *105*, 1118–1123. https://doi.org/10.1073/pnas.0706851105.

77. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

78. Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–1551. https://doi.org/10.1101/gr.121095.111.

79. Stroustrup, N., Ulmschneider, B.E., Nash, Z.M., López-Moyado, I.F., Apfeld, J., and Fontana, W. (2013). The Caenorhabditis elegans lifespan machine. Nat. Methods *10*, 665–670. https://doi.org/10.1038/nmeth.2475.

80. Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: bayesian analysis of single-cell sequencing data. PLoS Comput. Biol. *11*, e1004333. https://doi.org/10.1371/journal.pcbi.1004333.

81. Lun, A.T.L., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. *17*, 75. https://doi.org/10.1186/s13059-016-0947-7.

82. Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640. https://doi.org/10.1038/nmeth.2930.

83. Eling, N., Richard, A.C., Richardson, S., Marioni, J.C., and Vallejos, C.A. (2018). Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data. Cell Syst. *7*, 284–294.e12. https://doi.org/10.1016/j.cels.2018.06.011.

84. Wang, Y., Hicks, S.C., and Hansen, K.D. (2020). Co-expression analysis is biased by a mean-correlation relationship. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.13.944777.

85. Hashimshony, T., Feder, M., Levin, M., Hall, B.K., and Yanai, I. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature *519*, 219–222. https://doi.org/10.1038/nature13996.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli HT115	Caenorhabditis Genetics Center (CGC)	HT115(DE3)
E. coli NEC937 B	Baeriswyl et al.[69]	N/A
Chemicals, peptides, and recombinant proteins		
Betaine, 5M	Sigma-Aldrich	Cat# B0300
Dimethyl sulfoxide	Sigma-Aldrich	Cat# D8418
External RNA Controls Consortium (ERCC) spike-in mix	Invitrogen	Cat# 4456740
Exonuclease I	NEB	Cat# M0293L
5-Fluoro-2′-deoxyuridine	Sigma-Aldrich	Cat# F0503
KAPA HiFi Hotstart ReadyMix	Roche	Cat# KK2602
Levamisole hydrochloride	Panréac AppliChem	Cat# A4341.0010
PEG8000	Merck	Cat# 89510
Potassium 1-Naphthaleneacetate	TCI	Cat# N0006
Proteinase K	NEB	Cat# P8107S
RNasin	Promega	Cat# N2515
Sodium hypochlorite solution (6-14% active chlorine)	Sigma-Aldrich	Cat# 1056142500
SPRI paramagnetic beads	CRG Genomics Unit	N/A
Streptavidin magnetic beads	Pierce	Cat# 88817
Superscript II reverse transcriptase	Invitrogen	Cat# 18064014
T4 ssDNA:RNA ligase	NEB	Cat# E2040S
T7 transcriptase	NEB	Cat# M0204S
Tn5 transposase	CRG Protein Technologies Unit	N/A
Triton X-100	bioPLUS	Cat# 42030000
Tween 20	bioPLUS	Cat# 42030016
Yeast tRNA	Invitrogen	Cat# AM7119
Critical commercial assays		
Quant-iT dsDNA Broad-Range (BR) Assay Kit	Invitrogen	Cat# Q33130
Qubit dsDNA Broad-Range (BR) Assay Kit	Invitrogen	Cat# Q32853
Tapestation High Sensitivity D5000 Reagents	Agilent	Cat# 5067-5593
Tapestation High Sensitivity D5000 ScreenTape	Agilent	Cat# 5067-5592
Deposited data		
C. elegans Wormbase reference genome	release WS265; 2018	https://wormbase.org
Code for this project	Github	https://github.com/nstroustrup/asynch-seq-2024
mRNA sequencing data	This paper	NCBI Sequence Read Archive (SRA) BioProjectID PRJNA1015633
Experimental models: Organisms/strains		
C. elegans wild type N2	Fernandes de Abreu et al.[70]	QZ0 (N2)
C. elegans: Strain: AMP100: ieSi57 [eft-3p::TIR1::mRuby::unc-54 3'UTR] II; rpb-2::degron::3XFlag::GFP III	This study	AMP100
C. elegans: Strain: AMP101: [*eif-3p*::TIR1::mRuby::*tbb-4* 3'UTR + *Cbr-unc-119*(+)] II; *daf-2(syb1177)*(*daf-2*::AID::TEV::3xFLAG) III	This study	AMP101

(Continued on next page)

Cell
Article

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
C. elegans: Strain: AMP116: *ieSi57* [*eif-3p*::TIR1::mRuby::*tbb-4* 3'UTR + *Cbr-unc-119*(+)] II; *daf-2(syb1177)(daf-2*::AID::TEV::3xFLAG) II; *glp-1(e2141)* III	This study	AMP116
C. elegans: Strain: AMP254: *rpb-2*[*rpb-2*::EGFP::AID::*3xFLAG(ohm50)*] III; *glp-1(e2141)* III; ieSi57 [*eft-3p*::TIR1::mRuby::*unc-54* 3'UTR + *Cbr-unc-119*(+)] II	This study	AMP254
C. elegans: Strain: CB4037: *glp-1(e2141)*	CGC	CB4037
C. elegans: Strain: CER510: *rpb-2*::GFP(*dpiRNA*)::AID::3xFLAG)	CGC	CER510
C. elegans: Strain: DCL569: *mkcSi13* [*sun-1p::rde-1::sun-1* 3'UTR + unc-119(+)] II; *rde-1(mkc36)*	CGC	DCL569
C. elegans: Strain: QZ120: *daf-2(e1368)*	Fernandes de Abreu et al.[70]	QZ120
C. elegans: Strain: SS104: *glp-4(bn2)*	CGC	SS104
Oligonucleotides		
Biotinylated poly-dt primer: /5Biosg/AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN	This study and Picelli et al.[26]	N/A
Index 1 (i7) adapter primers: CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG	Illumina	N/A
Index 2 (i5) adapter primers: AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC	Illumina	N/A
Mars-seq2 1st RT primer: CGATTGAGGCCGGTAATACGACTCACTATAGGGGCGACGTGTGCTCTTCCGATCT[BC$_7$][UMI$_8$]T$_{20}$N	Keren-Shaul et al.[71]	N/A
Mars-seq2 2nd RT primer: CTACACGACGCTCTTCCGATCT	Keren-Shaul et al.[71]	N/A
Mars-seq2 barcode plate ligation adaptor: /5Phos/[BC$_9$]AGATCGGAAGAGCGTCGTGTAG/3SpC3/	Keren-Shaul et al.[71]	N/A
Mars-seq2 T7 Promoter: TAATACGACTCACTATAGGG	Keren-Shaul et al.[71]	N/A
Smart-seq2 IS primer: AAGCAGTGGTATCAACGCAGAGT	Picelli et al.[26]	N/A
Smart-seq2 template switching oligo: AAGCAGTGGTATCAACGCAGAGTACATrGrG+G	Picelli et al.[26]	N/A
Software and algorithms		
BEDTools	Quinlan and Hall[72]	https://bedtools.readthedocs.io/en/latest/
CellPose	Stringer et al.[73]	https://www.cellpose.org/
DESeq2	Love et al.[74]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
featureCounts version 2.0.0	Liao et al.[75]	https://subread.sourceforge.net/featureCounts.html
InfoMap	Rosvall and Bergstrom[76]	https://www.mapequation.org/
STAR version 2.6.0c	Dobin et al.[77]	https://www.encodeproject.org/software/star/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Nicholas Stroustrup (nicholas.stroustrup@crg.eu).

Materials availability

All nematode strains used in this study are available from the Caenorhabditis Genetics Center (CGC) and are available from the authors upon request.

Data and code availability

- Transcriptomic data is available from the NCBI Sequence Read Archive (SRA) under BioProjectID PRJNA1015633.
- Gene lists and model parameters are provided as supplemental data files. Analysis source code is provided in Data S8, with updated versions available from a github repository published at https://github.com/nstroustrup.
- Any additional information required to reanalyze the data reported is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Caenorhabditis elegans

The wild-type parent used in this study is the *C. elegans* Bristol strain N2 (QZ0). The relevant mutations used in this study are: QZ120: *daf-2(e1368)*, CB4037: *glp-1(e2141)*, SS104: *glp-4(bn2)*, Germline-specific RNAi line DCL569: *mkcSi13* [*sun-1p::rde-1::sun-1* 3'UTR + unc-*119(+)*] II; *rde-1(mkc36)*, AMP116: *ieSi57* [*eif-3p*::TIR1::mRuby::*tbb-4* 3'UTR + *Cbr-unc-119*(+)] II; *daf-2(syb1177)(daf-2*::AID::TEV::3xFLAG) II; *glp-1(e2141)* III, AMP100: ieSi57 [eft-3p::TIR1::mRuby::unc-54 3'UTR] II; rpb-2::degron::3XFlag::GFP III, AMP101: [*eif-3p*::TIR1::mRuby::*tbb-4* 3'UTR + *Cbr-unc-119*(+)] II; *daf-2(syb1177)(daf-2*::AID::TEV::3xFLAG) III, GFP-tagged RNA Polymerase II line CER510: *rpb-2*::GFP(*dpiRNA*)::AID::3xFLAG) and AMP254: *rpb-2*[*rpb-2*::EGFP::AID::*3xFLAG(ohm50)*] III; *glp-1(e2141)* III; ieSi57 [*eft-3p*::TIR1::mRuby::*unc-54* 3'UTR + *Cbr-unc-119*(+)] II. Unless otherwise specified, nematodes were maintained on NGM agar plates seeded with live NEC937 B (OP50 ΔuvrA; KanR)[69] at 20 °C. Populations were synchronized using hypochloride treatment. For auxin-inducible degradation experiments, populations were treated with 500uM Potassium 1-Naphthaleneacetate. RNA-interference constructs were obtained from the Ahringer library.

