Differences in protein dosage underlie nongenetic differences in traits

Nikolay Burnaevskiy[1], Bryan Sands[1], Soo Yun[1], Patricia M Tedesco[2], Thomas E Johnson[2], Matt Kaeberlein[1], Roger Brent[3,4], Alexander Mendenhall[1,4*]

[1] Department of Pathology, University of Washington, Seattle, Washington

[2] Department of Integrative Physiology, University of Colorado, Boulder, Colorado

[3] Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington

[4] Co-Senior Authors

[*] To whom correspondence should be addressed: alexworm@uw.edu

ABSTRACT

Genes and environments do not entirely explain variation in biological traits. Discrete and complex traits manifest distinctly in isogenic organisms in homogenous environments. One cause of variation in traits is variation in the expression of genes. We do not understand how variation in gene expression arises among isogenic individuals, nor do we fully understand the consequences. Here, for the first time in a metazoan, we quantified sources of cell-to-cell variation in gene expression to understand how it arises in the cells of *Caenorhabditis elegans*. We adapted an experimental approach and analytical framework we used previously in *Saccharomyces cerevisiae.* We measured cell-to-cell variation in gene expression attributable to: 1) stochastic molecular events in transcription and translation (a.k.a. intrinsic noise) and/or allele access, 2) pathway-specific signaling noise, and 3) general differences in protein dosage resulting from global differences in protein production or turnover. Surprisingly, and in contrast with single celled organisms, in *C. elegans* cells, general differences in protein dosage constitute the major source of variation. Cells with high protein dosage came from animals with high protein dosage. These differences in protein dosage were consequential. Animals that made more reporter proteins had discrete traits consistent with higher dosage of unrelated neomycin resistance and Ras gain-of-function proteins, consistent with global variation in protein dosage. Our results suggest that the natural axis of protein dosage underlies nongenetic, nonenvironmental differences in complex and discrete traits, including penetrance and expressivity of hypomorphic and hypermorphic alleles. They also raise the possibility that differences in penetrance caused by differences in protein dosage might be responsible for some of the missing heritability in genome wide association studies.

INTRODUCTION

Genes and environments are not the sole determinants of discreet or complex traits[1]. That is, complex and discreet traits vary among isogenic organisms existing in the same homogeneous environment. In the early 20th century, Timoféeff-Ressovsky[2] and Romaschoff[3], observed that, in isogenic populations of *Drosophila funebris*, some mutations affected only some individuals (*incomplete penetrance* of a mutation's phenotype, a discreet trait), and the severity of the phenotype varied among affected individuals (*expressivity*)[4]. Later, scientists characterized numerous differences in traits among isogenic unicellular organisms in the same environment, including differences in the lifespan of single yeast cells[5] (a complex trait). Similarly, researchers showed that individuals in isogenic populations of *C. elegans* in homogeneous environments manifest different lifespans[6-8] (a complex trait) and had incomplete penetrance of mutant alleles[9] (discreet traits).

Nongenetic differences in the manifestation of traits are not stochastic; they can be predicted by measured variables. Consider the, perhaps, seemingly peculiar, series of reports wherein differences in the measured expression level of a single gene correlates with and/or predicts differences in discreet and complex traits, in systems ranging from yeast to human stem cells. For example, there can be a nine-fold difference in the expression level of a pheromone reporter gene among individual yeast cells; cells that express more reporter gene increase in volume more rapidly[10]. In *C. elegans*, we have shown that high expression levels of a single gene, a GFP-based reporter of the molecular chaperone encoded by *hsp-16.2*, predicted increased lifespan and thermotolerance[7,8] (complex traits). The same reporter gene can also predict differences in the penetrance of hypomorphic mutations that cause defects in larval development[9] (discreet traits). Similarly, mammalian stem cells that express high levels of a cell surface protein, SCA-1, are more likely to differentiate into myeloid cells[11]. In these experiments where we and others measure a reporter gene and then see some other trait covary, there must be some

underlying causative gene or genes that covary with the measured gene. In cases like the measurement of endogenous SCA-1, the measured gene may be causative; however, this is not the case for the freely diffusing GFP reporter protein in *C. elegans*.

The molecular basis for variation in gene expression phenotypes (such as *hsp-16.2* reporter expression[12-14]) and their relation to variation in more complex traits they predict (such as lifespan[7,8]) is not well understood. In *C. elegans*, *hsp-16.2* reporters are expressed most abundantly in the 20 cells of the intestine, which form distinct segments (rings; Figure 1a)[12,14]. Genetic experiments suggest that the physiology of the worm intestine affects lifespan. For example, intestinal expression of the forkhead transcription factor homolog of FOXO3, *daf-16*[15], is sufficient to extend lifespan. Similarly, the expression level of a number of reporter genes of seemingly disparate function correlated with lifespan when measured from intestine cells[16]. Yet we do not entirely know how differences in reporter gene expression arise, nor why they are consequential.

Ultimately, differences in traits must be caused by differences in the expression of some causative gene or genes. Thus, it is important to understand how nongenetic, nonenvironmental differences in gene expression arise. One way to understand how variation in gene expression arises is to directly observe and quantify differences in gene expression using fluorescent reporter genes. The ability to directly observe the fluorescent signals in living cells allowed us[10] and others[17-19] to isolate and study different sources of variation in gene expression in single celled microbes.

To better understand variation in gene expression and its relationship to discreet and complex traits such as mutation penetrance and lifespan, we developed means to precisely quantify differences in reporter protein expression in intestine cells[14]. We measured animal-to-animal and cell-to-cell variation in the expression of a set of *hsp-16.2* based reporter genes. Our experiments showed, perhaps surprisingly, that differences in site of reporter integration, copy

number, identity of the fluorophore, or sequence of 3'UTR did not affect variation in gene expression[14]. Consistent with this observation, we also found that different *hsp-16.2* reporters have identical abilities to predict lifespan and heat tolerance[8]. Moreover, high or low expression of these reporters defined a whole-organism physiological state; that is, bright intestine cells came from animals composed of bright cells, and dim intestine cells came from animals composed of relatively dim cells[14].

Here, to better understand the relationship between the gene expression phenotype and other traits, we built on previous work in *S. cerevisiae* to dissect and quantify different sources of gene expression variation[10]. This is the first time these sources of variation have been quantified in a metazoan. The results were surprising.

RESULTS

Here we adapted an approach we used in yeast[10], wherein we compared the outputs of two differently colored (different fluorescent proteins) versions of the same reporter gene expressed from two identical loci on homologous chromosomes (Type I experiments) or the outputs of two different reporter genes (Type II experiments), shown in Figure 1. Measuring the pairs of reporter genes in these two different kinds of experiments allows us to quantify the amount of cell-to-cell variation in gene expression attributable to three distinct, experimentally tractable bins, each with distinct underlying molecular causes. Figure 1 shows how we can use this reporter gene measurement scheme to quantify differences in intrinsic noise[17,18] ($\eta^2(\gamma)$), signaling through different pathways[10] ($\eta^2(P)$), or general protein expression capacity[10] ($\eta^2(G)$). A complete description of the analytical framework is provided in Supplemental Materials Section 2: Analytical Framework.

Type I experiments allow us to measure how much cell-to-cell variation there is in apparently stochastic to-promoter or to-transcript binding events, or allele access. Type I

experiments reveal how much cell-to-cell variation there is in allele expression/access as uncorrelated variation, often referred to as intrinsic noise ($\eta^2(\gamma)$); Figure 1 shows an example of a scatterplot for the same cell type (same lineage/fate) measured from many different animals from one of these experiments. If there were a lot of uncorrelated variation (a cloud like appearance of points) it would mean that there were significant cell-to-cell differences in to-promoter binding, to-transcript binding, or allele access. If there was little uncorrelated variation (alignment of points along the major axis; Figure 1), it would mean that most cells would have had access to both alleles. It would also mean that there were few differences attributable to probabilistic differences in biochemical binding events (to-promoter/to-transcript) that tend to dominate when there are small numbers of molecules involved[18]; we would say that intrinsic noise was low.

In Type I experiments for *hsp-16.2*, and for all the other reporter genes we measured, intrinsic noise was low. The expression level of one allele accounted for 90% or more of the expression of the other allele in animals. Figure 2 shows the average intrinsic noise levels for a few genes we measured. Intrinsic noise was a minor component of the cell-to-cell variation. We show individual Type I scatterplots and bar graphs organized by cell type and experiment in Supplemental Figures 1&2.

Type II experiments compare the expression levels from two distinct genes to tell us about cell-to-cell differences in signaling through different pathways ($\eta^2(P)$). In the scatterplot from a Type II experiment shown in Figure 1, the uncorrelated variation (dispersion across the major axis) is a measure of cell-to-cell differences in signaling through distinct pathways. We subtract the average intrinsic noise ($\eta^2(\gamma)$) for the gene pairs we are measuring, so we know that the remaining dispersion is due to signaling/pathway noise ($\eta^2(P)$) (shown in Figure 1). Type II experiments also quantify how much cell-to-cell variation is attributable to differences in protein expression capacity ($\eta^2(G)$), which we define as the ability to express and maintain proteins. In

the scatter plot of a Type II experiment on the bottom right of Figure 1, the correlated variation (dispersion along the major axis) is a measure of ($\eta^2(G)$).