METHOD DETAILS

mRNA sequencing

To obtain an accurate and cost-effective method to capture population-scale gene expression variation, we started with Smart-Seq2 mRNA sequencing,[26] a technology recently explored for organismal measurements.[25] To optimize this method for population-scale nematode studies, we needed to alter the protocol as in *C. elegans* Smart-Seq2 cDNA libraries we found that approximately 80% of transcripts map to a 2300bp fragment of the 3509bp 26s rRNA rrn-3.1 (Figure S1A) in the genomic region I:15064838-15068346. This region terminates in a cytosine triplet—absent in mammalian genomes—which we realized had the potential to directly interact with the ribonucleotide guanosine triplet present in Smart-Seq2 template-switching oligomers (TSO). We confirmed the interaction by showing that rrn-3.1 was amplified by the Smart-Seq2 protocol even in the absence of poly-dT primers (Figure S1B). To eliminate this aberrant TSO priming, we modified the Smart-Seq2 protocol by introducing biotinylated poly-dT primers, allowing subsequent purification of mRNA and elimination of TSO-primed cDNA molecules. Our approach increased the effective read depth of *C. elegans* Smart-Seq2 libraries by 4.57 fold (Figure S1C; Methods S1), dramatically reducing the sequencing cost in preparation for population-scale experimentation. The purification also eliminates any *E. coli* derived transcripts which also lack poly-A tails. Our approach did not alter the fidelity of Smart-Seq2, as we saw only 293 transcripts (adjusted *p*-value < 0.001) (Figure S1D) differentially expressed between the standard and modified protocols, twenty-one of which were residual non-polyadenylated RNAs. The remaining 272 transcripts were polyadenylated mRNAs present in our libraries generated in the absence of poly-dT primers (Figure S1E), suggesting these transcripts were also affected by aberrant TSO priming in the standard protocol.

Lysis buffer was prepared according to published Smart-Seq2 protocols,[25,26] with the addition of ERCC spike-ins[78] to a final dilution of 1:40,000. Nematodes were individually picked into lysis buffer to achieve a ratio of one individual per 8 μL of buffer for individual transcriptomics and 30 individuals per 120 μL of buffer for population transcriptomics. In the latter case, four replicates were performed. Nematode suspensions were shock-frozen in liquid nitrogen and stored at -80 °C upon lysis, which was carried out for 10 min at 65 °C, followed by enzyme inactivation for 5 min at 85 °C.

On a DynaMag-96 Side Skirted magnetic plate (Invitrogen), 4 μL of streptavidin magnetic beads (Pierce) per sample were washed three times in 20 μL binding and washing buffer (5 mM Tris-HCl ph 7.5, 0.5 mM EDTA, 1 M NaCl) followed by two washes in 20 μL Solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 50 mM NaCL) and one wash in 20 μL Solution B (DEPC-treated 0.1 M NaCl). The beads were resuspended in 4 μL Solution B, 0.2 μL biotinylated poly-dT primer (100 μM; IDT) was added and the beads were incubated for 15 minutes at room temperature. The beads were blocked with the addition of 0.4 μL yeast tRNA (10 μg/μL; Invitrogen) and incubated for 15 minutes at room temperature. The supernatant was removed and the beads were resuspended in 4 μL Solution B with the addition of 0.1 μL RNasin (Promega). Then, 4 μL nematode lysate was added and mRNA was hybridized for 15 minutes at room temperature. The mRNA containing magnetic beads were washed once with 20 μL 2.5x SuperScriptII first strand buffer (Invitrogen) before cDNA libraries were prepared following the Smart-Seq2 protocol.[26] cDNA libraries were purified using in-house prepared SPRI paramagnetic beads that mimic AMPure XP beads (Beckman Coulter) with a bead to sample ratio of 0.8. The size

distribution of the resulting libraries were evaluated using a Tapestation 4150 (Agilent). cDNA concentration was determined by Quant-it (Invitrogen) on a plate reader (Tecan).

Nextera sequencing libraries were prepared from cDNA through tagmentation and subsequent PCR amplification with indexing primers as described in the Nextera DNA library prep protocol (Illumina). Nextera sequencing libraries were purified twice using in-house prepared SPRI paramagnetic beads that mimic AMPure XP beads (Beckman Coulter) with a bead to sample ratio of 0.9. The size distribution of the resulting libraries were evaluated using a Tapestation 4150 (Agilent). Library concentration was determined by Quant-it (Invitrogen) on a plate reader (Tecan). Nematode RNAseq libraries were pooled in equal masses. Paired-end Nextera libraries were sequenced on an Illumina NextSeq 500 with high-output 75 cycles v2.5 kits (Illumina) and read lengths of 38 bases at $> 2.0 * 10^6$ reads per sample.

MARS-Seq

A population of 96 individuals were lysed. Each individual's lysate was split in half and one half was purified using the poly-dT bead purification with an additional wash in 20 µL low-salt buffer, followed by elution in 4 µL of 10mM TRIS pH 7.5. The other half of each lysate was left unprocessed. Each single-worm lysate processed by either method was distributed into 4 different wells across two 384-well capture plates containing 2 µL of lysis solution: 0.2% Triton and RNAse inhibitors plus barcoded poly-dT reverse-transcription (RT) primers for RNAseq using a Bravo automated liquid handling platform (Agilent).

Single worm RNA libraries were prepared as previously described for the MARS-seq2 scRNA-seq protocol.[71] First, mRNA was converted into cDNA with an oligo containing both the unique molecule identifiers (UMIs) and cell barcodes. 0.15% PEG8000 was added to the RT reaction to increase efficiency of cDNA capture. Unused oligonucleotides were removed by Exonuclease I treatment. cDNAs were pooled (each pool representing the original 384-wells of a MARS-seq plate) and linearly amplified using T7 in vitro transcription (IVT) and the resulting RNA was fragmented and ligated to an oligo containing the pool barcode and Illumina sequences, using T4 ssDNA:RNA ligase. Finally, RNA was reverse transcribed into DNA and PCR amplified. The size distribution and concentration of the resulting libraries were calculated using a Tapestation 4150 (Agilent) and Qubit (Invitrogen). Single-worm RNAseq libraries were pooled at equimolar concentration and sequenced to saturation (>=6 reads/UMI, in most cases >10 reads/UMI) on an Illumina NextSeq 500 sequencer and using high-output 75 cycles v2.5 kits (Illumina).

Tissue-specific RNA Polymerase II knockdown

RNA knockdown of *ama-1* and *rpb-2* was started on day 1 of adulthood and the effects were measured on day 2 and again on day 8.

Perturbation screen

To study extrinsic sources of variation, we collected the gene-expression data describing the organismal effects of changes in temperature—moving individuals from 20 to 25 °C— and changes in diet including complete deprivation of food starting on day 2 of adulthood and exposure to a lifespan-extending diet of UV-inactivated bacteria. To study intrinsic sources of variation, we selected 94 genes for RNAi knockdown from our gene-expression variation atlas that showed a high-connectivity in the co-expression graph (Statistical Methods), hypothesizing that these genes' expression variability early in life might subsequently generate phenotypic variation later in life. To this corpus of different gene-expression profiles, we also added a canonical lifespan extending intervention in *C. elegans*, disruption of insulin/IGF signaling via the *daf-2(e1368)* allele, to link our results to previous research.

We exposed individuals to 104 different treatments, either at the start of life (*daf-2(e1368)* and *glp-1(e2141)*) or the start of adulthood (all RNAi knockdowns, changes in diet and temperature). We then allowed populations under treatment until day 8 of adulthood at which point we harvested mRNA. In our treatment corpus we also included various "pseudo-time" axes that describe the population-average change in gene expression between the following days: 1-2, 2-4, 4-6, 6-8, 8-10, 10-12, 1-4, 2-8, 4-6, 6-10, 8-12.

Lifespan Assays

All lifespan assays were performed either by hand or using the lifespan machine automated imaging platform,[53,79] on NGM agar plates seeded with the OP50-derived NEC937 or HT115 and 10 µg ml^{-1} 5-fluoro-2-deoxyuridine (FUDR). Experiments performed on UV-inactivated bacteria used 27.5 µg ml^{-1} FUDR. Figures 2H–2K: NEC937 at 25 °C via automated imaging. Figures 4A–4C: HT115 at 20 °C; individuals were placed on empty vector (EV) RNAi as eggs and transferred to the specified RNAi at either L4 stage on day 4 of adulthood as specified. Lifespan data was collected across four replicates, the first done by hand and the remainder via automated imaging. Figures 5G–5K. HT115 at 20 °C via automated imaging. *daf-2(-)* indicates AMP101 populations transferred onto agar plates containing the auxin analog 500uM Potassium 1-Naphthaleneacetate.

Fecundity Assay

Individuals developed together on agar plates seeded with HT115 EV. 12 individuals were exposed to each RNAi condition by separate transfer at L4 onto new agar plates seeded with HT115 carrying the respective RNAi construct, and allowed to lay eggs. Each individual was transferred to a fresh plate every 12 hours until egg laying stopped. The number of live progeny produced by each individual during each 12-hour interval was obtained by counting the number of larva on each plate.

CellPress

Cell
Article

Fluorescence Microscopy

To measure body size and germ cell nuclei number, nematode populations were synchronized via egg lay-off on NEC937 or HT115 live bacteria and transferred onto 40.6 μM 5-fluoro-2′-deoxyuridine NGM plates at the late L4 stage. Nematodes were immobilized in M9 buffer using 10mM levamisole hydrochloride and then imaged on 3% agarose pads. Three-dimensional tiled confocal images were acquired on the LSM980 inverted microscope (Zeiss) equipped with the Airyscan2 detector (Zeiss). Whole-body and gonad masks were drawn to select regions of interest for quantification. Germ cell nuclei were identified using the GFP signal from endogenously tagged RPB-2::GFP.

QUANTIFICATION AND STATISTICAL ANALYSIS

Image Quantification

Fluorescence microscopy image analysis was performed in Python and nuclear segmentation was performed using CellPose.[73]

Quantification of RNA-seq counts

RNA-seq reads were aligned using STAR version 2.6.0c[77] to the *C. elegans* Wormbase reference genome (release WS265) modified to include ERCC spike-ins sequences. Gene counts were quantified using featureCounts version 2.0.0.[75] The count matrix was subject to a detection threshold of 5 counts in at least half of the samples. Genome coverage of reads was computed using BEDTools.[72]

Specification of tissue-specific transcripts

Genes unique to somatic tissues (i.e. hypodermis, neurons, muscles, intestine) were obtained from published data.[45] Somatic unique genes are defined by the authors as "strongly expressed (logRPKM > 5) and significantly differentially expressed in comparison to the expression in each of the three other tissues (FDR <= 0.05, logFC > 2 for each comparison)". Genes unique to the germline were defined using the predicted tissue expression (normalized SVM) score from the same reference. We defined germline-specific genes as those with a predicted tissue expression score of more than 0.5 in the germline and less than 0.0 in every somatic tissue: hypodermis, neurons, muscles, intestine.

Normalization of RNA-seq counts and estimation of size factors

The mean read counts of a gene i and a sample j are modeled as the product of the true mean gene expression μ_i, a technical factor ν_j and a size factor ϕ_j. The technical factor is captured by sample differences present in ERCC spike-ins. The size factor accounts for differences in the RNA content between samples. Most often, normalization of RNA-seq counts does not distinguish between the technical and the size factor and simply estimates a normalization factor κ_j for biological transcripts. This will correspond to the product of the size factor and the technical factor. This is the same model as the model described by Vallejos et al.[80]

$$\log E[Y_{ij}] = \begin{cases} \phi_j \nu_j \mu_i = \kappa_j \mu_i & \text{(mRNA)} \\ \nu_j \mu_i & \text{(synthetic spike} - \text{in)} \end{cases}$$

Normalization factors and technical factors were estimated with scran[81] using the filtered count matrix. For technical factors $\widehat{\phi}_j$, we use the subset of the matrix corresponding to ERCC spike-ins. For normalization factors $\widehat{\kappa}_j$, we only use biological sequences. For normalization factors specific to tissues, we used the subset of the matrix corresponding to tissue unique genes. To obtain an estimate for the size factor ϕ_j, we compute the ratio $\widehat{\nu}_j = \widehat{\kappa}_j / \widehat{\phi}_j$.

We note, as an aside, that our multi-size-factor approach can be further generalized to analyze not only tissue-specific transcripts but the entire transcriptome because in principle the relative proportion of any mRNA expressed in each tissue can be estimated based on the gene's partial correlations with each tissue-specific scale-factor (Figure S2; Methods S1). Using the current data, such an approach functions well considering only the separation of germline and somatic transcripts (Figure S2) but performance is limited in distinguishing expression among somatic tissues due to the lower observable variance among those tissues. Furthermore, such analyses require the assumption that tissue-specific transcriptomic changes are somehow representative of global changes of all transcripts in each tissue—including those expressed across many tissues. Therefore, we focus our current efforts on the tissue-specific transcriptome.

PCA and Functional Enrichment Analysis

Principal component analysis was performed using log counts with a pseudo-count of 1 added and the prcomp procedure implemented in R with centering and scaling. Hierarchical clustering of samples or genes was performed using Spearman's correlation distances and the Ward algorithm. Enrichment analysis was done by using a Fisher exact test using annotations provided by WormCat,[31] WormExp[63] and WormMine. All p-values were adjusted for multiple testing using Benjamini-Hochberg.

Relating single and pooled nematode RNAseq counts

In Note S1, we derive an approximation for the distribution of read counts in pooled samples given that individual samples are distributed according to a negative binomial distribution. This allows us to compute the probability of observing a certain number of read

Cell
Article

CellPress

counts given our estimated parameters on individual nematodes and the number of pooled nematodes. We compute the probability of counts Y being 50% away from the true mean of the gene: $\Pr\left(y \notin \left[\frac{1}{1.5}\mu, 1.5\mu\right] \mid \mu, \delta, n\right)$

From this, we obtain one probability for each gene and pool size, that is, each set of parameters $\{\mu, \delta, n\}$. If this value is high, it means that there is a high probability that we observe read counts that are not a good estimate of the population mean. This suggests we should increase the number of pooled nematodes. To evaluate the quality of our population mean estimates, we considered the proportion of genes expected to fall outside of the 1.5 μ boundary, and compared this theoretic prediction with the empiric data obtained from pooled populations. This model also allows us to predict the pooled population size above which diminishing returns are encountered in estimating population mean gene expression—defined as the pool size above which the above probability does not increase substantially.

Transcriptome Scaling Models (TSMs)

Let w_i^k be the transcript abundance of gene k measured in individual i, taken from a population of N individuals each with G genes total. Let y_i^k be the transcript abundance of gene k in individual i, relative to the population mean such that $\log \overrightarrow{y_i} = \log \overrightarrow{w_i} - \log \frac{1}{N} \sum_j^N \overrightarrow{w_j}$.

Let $\overrightarrow{X_R}$ and $\overrightarrow{X_I}$ be the average gene expression count vectors measured in two populations, the first being a reference population (e.g. the control group) and the second describing a population exposed to some intervention. The fold-change effect of the intervention in gene expression is then $\log \overrightarrow{D} = \log \overrightarrow{X_I} - \log \overrightarrow{X_R}$.

TSM analysis estimates scale parameters λ_i that relate individual differences in gene expression $\overrightarrow{y_i}$ to the effect of interventions $\overrightarrow{D}$ by solving the equation $\log \overrightarrow{y_i} = \lambda_i \log \overrightarrow{D} + \epsilon_i$. Estimates of λ_i are obtained by minimizing the covariance $\mathrm{cov}(\log \overrightarrow{D}, \epsilon_i)$ across genes via the Brent–Dekker method, in R. The fraction of variance explained by the TSM is then calculated from the variance across individuals of the residuals of each gene g relative to the variance across individuals of the transcript abundance of g, $F = \frac{1}{G} \sum_{g=1}^{G} \frac{\mathrm{Var}(e^{g})}{\mathrm{Var}(y^g)}$. Because empirical variance estimates are sensitive to outliers, we also consider a more robust estimator based on 90-10 inter-quantile distance: $F^* = \frac{1}{G} \sum_{g=1}^{G} \frac{\mathrm{IQ}_{90}(e^{g})}{\mathrm{IQ}_{90}(y^g)}$

In many cases, intervention effects $\overrightarrow{D}$ do not explain any inter-individual variation in gene-expression, and the λ's fit for each individual will have no effect on or increase the variance of ϵ_g, leading to fractions F equal to or less than 0. In some plots, we crop F values less than one to zero, as interventions cannot explain less variance than zero. 95% confidence intervals on F estimates were calculated by running bootstrap replicates over subsets of both genes and individuals.

Consensus scale factor estimates

To obtain a "consensus" scale factor that describes each individual's location along a consensus scale defined in common by all non-genetic variation mimetics (NVMs), we calculated the average scale factor for each individual estimated by TSM using each target as a reference scale.

Estimating residual variance in gene expression

Negative binomial mean μ and overdispersion δ parameters were estimated via maximum-likelihood in R using the MASS package. Cohort variance was then derived according to $\sigma^2 = \delta \mu^2 + \mu$. To estimate the mean-variance confounding relationship of the underlying sequencing method,[82] we developed an approach based on residual overdispersion,[83] but focused on estimating the running minimum as a function of mean expression. All mRNAs and spike-ins were ordered by their mean expression and binned together into groups of approximately 200. The minimum variance was estimated for each bin, and these minimum were fit with a polynomial spline. Such spline fits provide continuous estimates of the minimum variance produced by technical noise as a function of mean expression. The residual cohort variance of each transcript is then defined as the difference between the transcript's measured variance and the minimum variance at that transcript's mean expression. Confidence intervals on mean and variance estimates are obtained by running bootstrap replicates of full variance calculation including minimum variance estimation across subsets of individuals.

Statistical tests for per-gene changes in mean and cohort variance in Gene Expression

To estimate differences in the mean and variance of each gene between two sets of single-individual transcriptomes, we repeat the full residual variance estimation procedure across 250 bootstrap replicate populations. p-values are calculated using a two-sided studentized bootstrap method. For differential expression analysis of standard batch samples, we obtain p-values using the Wald test on coefficients estimated using a negative binomial generalized linear model implemented by DESeq2,[74] with p-values adjusted for multiple hypothesis testing using the Benjamini-Hochberg method.