In Type II experiments, we found that cell fate (which ring a cell was in) was the primary determinant of a rigid cell-fate specific ratiometric setpoint for any given gene pair, shown in scatterplots of cells grouped together and by ring in Supplementary Figure 3. Thus, we grouped cells we measured by the ring from which they came to prevent artificial inflation of ($\eta^2(P)$). We found that there was little cell-to-cell (and thus, animal-to-animal) variation in signaling through distinct pathways to activate distinct genes. Figure 3 shows that there was little signaling noise ($\eta^2(P)$). Figure 3 also shows that we found that most of the variation in gene expression was attributable to cell-to-cell (and thus, animal-to-animal) differences in protein expression capacity ($\eta^2(G)$). These intestine cells express more protein from other, distinctly regulated genes, even if they are on other chromosomes, or fused to other genes (Emerin::GFP; Figure 3e). Figure 3f shows ($\eta^2(G)$) visually; two different types of animals expressing two distinct sets of reporter genes from Type II experiments are arranged from dimmest to brightest. Note that the patterns of expression are maintained but that the overall expression level of both reporters is different between animals; to observe this phenomenon, there must be little intrinsic noise ($\eta^2(P)$) and little signaling noise ($\eta^2(\gamma)$). We show individual Type II scatterplots and bar graphs organized by cell type and experiment in Supplemental Figures 4, 5 & 6.

To corroborate the finding that protein expression capacity is a major mechanism of cell-to-cell (and thus, animal-to-animal) variation, we conducted two additional types of experiments where we sorted animals on the expression level of a fluorescent reporter gene (approximately top or bottom 5-10% of animals), and measured phenotypes indicative of the activity of particular non-fluorescent genes. First, we sorted larval (L1) animals on GFP expressed from a site on chromosome I, controlled by the promoter of another ubiquitously expressed gene, *eft-3*; these animals also expressed the neomycin resistance gene (Neo^R) on a different chromosome

(chromosome II) under control of the *rps-27* promoter. If general protein expression capacity (G) was a major mechanism of cell-to-cell (and animal-to-animal) variation, then animals that have more GFP should have more NeoR and resist extreme neomycin concentrations to a greater extent. Figure 4A shows that we found this to be the case; brighter green animals can grow to adulthood in a high neomycin environment better than dimmer animals.

Next, we used the *hsp-16.2* lifespan biomarker gene to sort animals expressing an incompletely penetrant, dominant Ras gain of function mutation that caused animals to develop between 0-4 hypodermal neoplasias[20]. If animals that made less biomarker had less protein expression capacity, we would expect them to have lower penetrance and expressivity for this gain of function mutation. In fact, Figure 4B shows that animals that make less of the biomarker have a lower penetrance; brighter animals have more hypodermal neoplasias and a higher probability of developing them. Thus, animals seem to vary in effective gene dosage for multiple different genes. Supplemental Figure 7 shows images of animals in which the nuclear envelope of the somatic cells, including the neoplastic hypodermal cells, is delineated by a green fluorescent protein fusion to Emerin, a nuclear envelope component.

DISCUSSION

Here, for the first time, we quantified different sources of cell-to-cell variation in gene expression in a metazoan. Work was facilitated by development of precise microscopic methods that enabled accurate measurements of gene expression in individual *C. elegans* intestine cells[14]. Our experiments allowed us to independently quantify three possible sources of variation in gene expression. We found that the major contribution to cell-to-cell variation among isogenic animals in the same environment was correlated variation in the expression of reporter genes (protein expression capacity; G; Figures 1&3). Our previous work identified whole animal differences in reporter gene expression and suggested that these differences defined a whole organism physiological state[14]. Our work here shows that these differences are caused by differences in

general protein expression capacity (quantified by a variable, G), which define differences in protein dosage in cells throughout the animal. Our experiments here showed that animals with higher G showed greater penetrance of two single gene traits: antibiotic resistance conferred by a Neo^R transgene, and developmental hypodermal neoplasia caused by a gain of function in the sole *C. elegans* Ras gene (*let-60*[20]).

The fact that, in *C. elegans*, differences in G constitute the major contribution to organism-to-organism variation in gene expression, was somewhat surprising. In *S. cerevisiae*, differences in G are also important for variation, but differences in *P*, the ability to send signal to activate different promoters, are of similar magnitude[10]. Our findings do not exclude the possibility that variation in other cell types at other points in the life of an organism might be dominated by other sources of variation in gene expression. In this vein, it will be interesting when future researchers quantify the different contributions to the increased cell-to-cell variation in gene expression observed in aged mammalian cardiomyocytes[21].

Work in simple metazoans like *C. elegans*, whose cells are observable in intact animals via light microscopy, is an essential complement to work in mammalian cell culture. The work in simple metazoans allows us to observe and understand types of cell-to-cell variation, such as variation in G for the whole intestine tissue (and effectively whole animals), which might not be observed in cultures of cells lacking the proper milieu and control signals. For example, in *C. elegans*, intestinal proteostasis is known to be controlled by insulin signaling and neurons[22,23], and variation in general protein dosage, G, might stem from differences in neuronal and/or insulin signals[24].

Our findings on animal-to-animal differences in protein dosage cause us to revisit previous results on a complex trait, lifespan. Earlier, we found that young adult *C. elegans* with greater expression of a *hsp-16.2* reporter ($P_{hsp16.2}$*::gfp*) lived longer and were more tolerant to stress[7,8]. We asserted that high reporter gene expression in these animals defined a physiological state

that protected the animals from the "rigors of living". The *hsp-16.2* gene encodes a small molecular chaperone, and it seemed possible that overexpression of *hsp-16.2* and other chaperones[13] might account for the longevity phenotype[25,26], perhaps by protecting cells from the consequences of damage due to accumulated misfolded proteins. However, Sanchez-Blanco and Kim identified a number of reporter genes that predict lifespans, and not all of them were chaperones (for example, superoxide dismutase, *sod-3*). Significantly, in these experiments lifespan prediction power increased when the investigators quantified correlated expression of any two reporters (thus measuring the variable we call G).

Our finding that that individual animals (and the cells that comprise them) show general differences in protein dosage may help explain incomplete penetrance. Burga et al. showed that animals with increased expression of a *daf-21* reporter gene ($P_{daf\text{-}21}$*::mCherry*) had decreased penetrance of a hypomorphic *tbx-9* mutation[27]. Since *daf-21* encodes the Hsp90 chaperone, it seemed possible that the diminished penetrance of the *tbx-9* mutant phenotype might be caused by increased chaperone activity that increased the activity of the hypomorphic TBX-9 protein. Consistent with this idea, Casaneuva et al. showed that animals with chaperone levels increased by three different experimental means exhibited decreased penetrance of a few different hypomorphic mutant alleles. These observations are also consistent with the idea that animals that make more chaperone reporters have higher doses of most of the proteome. Furthermore, their experiments suggest that only mutant proteins with some residual function could benefit from the physiological state defined by increased chaperone reporter expression level. For example, penetrance of the hypomorphic *lin-31(n1053)* allele (a point mutation changing a single amino acid) showed differential penetrance in animals with different measured expression levels of a *hsp-16.2* reporter gene, but the *lin-31(gk569)* null deletion (presumed null, first three of four total exons deleted) mutant did not[9]; this is consistent with the idea that high G animals that make more of a hypomorphic mutant protein get some increased functional capacity from increased

dosage of that partially functional protein, but animals that make more of a null/truncated protein do not.

Thus, taken with previous work, our results suggest a complementary interpretation for the relationship between reporter gene expression, longevity (complex trait), and incomplete penetrance (discreet trait). We suggest that: 1) the physiological state first defined operationally by high expression of a $P_{hsp16.2}$*::GFP* reporter is a global increase in protein dosage defined here (increased G), 2) that this state may cause increased longevity by its general increase in dosage of proteins (in addition to chaperones), and 3) that changes in penetrance of particular hypomorphic and hypermorphic mutations may be due to changes in the dosage of these mutant proteins.

We do not understand the molecular origins of organism-to-organism differences in protein dosage. In principle, animals might differ in their ability to transcribe DNA into RNA, translate mRNA, or in the rates at which mRNA and proteins are degraded. Regarding how differences in mRNA production may come to be, in mammals, expression of c-Myc proteins globally increases transcription [28,29]. Thus, in *C. elegans*, differences in G could be due to differential expression of genes such as *mml-1*, the (sole) Myc family gene[30], which might lead to global differences in transcription. For instance, measurements of single mRNAs suggest that incomplete penetrance of *skn-1* hypomorphic mutants (which fail to form intestines) in some animals might be due to high expression of mRNA for *end-1*[31]. It is possible that global differences in mRNA caused differences in *end-1*. Alternatively, increased chaperone abundance in some animals[13], might lead to global increases in translation, for example by increasing the protein folding efficiency of the nascent protein associated complex[32].

However, we are tempted to speculate that a major contribution to animal-to-animal differences in G might be differences in the rate of protein turnover. Activation the *daf-2*/*daf-16* insulin signaling system decreases general protein turnover[33-35]. We showed recently that

activation of this system increases mean expression of the *hsp-16.2* biomarker gene[24], consistent with the idea that decreased protein turnover conveyed by the *daf-2(e1370)* mutation would lead to increased expression level. In the same report, we also showed that the AFD sensory neuron pair and insulin signaling controlled the animal-to-animal variation in reporter gene expression[24]. Taken together with this report, our prior investigation suggests that variation in G might be controlled by sensory neurons and insulin signaling (see working model in Supplemental Figure 8). Perhaps significantly, Sanchez-Blanco and Kim reported that the ability of a lifespan biomarker ($P_{sod\text{-}3}$*::GFP*) to predict lifespan depended on an intact insulin signaling system.