Cell
Article

Weighted correlation network analysis and the identification of a consensus network

For each genotype, we compute the Spearman's correlation between genes unique to tissues using tissue-specific normalized read counts. We observed a relationship between the mean counts and the squared gene-gene correlation previously described in bulk and single-cell RNA-seq data.[84] To eliminate low-count biases, we exclude from analysis genes with mean read counts lower than 30. Analysis focusing on genes primarily expressed in single tissue types focused on 658 well-measured genes. For the whole-transcriptome analyses, between 3,000 and 4,000 transcripts typically passed the threshold in each sample.

We defined a co-expression network such that edges corresponded to squared Spearman correlation values. The square correlation metric lets us group genes closely in the network based on the strength of their interaction, irrespective of whether that interaction involves an activating or inhibitory action. To identify communities of genes that are present in either wild-type or *daf-2(e1368)* genetic backgrounds, we calculate graphs for each genotype separately and then take the union of both graphs. Communities were identified in this graph using the InfoMap algorithm[76] as implemented by the igraph R package. We excluded communities smaller than 10 genes.

Our approach to network analysis requires we choose a minimum threshold for gene-pair Spearman's correlations to be included in the network analysis. To identify communities robust to this choice of threshold, we tested a range of thresholds between 0.5 and 0.8 in increments of 0.05. We performed network community analysis at each threshold, and then counted the number of times each gene-pair appeared in the same co-expression group across the different correlation thresholds. These counts formed a square, symmetric matrix with each element i,j taking values ranging from 1 (genes i,j always clustered together) to 0 (genes i,j never clustered together). We then clustered this matrix using Ward's algorithm to identify communities of genes stable across a range of different correlation thresholds. It is worth noting that a subset of genes were identified as not robustly clustering in a single co-expression group across different correlation thresholds, so we removed these genes from consideration.

Statistical tests involving communities

To compute confidence intervals of the squared correlations between genes within each community of the consensus network, we perform 500 bootstraps of normalized read counts and estimate the empirical cumulative distribution function (ECDF) of the lower triangle of the squared correlation matrix. We summarize these bootstrapped ECDFs by taking the 1% and 99% quantile. During bootstrapping, we applied a gene-pair Spearman's correlation threshold of 0.6.

To statistically test if a community present in the consensus network is present in a particular population, we estimate the null distribution by randomly selecting genes, selecting the same number of genes in that community. We then use the Kolmogorov–Smirnov statistic to compare the observed ECDF with the ECDF corresponding to the null distribution to obtain a p-value.

To statistically test if correlations of one community are stronger in one condition compared to another, we use a Wilcox test on the squared correlations of the lower triangle of the squared correlation matrix.

Community bootstrap

To test the stability of communities, we run the community detection pipeline on 1000 sets of bootstrapped counts. For every pairwise combination of genes, we compute the proportion of times that two genes are in the same community across bootstraps. We also compare bootstrapped communities and the communities obtained using the original dataset using the maximum Jaccard index.

Co-expression group cohesion

To quantify the degree to which a set of genes all co-express together between individuals in a cohort, we calculate a quantity we call their "cohesion": defined as the mean across all pairs of genes of the Pearson correlation of each pair's expression across individuals. In other words, the cohesion of a set measures the average correlation in expression across a population of any two genes in that set.

Analyzing the relationship between residual co-expression and the effect of interventions

To identify significant correlations between the residual co-expression between two genes and the fold-change response of one gene in respect to knockdown of the second, we fit a generalized linear model with a log link function and normal residuals using *glm* in R. P-values were calculated using the Wald test.

Analysis of fecundity and lifespan

The effect of each RNAi on fecundity was estimated using a simple linear regression model with p-values obtained using the Wald test. The effect of each RNAi on lifespan was estimated using Accelerated Failure Time (AFT) models in R using the Buckley-James implementation provided by the RMS package. p-values were obtained form the Wald test. In Figures 4C and 4F, multiple replicate experiments were performed and in this case RNAi effect and replicate batch effects were modeled as separate covariates in the multivariable regression AFT model: $\log(y_i) = \beta_{RNAi} X_i + \beta_{batch} z + \varepsilon_i$. Device-corrected AFT residuals used for subsequent statistical analysis were calculated following the approach of Stroustrup et al.[52]

Cell
Article

∂ **CellPress**

Quantifying lifespan variance

Lifespan distributions were fit with Weibull with frailty models using flexsurv in R as described in Stroustrup et al.[52] Estimates of lifespan mean μ and variance σ^2 were obtained from Weibull shape α and scale β parameters according to $\mu = \beta \Gamma(1 + 1/\alpha)$ and $\sigma^2 = \beta^2 (\Gamma(1 + 2/\alpha) - \Gamma(1+1/\alpha)^2)$ with Γ as the gamma function. Separately, to acquire non-parametric estimates of disproportional difference in lifespan variance between two populations that differ in mean lifespan, the residuals of an Accelerated Failure Time (AFT) model were compared across 200 bootstrap replicates—repeating the full AFT regression on each bootstrap sample population and

calculating the ratio of the variances between groups a and be according to $\Delta\sigma^2 = \frac{\sigma_b^2}{\sigma_a^2} = \frac{\sum_1^{N_b} (y_i - \mu_b)^2 / N_b}{\sum_1^{N_a} (y_i - \mu_a)^2 / N_a}$ with N_a and N_b as the size and

μ_a and μ_b as the means of each population.

Supplemental figures
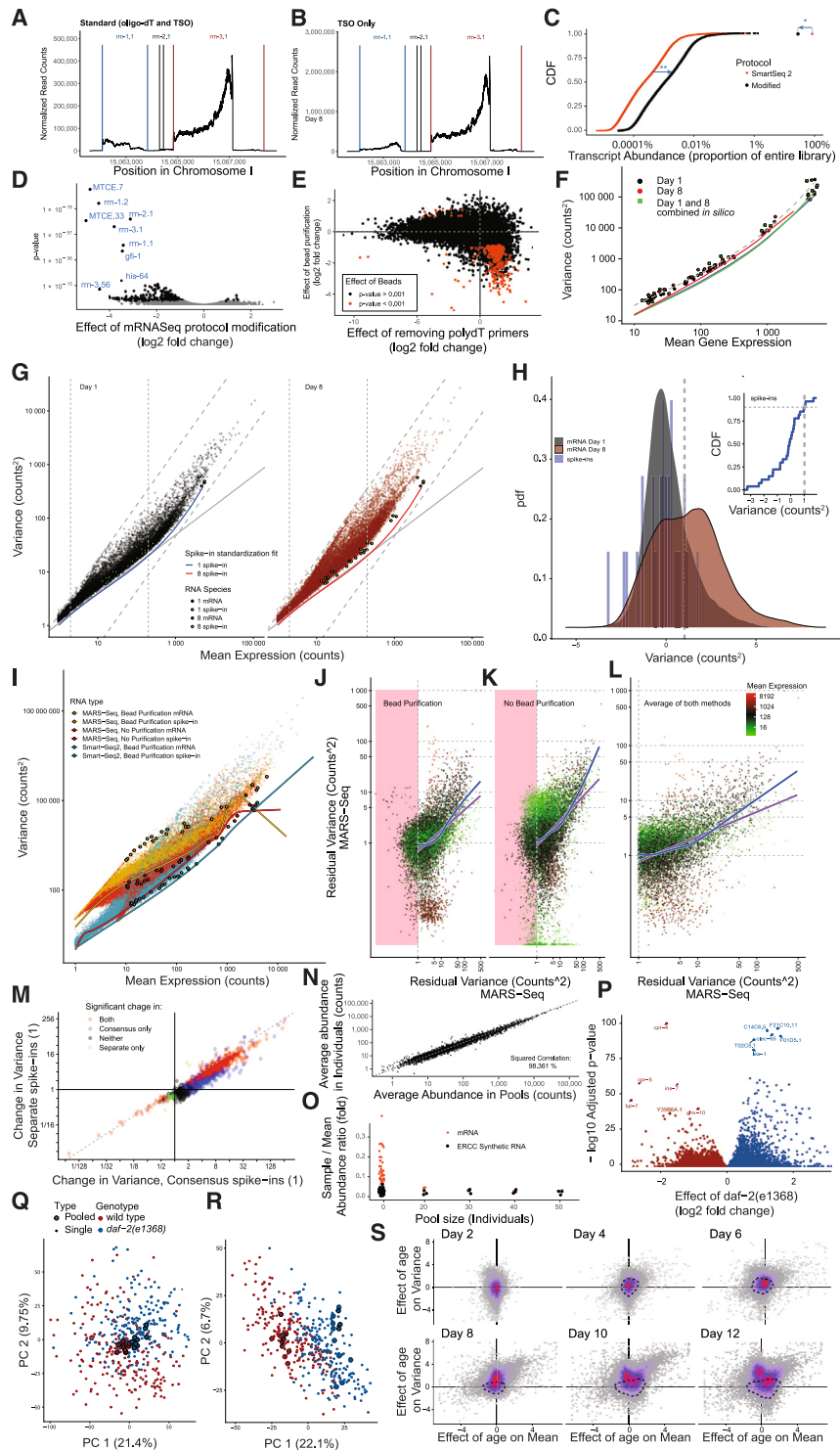
Figure S1. A population-scale atlas of non-genetic heterogeneity, related to Figure 1

(A and B) (A) Expression at the *rrn-1/2/3* locus in libraries, made using the standard Smart-Seq2 protocol or (B) using Smart-Seq2 with poly-dT primers omitted.

(legend continued on next page)

(C) The transcript abundance distribution in Smart-Seq2 libraries, as a fraction of the total library size, for standard Smart-Seq2 (orange) and the new biotinylated poly-dT (BpdT) protocol. A decreased *rrn-3* abundance (blue star) increases the relative abundance of all other transcripts (double blue star).