In animals ranging from humans to flies to worms, the manifestation of traits conferred by specific alleles is often incomplete. Because selection acts on phenotype, incomplete penetrance has long been recognized as a factor affecting the rate of evolution[36]. Our results suggest that in other multicellular organisms, global differences in protein dosage might account for some organism-to-organism phenotypic difference in discreet and complex traits, such as lifespan (Figure 5). Variation in allelic penetrance bedevils detection of gene-trait associations in human genome wide association studies[37] (GWAS). Differences in the phenotypic penetrance of alleles contributes to the phenomenon called "missing heritability[38]; specifically, the summed contributions of loci identified in GWAS studies cannot account for all of the heritability in the disease or trait. Missing heritability might have multiple causes including, for example, epistatic gene-gene interactions whose effect (on fitness) can now be comprehensively surveyed and quantified (in *S. cerevisiae*[39]). We suggest that differences in protein dosages causing cells and clones of cells to manifest different phenotypes and exist in different physiological states[7,40,41] might also account for some of the missing heritability of traits.

Acknowledgements: Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We thank the other laboratories and scientists that have contributed to the understanding of cell-to-cell variation in gene expression not referenced in this manuscript; a complete literature review on non-genetic, non-environmental biological variation was beyond the scope of this study. We thank George Martin for thoughtful discussions about the manuscript.

Funding: Funding was provided by the National Institutes of Health, National Institute on Aging R01AG039025 to TEJ, National Institutes of Health, National Institute of General Medical Sciences R01GM97479 to RB, the National Institutes of Health National Institute on Aging K99/ R00AG045341 to AM, P50 AG005136 to MK, a training grant from the National Institute on Aging, T32AG000057, to support NB, and Pilot grant from the Nathan Shock Center for Excellence in the Basic Biology of Aging to AM (NIA Grant P30AG013280 to MK).

Author Contributions: AM, TEJ and RB designed the study. NB, AM and RB analyzed and interpreted the results. AM, RB, TEJ and MK provided funding. BS, PT, AM and SY generated DNA and transgenic nematode strains. AM and NB conducted the imaging experiments and performed image cytometry. AM, PT and BS conducted the flow sorting and phenotyping experiments. NB and AM analyzed the data. AM, RB and NB wrote the initial manuscript. RB, MK, TEJ, PT, BS, AM and NB reviewed and revised the manuscript.

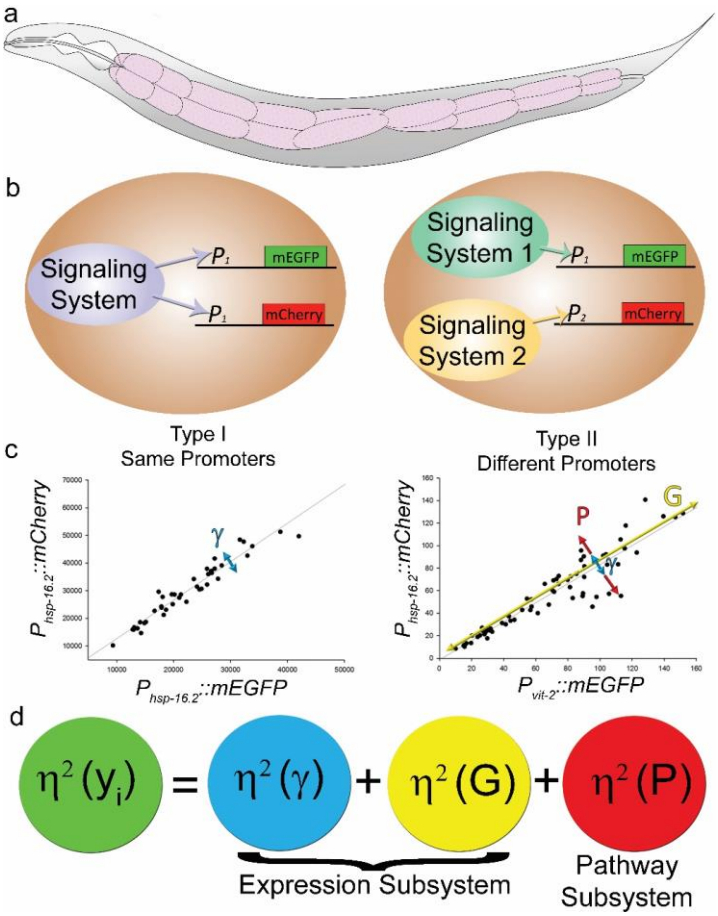
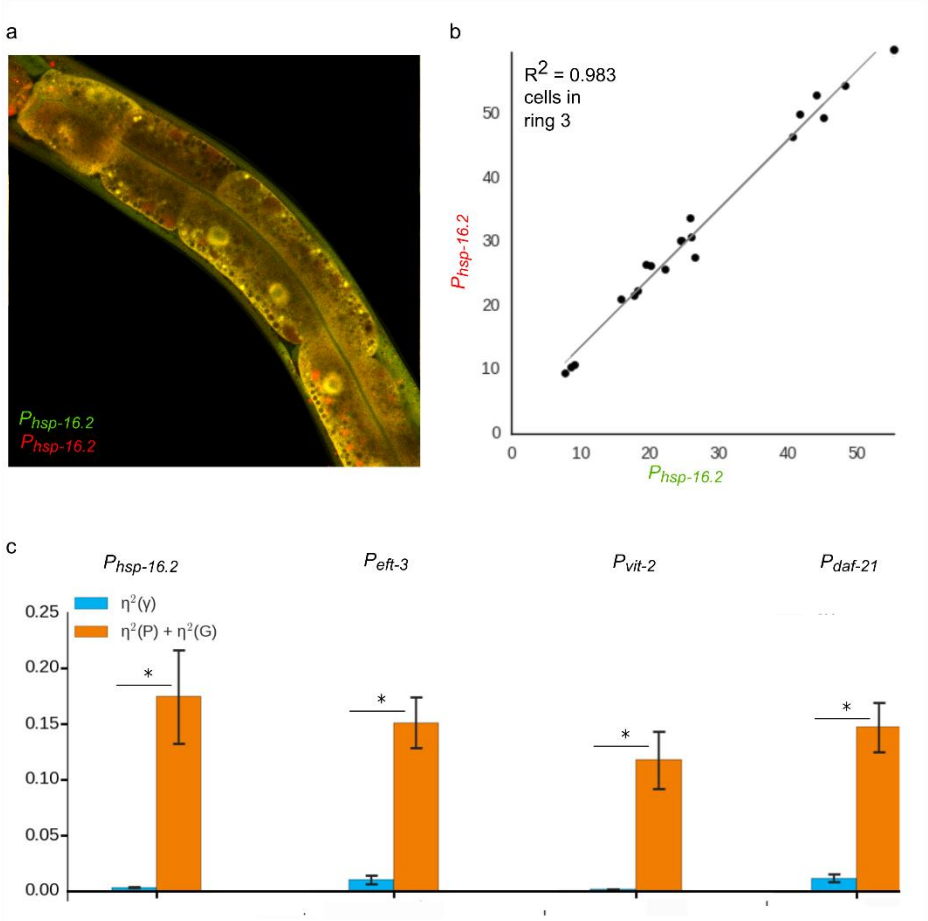
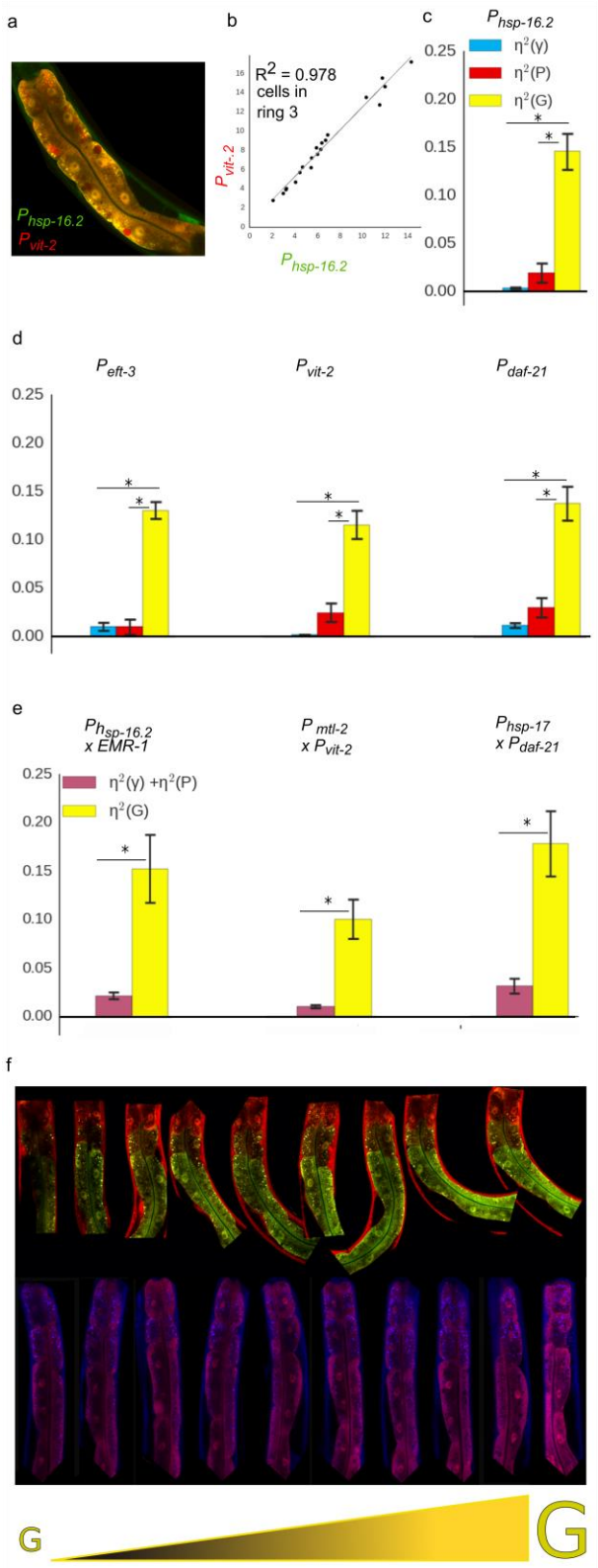


Figure 1. Experimental design and analytical framework. **a.** Schematic view of the *C. elegans* body. Intestine is colored pink. The twenty cells of the intestine are organized into nine segments, called rings. **b.** Overview of reporter gene measurement scheme. In the left panel, two differently colored copies of the same gene respond to the same signaling system (Type I experiments), or, in the right panel two differently colored reporters controlled by two different promoters respond to two distinct signaling systems (Type II experiments)**.** **c.** Data plotting schematic. We imaged intestine cells in rings one through four (the cells that fit in a single field of view with a 40x objective) in animals on day two of adulthood, at the point when lifespan can be predicted from $P_{hsp\text{-}16.2}$*::gfp* expression levels. We then plotted expression levels of reporter pairs expressed in the intestine cells, grouped by cell type; the same cell types (e.g., cells in ring three; int3 cells) from different animals are plotted. Left scatterplot shows typical results of type I experiment. Since both genes receive the signals from the same upstream regulators and share the same downstream expression machinery, uncorrelated variation of expression results only from stochastic noise of transcription/translation or variable allele access $\eta^2(\gamma)$. Right scatterplot shows typical results of type II experiment. In the case of two different genes, uncorrelated variation of expression can result from stochastic noise of transcription/translation or variable allele access $\eta^2(\gamma)$ and variation in activation of particular signaling pathways $\eta^2(P)$. *C*orrelated variation results from variation in shared gene expression machinery $\eta^2(G)$. **d.** A simplified version of the analytical framework derived from Colman-Lerner et al 2005 is shown; it depicts the three experimentally tractable bins into which we can attribute cell-to-cell variation in gene expression. Each bin has different underlying molecular causes. A complete, mathematically detailed description of the analytical framework is shown in Supplemental Materials Section 2.

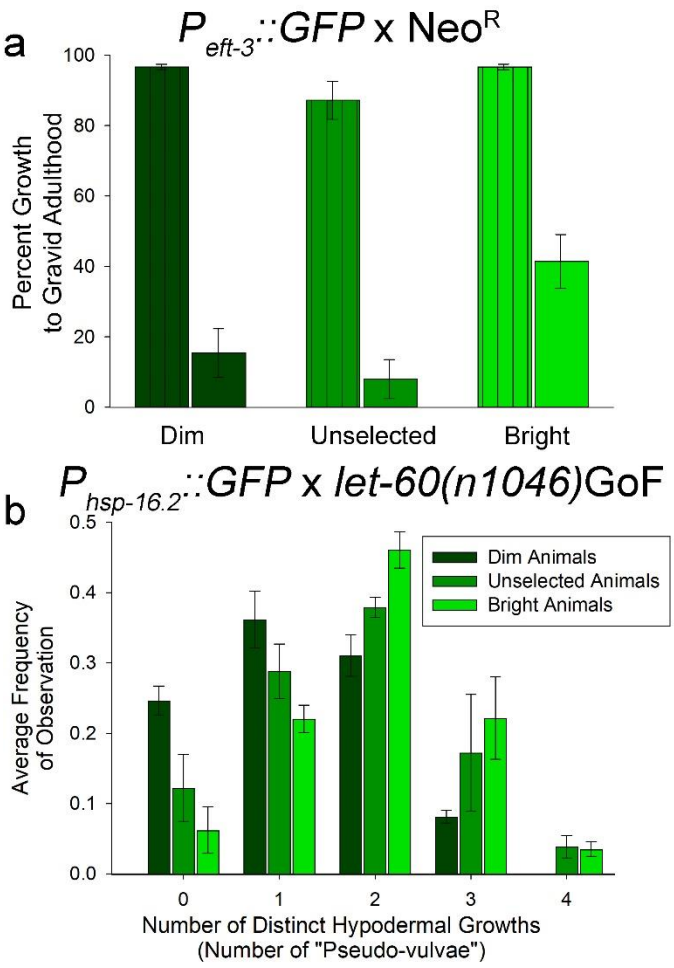
Figure 2



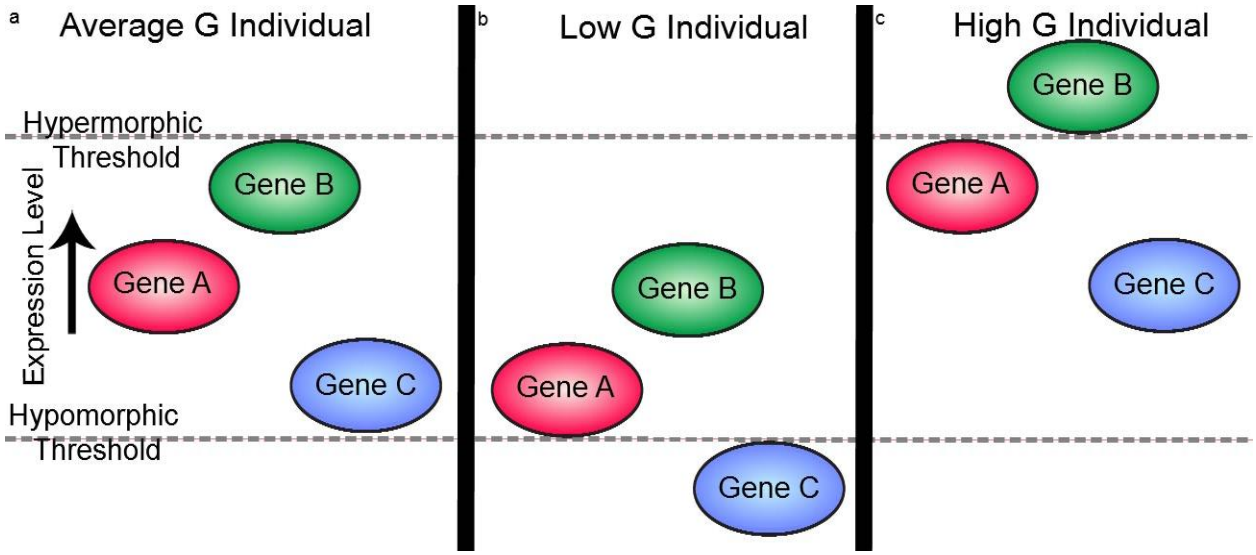
<u>Figure 2. Type I experiments reveal low intrinsic noise.</u> a. An example of the animal producing fluorescent reporters (mEGFP and mCherry) from two identical promoters ($P_{hsp-16.2}$). We did not observe patchwork of green and red cells. Cells were nearly uniformly yellow, indicating low intrinsic noise. **b.** Scatterplot of $P_{hsp-16.2}$::GFP and $P_{hsp-16.2}$::mCherry expression in cells in ring 3 of the intestine from ten animals examined in one experiment. **c.** Average correlated and uncorrelated variations for $P_{hsp-16.2}$, P_{eft-3}, P_{vit-2} and P_{daf-21} in the intestine cells in rings 1-4. In type I experiment uncorrelated variation arises from stochastic noise of transcription/translation or variable allele access – $\eta^2(\gamma)$; correlated variation is a combined result of variation in gene expression capacity $\eta^2(G)$ and variation in pathway activation $\eta^2(P)$. * indicates statistical significance P <0.05 by t-test.



<u>Figure 3. Variation in protein expression capacity, $\eta^2(G)$, is the major source of inter-individual variation in expression</u>. **a**. An example of the animal producing fluorescent reporters (mEGFP and mCherry) from two different promoters ($P_{hsp-16.2}$ and $P_{vit-.2}$). **b**. Scatterplot of $P_{hsp-16.2}$::GFP and P_{vit-2}::mCherry expression in the intestine cells in ring 3. **c**. Average stochastic noise $\eta^2(\gamma)$, variation in pathway activation $\eta^2(P)$ and variation in protein expression capacity $\eta^2(G)$ for $P_{hsp-16.2}$ promoter in cells in the intestine rings 1-4. * indicated statistical significance $p<0.05$ analyzed by one-way ANOVA with post-hoc Tukey's HSD test. **d**. Average stochastic noise $\eta^2(\gamma)$, variation in pathway activation $\eta^2(P)$ and variation in protein expression capacity $\eta^2(G)$ for P_{vit-2}, P_{eft-3}, P_{daf-21} in the intestine cells in rings 1-4. * indicated statistical significance $p<0.05$ analyzed by one-way ANOVA with post-hoc Tukey's test. **e**. Average correlated and uncorrelated variations for different promoters couples in the intestine cells in rings 1-4. Uncorrelated variation combines stochastic noise of transcription/translation or variable allele access – $\eta^2(\gamma)$ and variation in pathway activation $\eta^2(P)$. Correlated variation results from variation in protein expression capacity $\eta^2(G)$. * indicates statistical significance $p<0.05$ by t-test. **f**. G is a whole organism property. Multiple pairs of reporter genes covary; animals are generally brighter or dimmer for any pair of reporter genes we measure. Shown are individual animals that vary in expression level for both P_{mtl-2}*::GFP* and P_{daf-21}*::mCherry* (top panel), or, P_{eft-3}*::BFP* and P_{vit-2}*::mCherry* (bottom panel).