(D) Differentially amplified transcripts in the BpdT protocol relative to standard Smart-Seq2.

(E) The effect of the BpdT protocol relative to standard Smart-Seq2, compared with the effect of removing poly-dT primers. The BpdT protocol significantly depletes transcripts that are also significantly upregulated in poly-dT-removed libraries (orange).

(F) The mean-variance relationship of ERCC spike-ins added at a fixed concentration across samples, and thus reporting the technical variability of the experimental method for two cohorts, collected on day 1 (black), day 8 (red), or days 1 and 8 pooled together (green). The running minimum variance was fit as a function of the mean for day 1 (black), day 8 (red), and both days pooled together (green).

(G) The mean-variance relationship of all transcripts measured, including mRNA from days 1 and 8 (black, red) and spike-ins from days 1 and 8 (green, orange), and the running minimum of all transcripts on days 1 and 8 (blue, red).

(H) The distribution of variation across the transcriptome on day 1 (gray) and day 8 (red), with the pooled-variance of all spike-ins overlaid (blue). Inset: the cumulative distribution function of variation measured across all spike-ins shows a variation of 1 as a reasonable minimum variation measurable using our approach.

(I) The mean-variance relationship of samples collected using Smart-Seq2 and MARS-seq, showing that MARS-seq provides systematically higher variation estimates for both spike-ins and mRNA, both with and without bead purification of MARS-seq samples prior to processing.

(J and K) (J) The residual variation of each mRNA, as measured using Smart-Seq2 and MARS-seq, with linear (purple) and spline (blue) fits on samples above the technical minimum (unshaded area). MARS-seq samples were pre-treated, either using bead purification or (K) without bead purification.

(L) The same analysis, but plotting the average residual-overdispersion estimates for each gene across the bead-purified and unpurified MARS-seq approaches.

(M) Estimates of the effect of aging on each mRNA, as measured using MARS-seq with two different approaches for removing the technical mean-variance confound. Each transcript was normalized either to the spike-ins specific to that sample's cohort (y axis) or alternatively normalized to the pooled spike-ins across both day 1 and day 8 cohorts (x axis). Samples colored by statistical significance, with changes significant using both analytic methods (red), significant only in the pooled spike-in method (blue), only significant in the cohort-specific spike-in method (green), or significant in neither method (black).

(N) The transcriptomes of 205 individually sequenced samples are compared with 53 pooled samples containing 10+ individuals each. The average expression of each gene across each set of samples is compared.

(O) The inter-sample variability between the same samples, plotted as a function of pool size. Variability for each sample was measured as the average correlation distance of each transcript relative to the population-average abundance (statistical methods), estimated separately for mRNA (orange) and synthetic ERCC RNA spike-in standards (black).

(P) A volcano plot showing genes differentially regulated on average between *daf-2(e1368)* and wild-type single-individual samples.

(Q) Wild-type (red) and *daf-2(e1368)* (blue) individuals plotted in respect of their magnitude along the first and second principal components (PCs) derived from the correlation matrix of single-individual samples. Small points indicate single worms, large points indicate pooled samples.

(R) The same individuals, but using PCs derived from the correlation matrix of pooled samples.

(S) A density plot showing the same changes in mean-variance relative to day 1 as in Figure 1D, but here across a time series of 1,318 individuals sampled every second day from days 2 until 12 of adulthood. Dotted contour indicates the region outside of which 90% of genes show significant changes in variance.
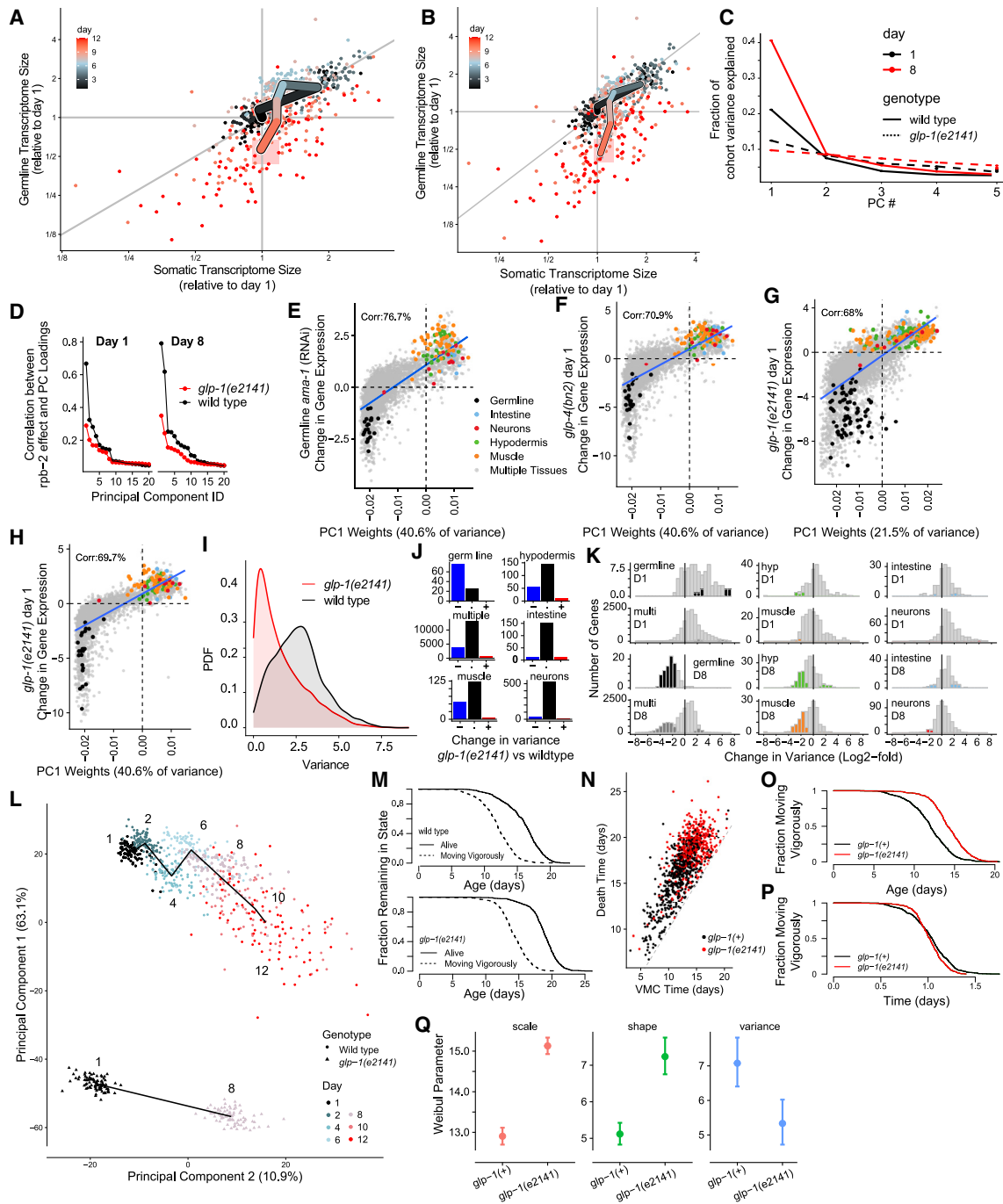
Figure S2. Aging involves a decoupling of germline and somatic mRNA content, related to Figure 2

(A) The absolute size of the germline-specific and soma-specific transcriptomes of each individual (points) relative to day 1, with lines connecting the daily means. Pink square is bootstrap 95% confidence interval of day 12 mean.

(B) The same age-dependent trajectory of germline and somatic scale factors, but this time excluding all transcripts expressed in embryos.[85]

(C) A diagnostic of the principal-component analysis (PCA) of the single-individual gene-expression matricies for four populations, wild type (solid line) and *glp-1(e2141)* (dotted line) individuals on day 1 (black) or day 8 (red) of adulthood, comparing the fraction of gene-expression variance explained for each population by each principal component (PC).

(D) The Pearson correlation between the effects of germline-specific *rpb-2(RNAi)* and each PC of either the single-individual gene-expression matrix for the wildtype (black) or for the *glp-1(e2141)* (red) population.