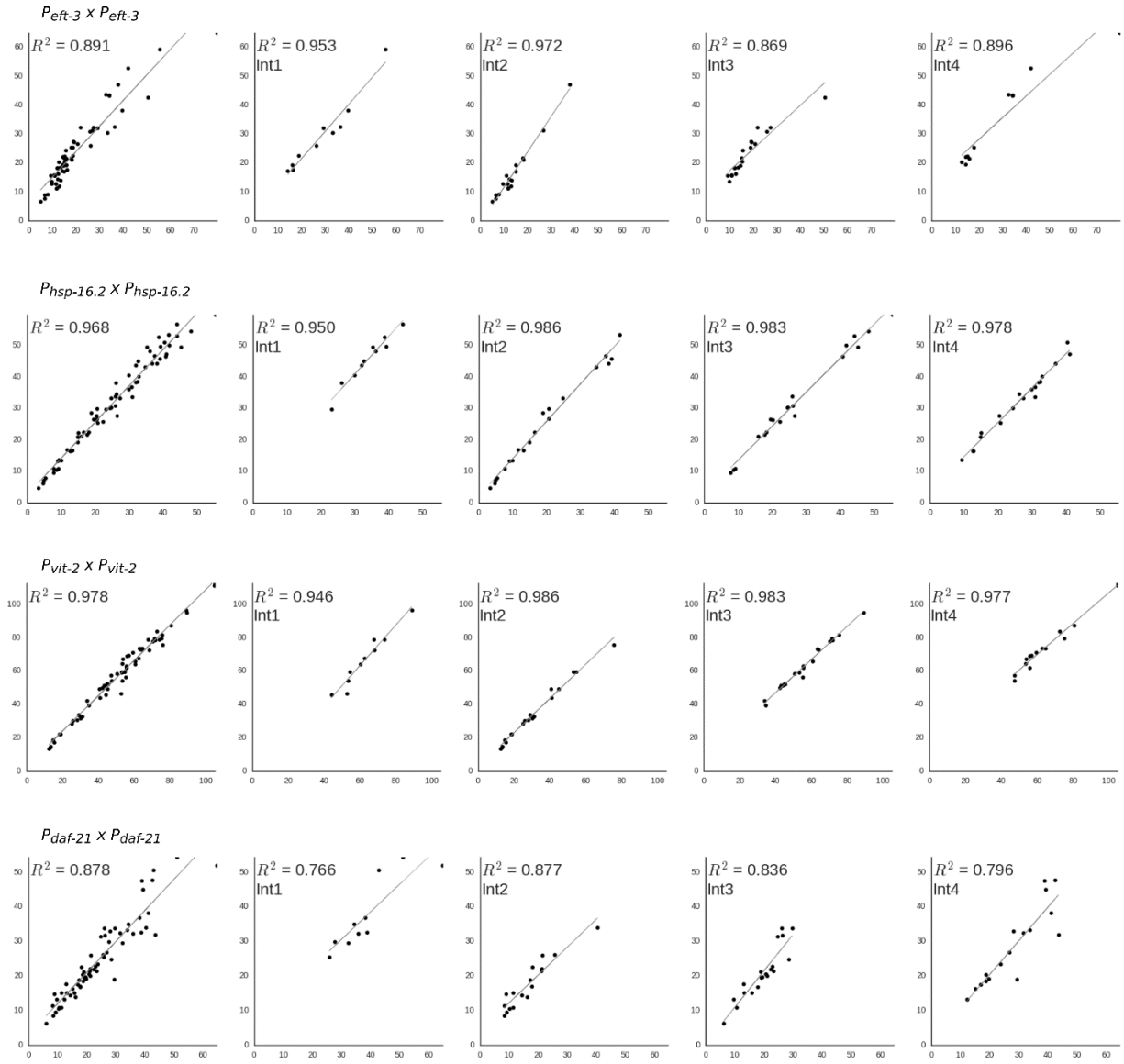
<u>Figure 4 Differences in gene expression capacity are associated with differences in penetrance and drug resistance</u>. a. Drug resistance conferred by Neo^R expressed from P_{rps-27} promoter correlates with expression of P_{eft-3}*::GFP* reporter. P_{eft-3}*::GFP* reporter does not predict ability to grow in the absence of neomycin (striped bars). High P_{eft-3}*::EGFP* expression predicts a significant difference in the ability to grow to adulthood on a high neomycin concentration (solid bars; $P < 0.05$, Paired t-test). See Tables S1 and S2 for numerical data. **b**. Penetrance of Ras/let-60 gain of function mutation *let-60(n1046)* correlates with the ability to express $P_{hsp-16.2}$*::GFP* reporter. Chi Square for distributions of bright or dim vs unselected $P < 0.0001$. Difference in penetrance between bright and dim $P < 0.05$, Paired t-test. See Tables S3a-c for numerical data.



<u>Figure 5 Phenotypic expression of allelic forms of Gene A, Gene B, and Gene C</u>. a. An individual with average G and wild-type phenotypes. **b.** An individual with low G and penetrance of a hypormorphic phenotype for the allele of Gene C. **c.** An individual with high G and penetrance of a hypermorphic phenotype for the allele of Gene B.

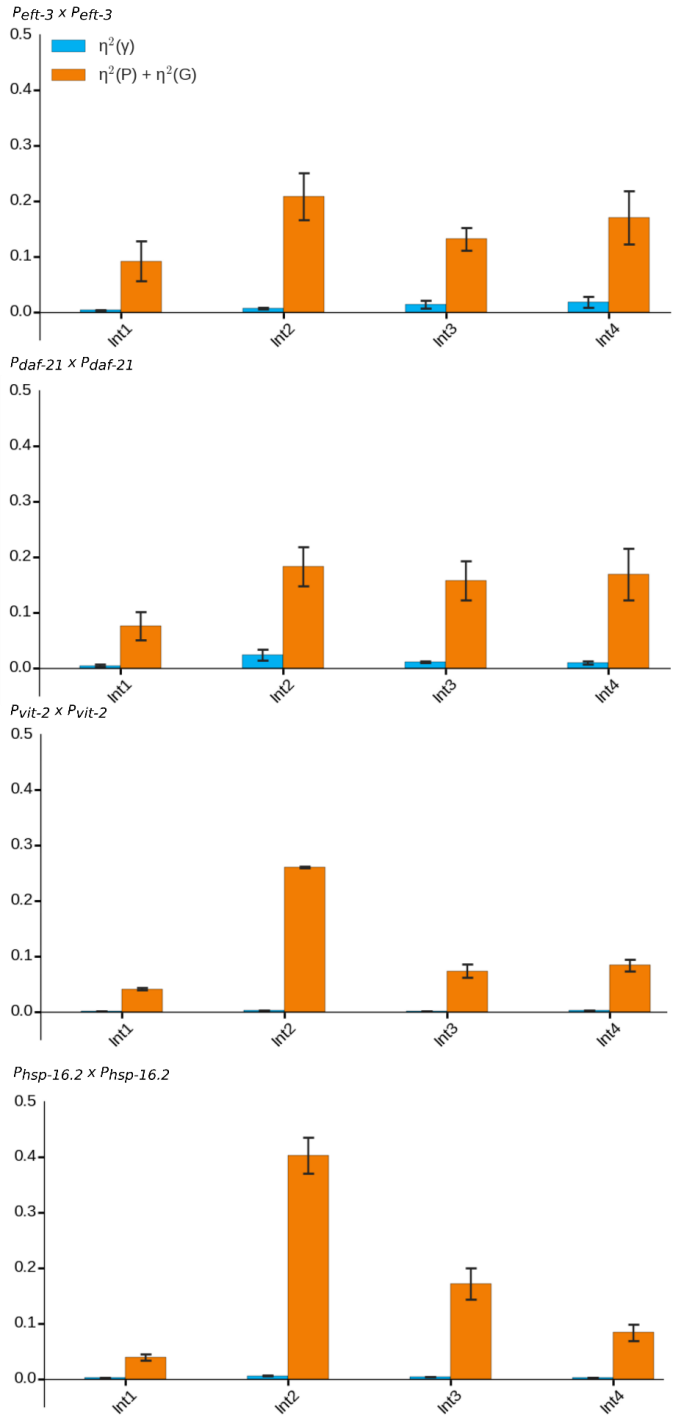
Supplemental Figure 1. Type I experiments supplemental scatterplots.

Scatterplots of two identical promoters with different fluorescent protein outputs expressed from homologous chromosomes (type I experiment). Far left scatterplots show all cells measured in a given experiment. Scatterplots on the right show expression of reporters in the cells from particular intestine rings.



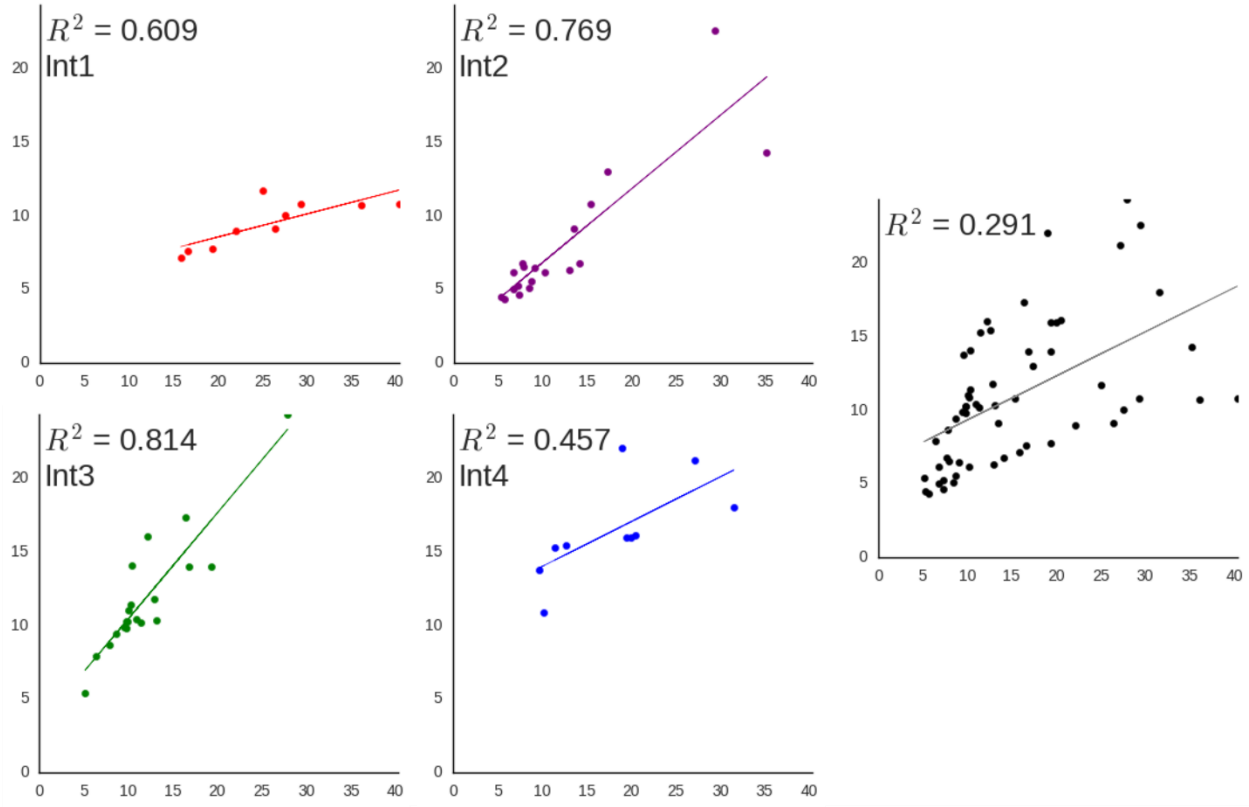
Supplemental Figure 2. Type I experiments supplemental bargraphs.

Correlated and uncorrelated variation for $P_{hsp\text{-}16.2}$, $P_{eft\text{-}3}$, $P_{vit\text{-}2}$ and $P_{daf\text{-}21}$ based reporter genes in intestine cells in rings 1-4. In type I experiment uncorrelated variation arises from stochastic noise of transcription/translation or variable allele access – $\eta^2(\gamma)$; correlated variation is a combined result of variation in gene expression capacity $\eta^2(G)$ and variation in pathway activation $\eta^2(P)$.



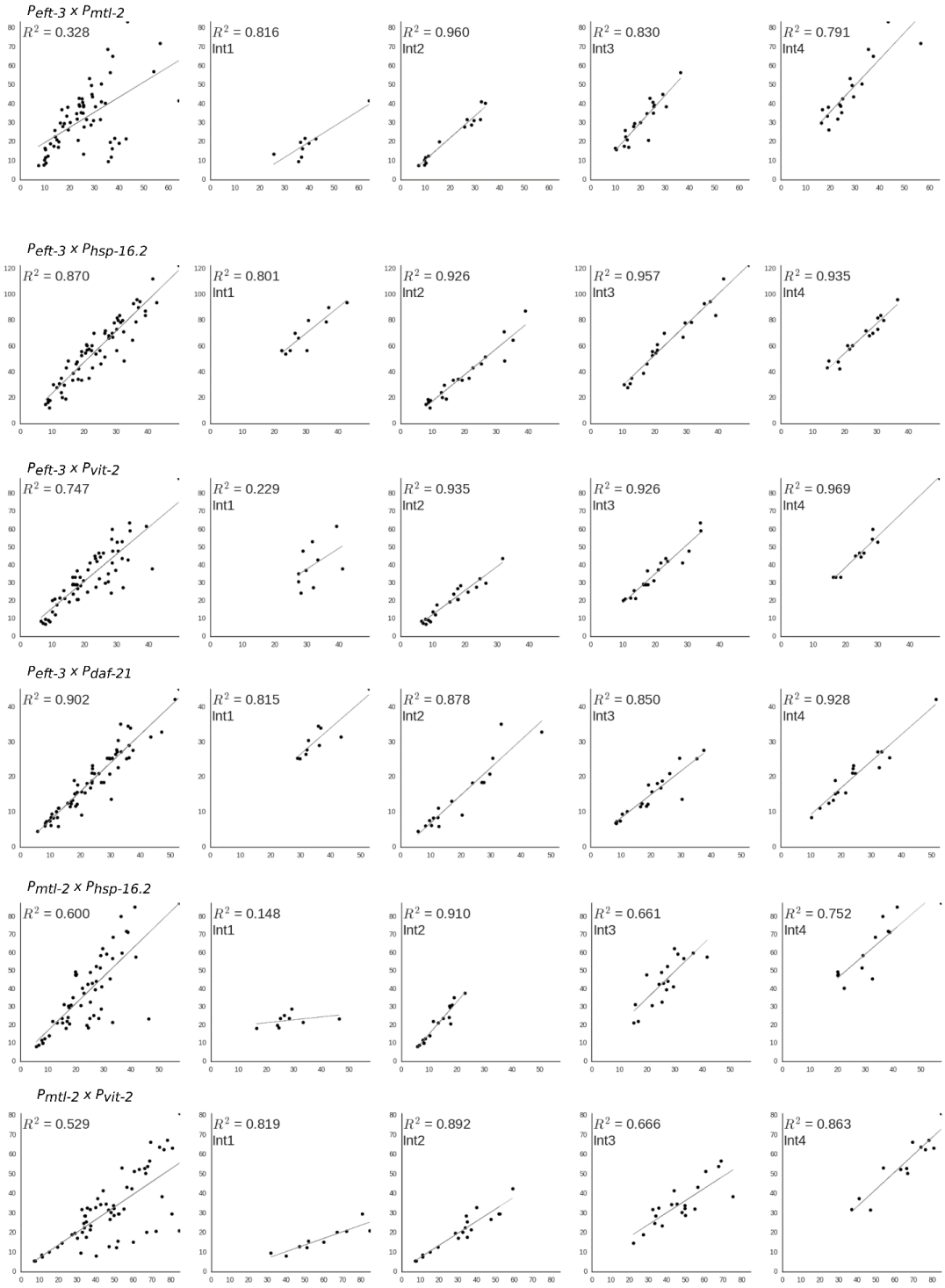
Supplemental Figure 3. Evidence for splitting cells by fate for analysis in type II experiments.

Scatterplots of expression of two reporters from the different reporter genes, $P_{hsp\text{-}17}$ and $P_{mtl\text{-}2}$, grouped by ring (one through four) or combined in one plot (right panel). Cell fate determines ratiometric setpoint for expression of two distinct genes. When cells are not split by fate, correlation of expression is quite low. Correlation of expression is much higher when cells are grouped by fate.



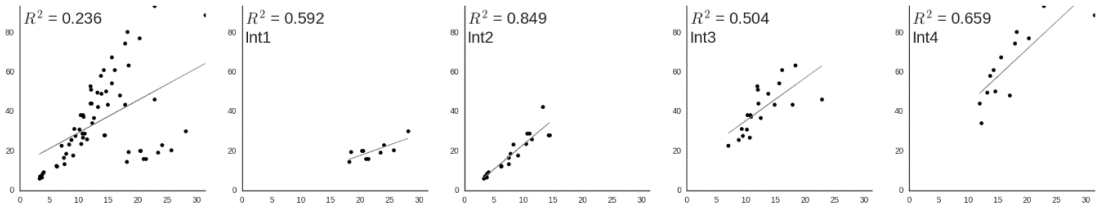
Supplemental Figure 4. Type II experiments supplemental scatterplots.

Scatterplots of expression of two distinct reporter genes (type II experiment). Far left scatterplots show all cells measured in a given experiment. Scatterplots on the right show expression of reporters in the cells from particular intestine rings.