(E and F) (E) The correlation between the first PC weights calculated from wild-type individuals on day 1, the average effect of germline-specific *ama-1* RNAi measured on day 8, and (F) the *glp-4(bn2)* mutation measured on day 1.
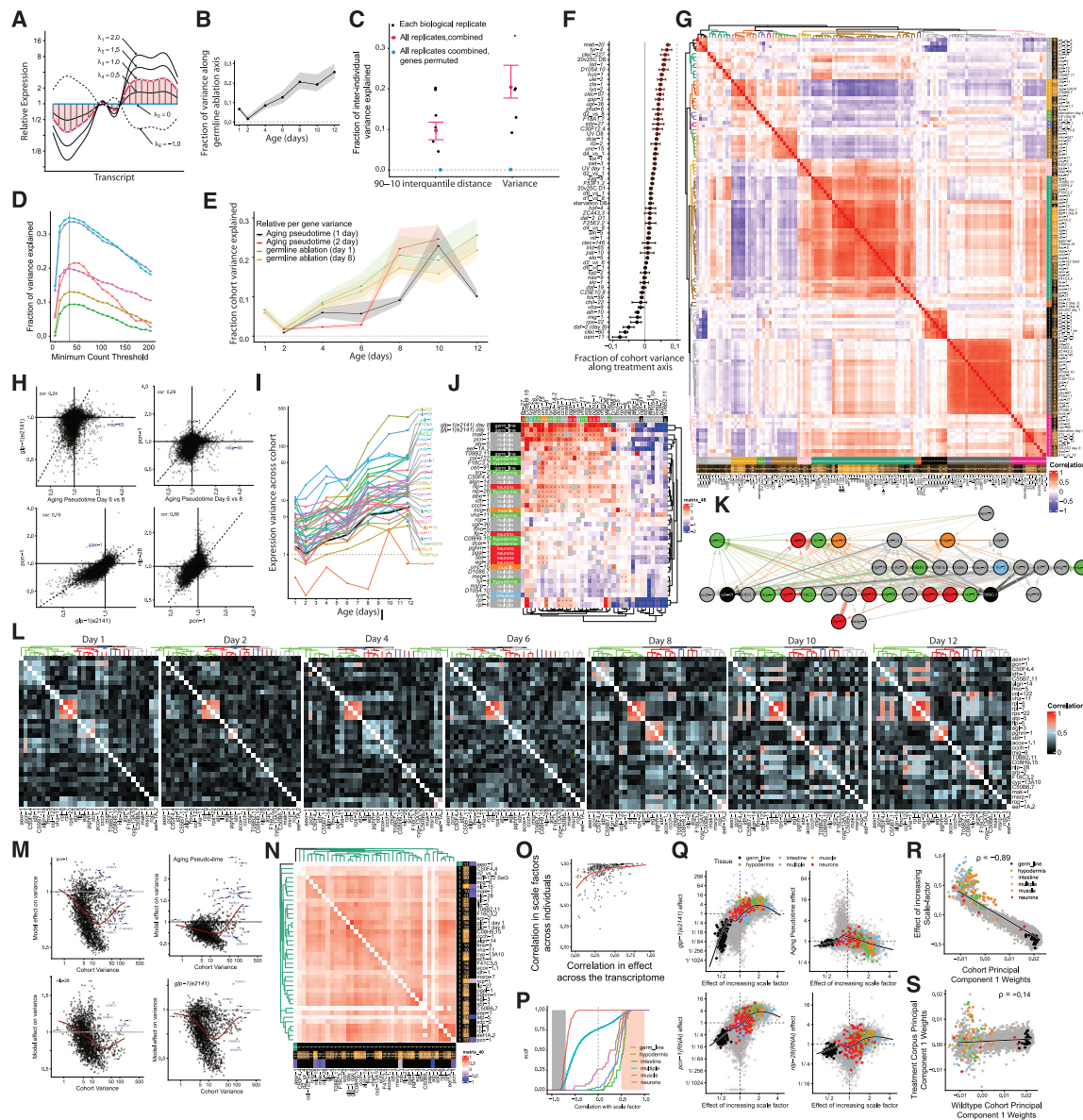
(legend continued on next page)

Figure S3. A perturbation screen to identify molecular drivers of non-genetic heterogeneity, related to Figure 3

(A) Transcriptome scaling models (TSMs) align the profile of each individual's gene expression across the transcriptome (black lines) to the profile of the gene-expression changes produced by some treatment (pink line) relative to the control (blue line). Scale factors (lambda values) are assigned to each individual.

(B) Many of the same results obtained indirectly using conventional PCA can be obtained more directly using TSMs. In Figure 2, we show that the first principal component (PC) of the day 8 wild-type cohort explains 40.5% of all gene-expression variance (Figure 2D), and this PC shows a 70.9% correlation with the *glp-1(e2141)* effect. We can repeat the same analysis by fitting a TSM to the wild-type day 8 population using *glp-1(e2141)* as the reference axis. In a young cohort on day 1, this TSM finds that 10% of all wild-type inter-individual gene-expression variance can be explained by variation along the *glp-1(e2141)* axis. During aging, this percentage increases to 26% by day 12, as shown by TSM estimates of the fraction of total gene-expression difference within each age-cohort, explained by individual variation along the *glp-1(e2141)* axis, with bootstrap 95% confidence region.

(C) Variation within wild-type day 8 population, including six independent replicate cohorts (black dots) was analyzed via TSM along the day 1 *glp-1(e2141)* axis. Two measures of gene-expression variation were calculated, variance and the 90–10 inter-quantile distance, with the fraction of variation explained for each cohort (black) and the entire pooled population (red) with bootstrap 95% CI. The same TSM was fit to the same pooled population, but with the names of each gene randomly shuffled in the *glp-1(e2141)* reference axis (blue).

(D) The effect of applying different minimum gene expression thresholds prior to TSM analysis of the same wild-type day 8 population along the same *glp-1(e2141)* axis. Each color shows the results of a TSM fit to a different biological replicate experiment within the larger wild-type dataset.

(E) We can also apply TSMs to the separation we had observed between the axes of inter-individual variation and population-average dynamics during wild-type aging (Figure 1E). TSM analysis can be directly applied to align individual variation along the pseudo-time axis by setting the treatment to be the average expression of an aged cohort and the control to be the average expression of a young cohort. In that case, scale factors quantify the relative biological age of each

(legend continued on next page)

individual between young and aged averages. TSMs fit this way show that individual variation along pseudo-time axis explains much less than 10% of all gene-expression variance early in life and only rises above 20% by day 8, shown as the fractions of variance explained by pseudo-time TSMs along an axis defined by consecutive time points (black) or time points separated by 4 days (red). This confirms in a different way our previous result obtained qualitatively by PCA: most non-genetic differences in aging do not arise from individuals navigating a single stereotyped aging trajectory at different rates.

(F) In a screen of 104 treatments, 52 treatments with the lowest fraction of variance explained within the wild-type day 8 cohort.

(G) All knockdown targets hierarchically clustered according to the pairwise Spearman correlation between each knockdown's fold-change effects on the transcriptome as measured on day 8 of adulthood.

(H) A comparison of the fold-effect change in gene expression of representative RNAi-target knockdowns selected from (G): *glp-1(e2141)*, *pcn-1*(RNAi), *nlp-28* (RNAi), and the average gene expression on day 6 relative to day 8.

(I) For the 40 top knockdown target hits from the cross-sectional transcriptome scaling model, which act as non-genetic variation mimetics (NVMs), we plot the cohort variance of the mRNA of each NVM at 7 time points across adulthood.

(J) To identify any mutual transcriptional regulation of NVMs, the fold-change in expression of each target (top) was measured after application of each treatment (left), with significant changes marked with black circles.

(K) The directed graph reflecting gene-regulatory relationships identified in co-expression data for all NVMs. Node colors correspond to the tissue of expression, as before, with line thickness indicating correlation magnitude.

(L) The co-expression between all pairs of hit target genes across cohorts measured on 7 time points across adulthood shows that gene regulation among these targets is not tightly coupled as one group.

(M) TSMs were fit to the wild-type day 8 population using four different treatments as reference—*glp-1(e2141)*, the pseudo-time difference between days 6 and 8, *pcn-1* (RNAi), and *nlp-28* (RNAi). For each gene (points) the fraction of variance explained by individual variation along the reference TSM axis (y axis) is compared with that gene's variation in the wild-type day 8 population. TSM models tend to explain better the variance of more highly variable genes (red line), but this is not true for all highly variable genes (blue labels).

(N) Using each RNAi knockdown as a reference, TSM was fit to the wild-type population on day 8. The correlation of the change in variance across each transcript was estimated for each pair of scaling model residuals, and hierarchically clustered.

(O) We compare the correlation between each pair of NVM knockdowns (points) in respect of their effect on gene expression across the transcriptome (x axis) and the correlation between the same NVM knockdown pair in respect of the scale factors assigned by TSMs (y axis) to individuals in the wild-type day 8 cohort. In doing so, we compare the values plotted in Figures 3C and S3G to the values plotted in Figure 3D, and show that the similarity of all NVM targets in the latter is not reflected in a similarity in the former (locally estimated scatterplot smoothing [LOESS] fit, red).

(P) The correlation in expression of each gene with the scale factor was calculated. These correlations are plotted as a cumulative distribution function, grouped by tissue of expression. Genes showing the largest negative (gray shaded) and largest positive (red shaded) correlations are highlighted.

(Q) To understand the common set of gene-expression targets influenced by all NVMs, we then considered all genes whose expression across wild-type individuals showed a correlation greater than or equal to 70%, with scale factors corresponding to all NVMs (statistical methods). We find that 617 genes, 3% of the transcriptome, correlate to this extent (Data S4). In particular, we find that the abundance of an individual's germline-specific genes is strongly negatively correlated with their scale factor, and the abundance of soma-specific genes is strongly positively correlated. Each individual in the wild-type day 8 population was ordered according to the mean of the scale factors assigned by all treatments' transcriptome scaling models. In this panel, we highlight the differential analysis that compared the top 25% and bottom 25% of individuals to identify gene-expression profiles corresponding to "high" and "low" scale factor individuals in respect of the entire corpus of NVM treatments, defining a set of fold-changes across the transcriptome that we call "the effect of increasing the scale factor" (Data S6). The effect of four treatments, *glp-1(e2141)*, the effect of aging between 6 and 8 days, *pcn-1(RNAi)*, and *nlp-28 (RNAi)* across the transcriptome are compared with the effect of increasing the scale factor. We conclude that individuals with larger-scale factors have increased ratios of somatic to germline mRNA content and, vice versa, that individuals with smaller-scale factors have decreased ratios of somatic to germline mRNA content.

(R) This scaling effect was compared with the weights of PC1 of the wild-type day 8 population.