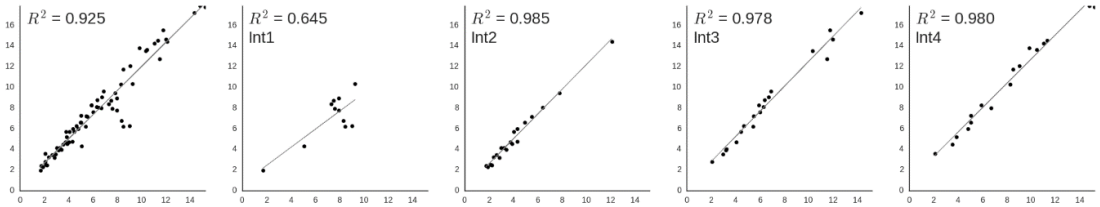


Type II experiment scatterplots Part II

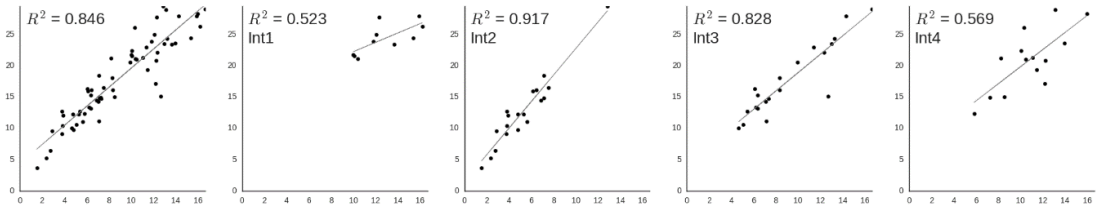
P_{mtl-2} x P_{daf-21}



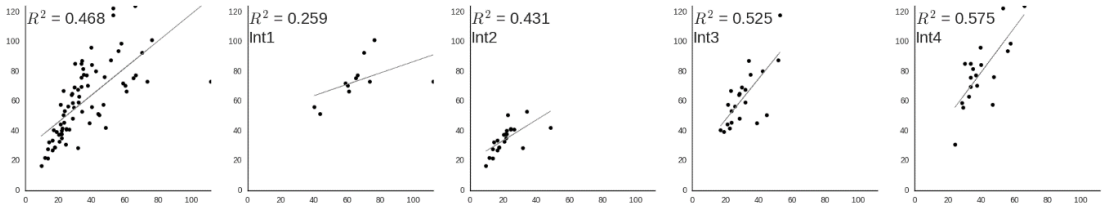
$P_{hsp-16.2}$ x P_{vit-2}



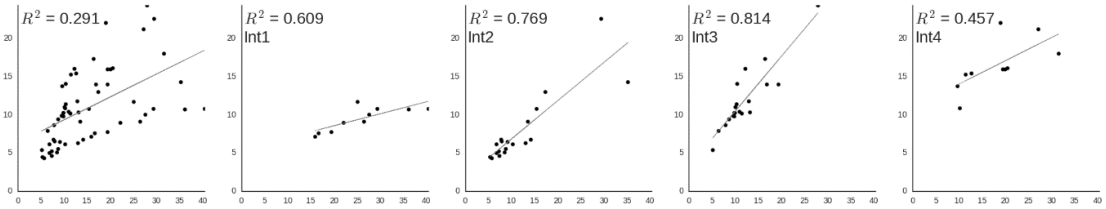
$P_{hsp-16.2}$ x P_{daf-21}



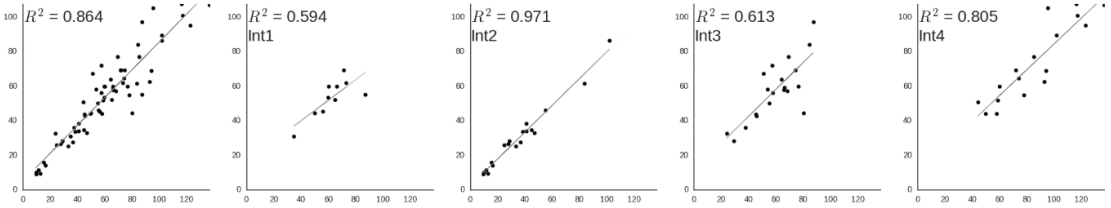
P_{daf-21} x P_{vit-2}



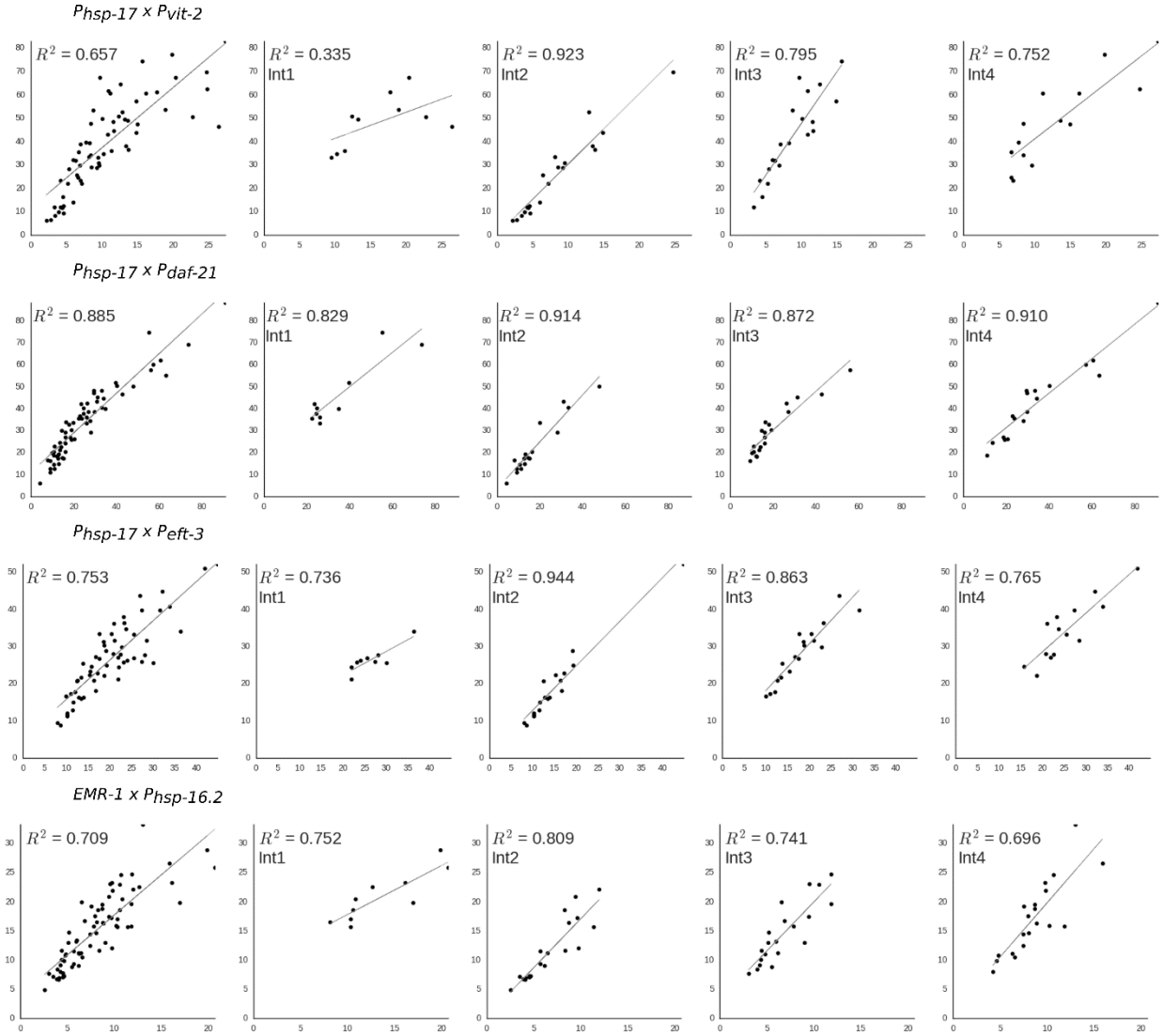
P_{hsp-17} x P_{mtl-2}



P_{hsp-17} x $P_{hsp-16.2}$

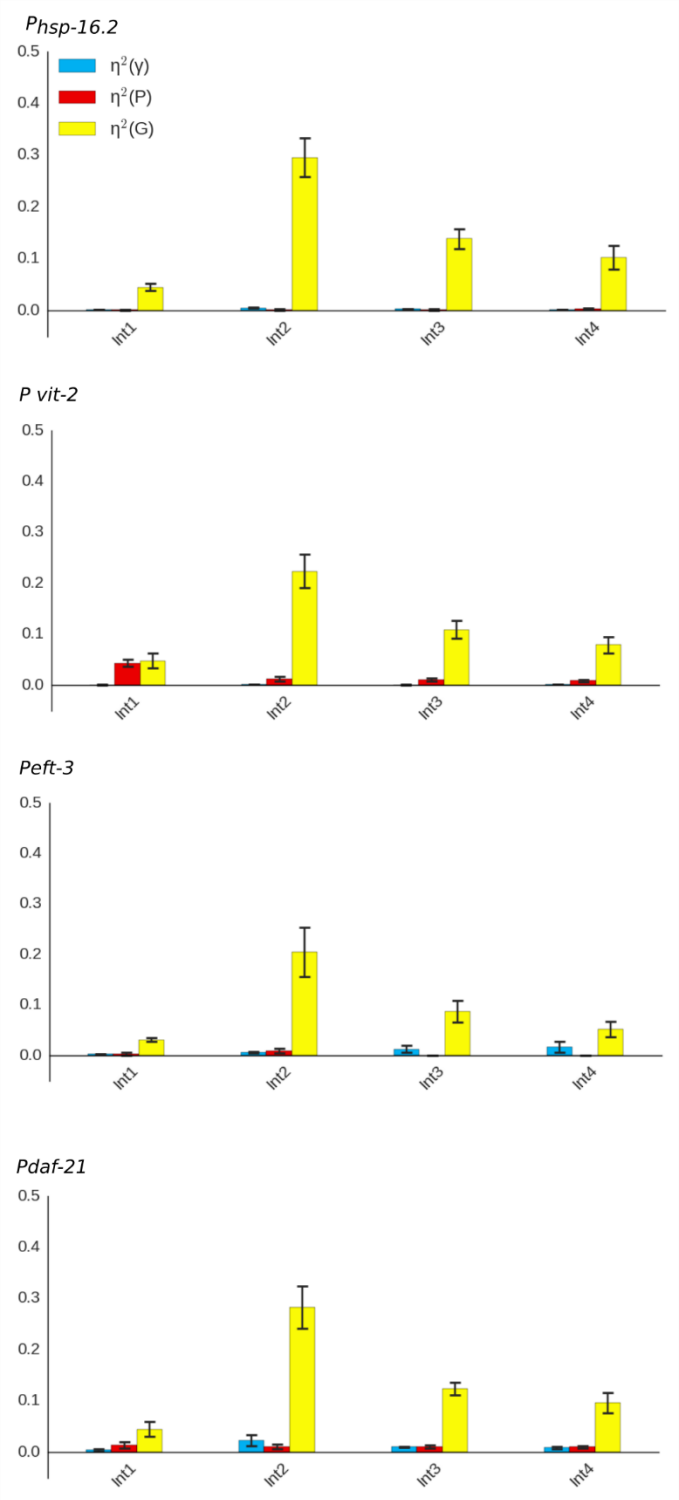


Type II experiment scatterplots Part III



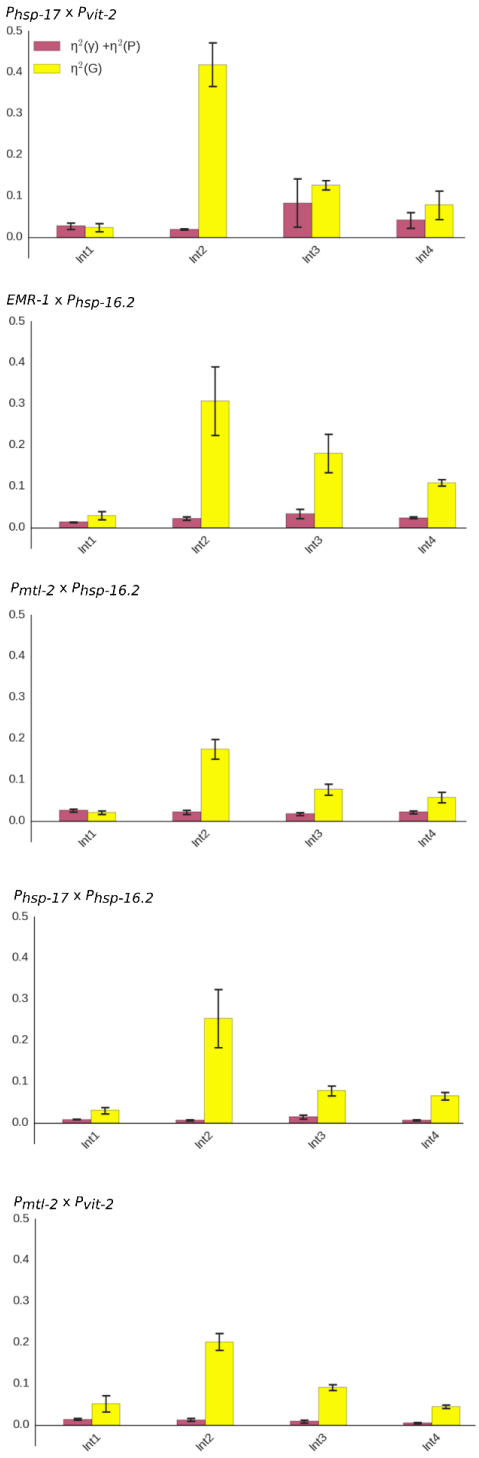
Supplemental Figure 5. Type II experiments supplemental bargraphs.

Stochastic noise $\eta^2(\gamma)$, variation in pathway activation $\eta^2(P)$ and variation in gene expression capacity $\eta^2(G)$ for $P_{hsp\text{-}16.2}$, $P_{vit\text{-}2}$, $P_{eft\text{-}3}$ and $P_{daf\text{-}21}$ for cells in intestine rings 1-4.



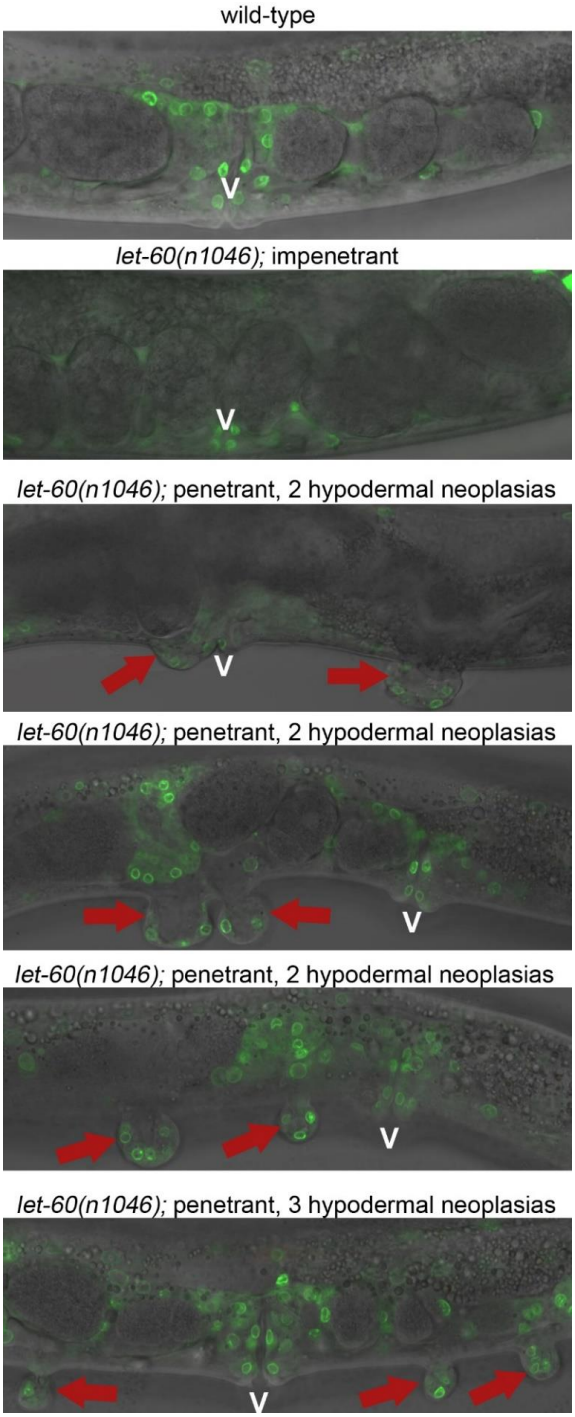
Supplemental Figure 6. Type II experiments supplemental bargraphs (no promoter splitting)

Correlated and uncorrelated variation for different gene pairs in cells in intestine rings 1-4. Uncorrelated variation combines stochastic noise of transcription/translation or variable allele access – $\eta^2(\gamma)$, and variation in pathway activation $\eta^2(P)$. Correlated variation results from variation in gene expression capacity $\eta^2(G)$.



Supplemental Figure 7. Penetrance and Expressivity of Ras Gain of Function Mutation in *C. elegans.*

Animals express EMR-1::GFP to mark cell nuclei. Wild type and *let-60(n1046)* animals are shown in the top two panels. Some *let-60* mutants do not develop extra pseudo-vulvae/neoplasias (impenetrant, compare to wild type). Other *let-60* animals exhibit variable number of hypodermal neoplasias.