(S) The first principal component (PC) of the corpus of gene-expression changes produced by treatments that are not NVMs—that explain less than 10% of wild-type variance—compared with the first PC of the wild-type day 8 single-individual cohort. Tissue-specific genes identified by color, hypodermis (green), muscle (orange), neurons (red), germline (black), and multiple tissues (gray).
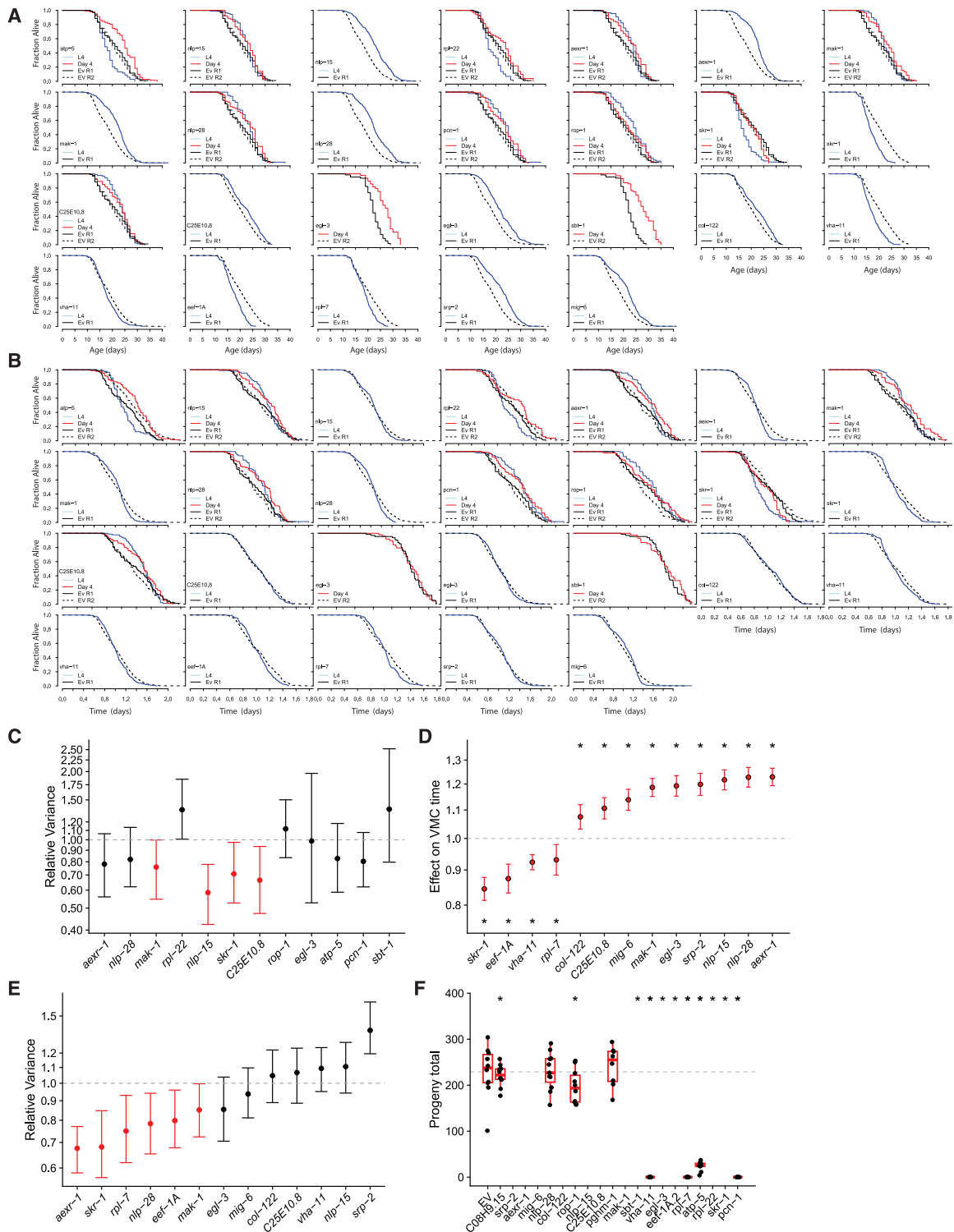
Figure S4. A new class of molecular drivers of non-genetic variation in lifespan, related to Figure 4

(A) The full survival curves for each RNAi knockdown lifespan experiment shown in Figure 4C, showing knockdowns started immediately before the start of adulthood (L4 larval stage) (blue), or on day 4 of adulthood (red). Empty vector (EV) controls (black) are shown for the L4 treatment (solid) and day 4 (dotted) conditions. Independent biological replicates are plotted separately.

(B) The AFT regression residuals comparing the lifespan of each RNAi to EV across all experiments.

(C) The relative variance of the AFT residuals of all RNAi knockdowns, as in Figure 4F, but in populations exposed to each RNAi starting later, on day 4 of adulthood. Bootstrap 95% confidence intervals (CIs) are shown.

(D) AFT regression was performed to compare the vigorous movement span (VMC) of individuals exposed to each RNAi and empty vector. AFT estimates quantify the fold-change in VMC of each RNAi relative to EV, shown with 95% bootstrap CIs.

(E) The change in variance of the residuals from the same AFT regression model fit to VMC times for each RNAi compared with EV, with 95% bootstrap CI.

(F) The effect on the total number of eggs laid for each RNAi knockdown shown in Figure 4G, but for RNAi knockdowns started at the time of hatching.
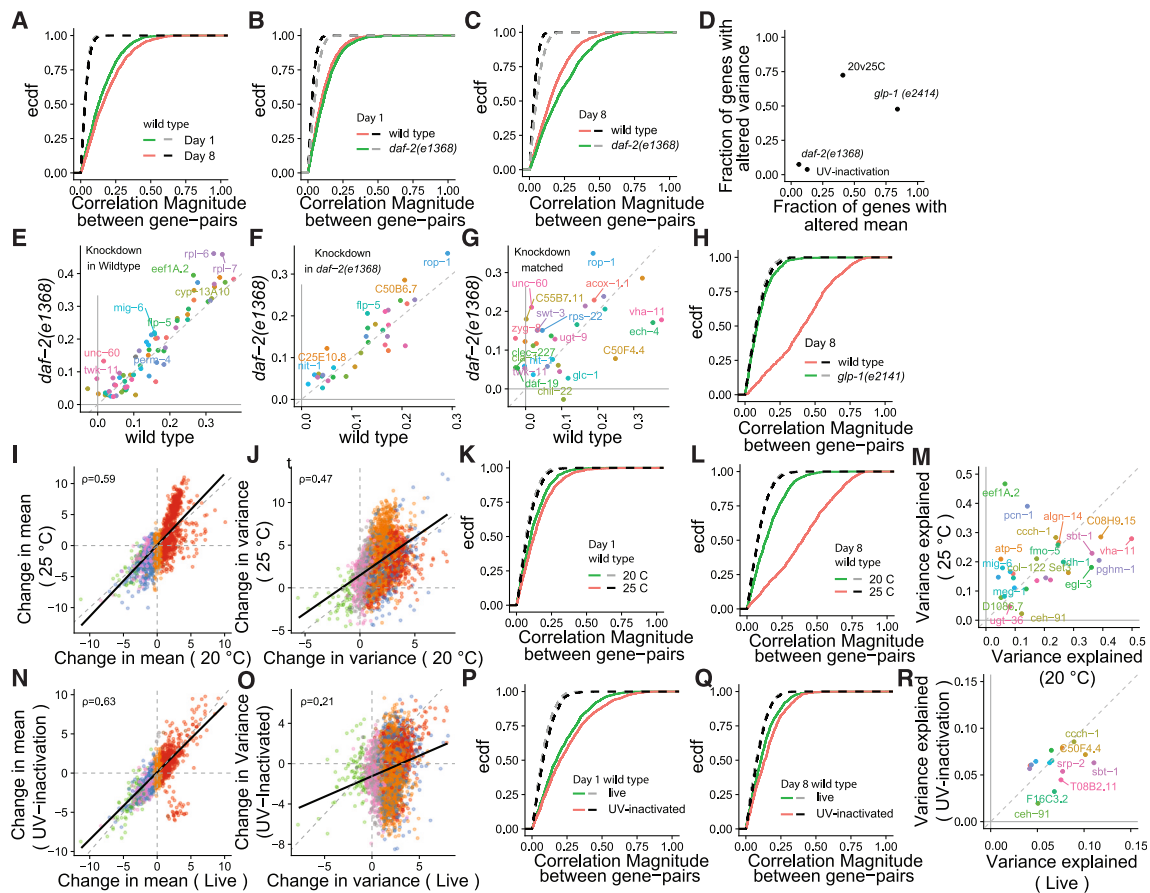
Figure S5. A partial decoupling of lifespan mean and variance, related to Figure 5

(A–C) (A) The distribution of pairwise correlation magnitudes (absolute value of co-expression) in gene expression across the transcriptome is compared between wild-type populations on day 1 (green) and day 8 (red) (B) compared between wild-type (red) and *daf-2(e1368)* (green) populations on day 1 and (C) on day 8. Gray and black dotted lines represent the distribution of pairwise correlation magnitudes of the respective distribution (green and red) but with gene identities permuted to approximate a random matrix.

(D) The fraction of the transcriptome with significantly altered mean and variance ($p < 0.001$; studentized bootstrap) on day 8 in *daf-2(e1368)* populations relative to wild type, in *glp-1(e2141)* populations relative to wild type, in populations housed at 25°C relative to 20°C, and in populations fed UV-inactivated bacteria relative to populations fed live bacteria.

(E) Transcriptome scaling models (TSMs) were fit to both wild-type and *daf-2(e1368)* populations for each NVM gene. Reference RNAi knockdowns were performed in a wild-type genetic background. Plotted are the fraction of gene-expression variance explained by each reference.

(F) The same analysis as in (E), but this time using reference RNAi knockdowns performed in a *daf-2(e1368)* background.

(G) The same analysis as in (E), but this time each scaling model was fit using the reference RNAi knockdown matching the genotype of the subject population.

(H) The distribution of pairwise correlation magnitudes (absolute value of co-expression) in gene expression across the transcriptome is compared between *glp-1(e2141)* (green) and wild-type (red) populations on day 8.

(I and J) (I) The fold-change in gene-expression mean and (J) variance across the transcriptome between days 1 and 8, comparing the effect of a 5°C increase in body temperature, from 20°C to 25°C. Genes (points) are colored according to statistically significant changes with age ($p < 0.001$, Wald test), using the same colors as in Figure 5A.

(K and L) (K) The distribution of pairwise correlation magnitudes (absolute value of co-expression) in gene expression across the transcriptome is compared between populations housed life-long at 20°C (green) and 25°C (red) on day 1 and (L) on day 8.

(M) The same as in (E), but comparing TSMs fit to populations at 20°C and 25°C.

(N and O) The same as (I) and (J), but showing the effect of a lifespan-extending change in diet, comparing animals fed live bacteria to those fed UV-inactivated bacteria.

(P and Q) (P) The same as in (K) but comparing populations fed life-long live (green) or UV-inactivated (red) bacteria on day 1 or (Q) on day 8.

(R) The same as in (M), but comparing TSMs fit to populations on day 8 fed life-long live (green) or UV-inactivated (red) bacteria.
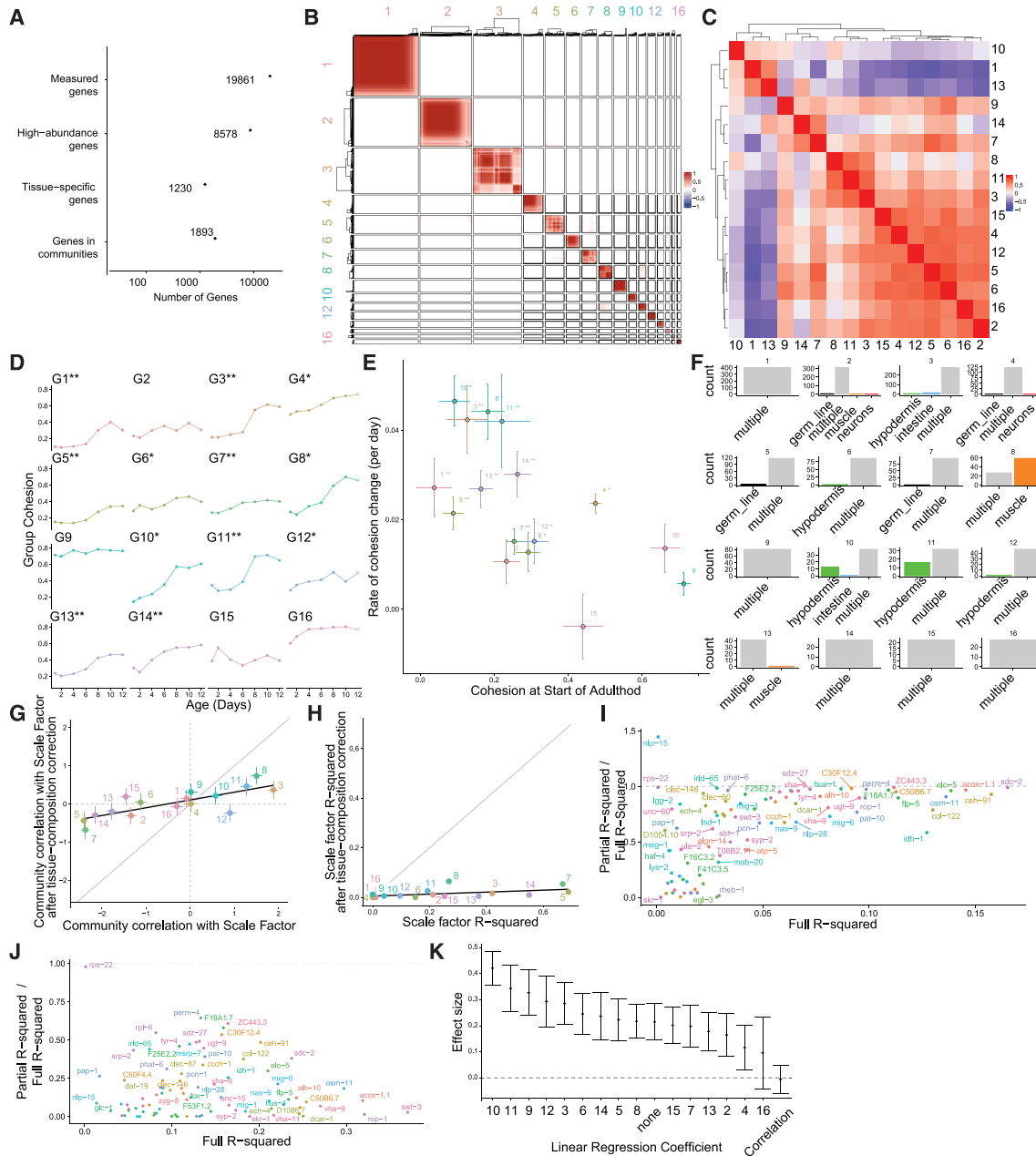
Figure S6. Inter-individual heterogeneity within cohorts can be used to predict the effect of interventions on aging, related to Figure 6

(A) The number of genes identified in the RNA sequencing data (first row), the number of genes exceeding the minimum abundance threshold (second row), the number of tissue-specific genes used for tissue-specific regression (third row), and the number of genes found to exist within co-expression groups (fourth row).

(B) Across 1,000 bootstrap replicates across individuals from the day 8 wild-type cohort, the frequency of co-occurrence of each pair of genes in the same co-expression group—on a scale from 0 (white), where two genes never are assigned to the same co-expression group, to 1 (red), where two genes always fall within the same co-expression group. Well-defined separation between each co-expression groups' members indicates a robust group assignment.

(C) Across the day 8 wild-type population, the mean abundance of co-expression group members was calculated for each group in each individual. Then, the correlation between group abundances across individuals was estimated, and these correlations were hierarchically clustered and plotted as a heatmap.

(D) The cohesion of each group—the average covariation among all pairs of genes in the group—was estimated for each time point of the wild-type time series.

(E) The initial group cohesion and its rate of change with age, based on linear fits of the cohesion time series for each group. Bars mark standard errors.

(F) Histograms show the tissue localization of the members of each co-expression group.

(G) A linear regression model was fit to compare the average expression of all members of a co-expression group in each wild-type individual to the NVM scale factor assigned to that individual, across the day 8 cohort. The effect sizes measure the fold-change in expression on average across group members associated with a fold-change increase in scale factor—large magnitudes indicate a strong correlation between an individual's scale factor and that individual's average

(legend continued on next page)

expression of the co-expression group. The regression was run using gene-expression data of co-expression group members before (x axis) and after (y axis) tissue-compositional normalization.

(H) Results from the same linear regression models, but this time plotting the coefficient of determination (r^2) of the model rather than the effect size for each co-expression group.

(I) To understand why the residual covariation matrix can be used to predict the effects of interventions on aging, we performed a series of additional statistical diagnostics. First, we considered the possibility that our predictions depended on gene-expression changes in a particular tissue. We find that knowing the tissue localization of a transcript does not help us predict its response to interventions on aging, as shown by the coefficient of determination of the full linear regression model (log(RNAi_effect) $\sim$ co-expression + tissue) compared with the partial r-squared explained by the reduced model (log(RNAi_effect) $\sim$ co-expression) for each RNAi-target knockdown (colored points). The horizontal dotted line indicates the level where partial r-squared/full r-squared == 1, i.e., the case where tissue localization of transcripts are not informative about the effect of RNAi knockdown.

(J) So, we then performed a regression in which each transcript measured in the RNAi knockdown experiment was labeled according to its membership in a co-expression group identified in the wild-type aged population. Labeling transcripts in this way greatly improved our ability to predict their response to interventions in aging, shown using the same plot as in (I), but here comparing the partial model (log(RNAi_effect) $\sim$ co-expression) to the full model (log(RNAi_effect) $\sim$ co-expression + co-expression-group). The horizontal dotted line indicates the level where partial r-squared/full r-squared == 1, i.e., the case where collective changes among all group members is not informative about the effect of RNAi knockdown.

(K) When we include group labels in our prediction model, the residual co-expression matrix then contributes no additional benefit over the labels alone, as shown by the effects of RNAi knockdown on each community's members in the linear model (log(RNAi_effect) $\sim$ co-expression + co-expression-group), averaged over the top hits from regression analysis where co-expression alone explained 25% of variation in log(RNAi effect). "none" indicates the effect of a gene not being in any co-expression group. The "correlation" parameter models the effect of the pairwise covariation between two genes—a small value indicates that this measurement of covariation does not have additional predictive power beyond the co-expression membership group of the two genes considered. From this, we conclude that the residual covariation matrix can be used to predict intervention on aging because the co-expression groups we identify in aged cohorts act as gene-expression modules that respond cohesively to RNAi. In other words, the physiologic axes of non-genetic, inter-individual variation observable in a single aged cohort are the same physiologic axes along which individuals respond to interventions in aging. Therefore, we can use the former to predict the latter.