<u>Supplementary Figure 8. Working Model for how Differences in G Might Arise.</u> a) Genetic experiments from Mendenhall et. al. [24] show that interindividual variation in the *hsp-16.2* reporter gene that is a biomarker for mutation penetrance and lifespan arises from differences in insulin signaling and depolarization of the AFD neuron pair. Thus, animals with low insulin signaling and high neuronal depolarization would have the highest expression of the *hsp-16.2* lifespan/penetrance biomarker and have the longest lifespans, highest penetrance of hypermorphic phenotypes and lowest penetrance of hypomorphic phenotypes. **b)** Diagram of the possible molecular differences in cells that might underlie global differences in protein dosage. Solid arrows are directly supported by experiments in *C. elegans* using *hsp-16.2* reporter genes, and arrows with question marks are possibilities supported by less direct experimental evidence (see discussion). Here we currently favor the interpretation that decreased protein turnover, and not protein production, results in higher protein dosage, at least in part because we showed animals with high biomarker expression transcribe less of a AAATPase ribosome export homolog, *mac-1*[13], consistent with the idea that brighter animals are not making more protein. Furthermore, isogenic animals with lower insulin signaling, measured by the duration of DAF-16 nuclear translocation (which would also have higher *hsp-16.2* biomarker expression[24]), produce fewer progeny[9], also consistent with the idea that animals with high biomarker expression do not produce more protein, but instead, maintain existing protein longer, resulting in higher effective protein dosages.

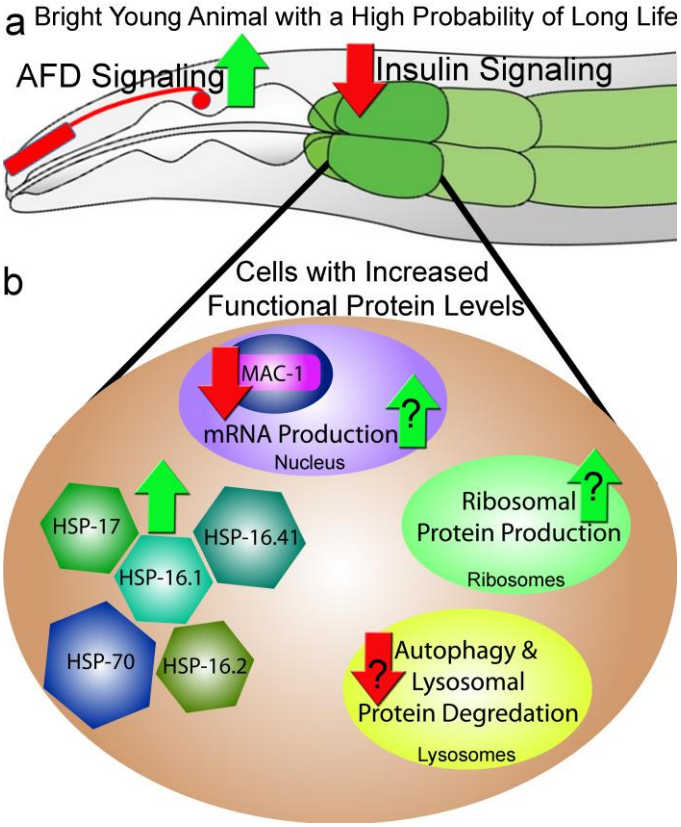


Table S1 Growth to gravid adulthood after 72 hours development on regular NGM

Experiment	Bright adulthood	Bright total	Bright % adulthood	Dim adulthood	Dim total	Dim % adulthood	Unselected adulthood	Unselected total	Unselected % adulthood
1							37	45	82%
2									
3	46	48	96%	46	48	96%	48	49	98%
4									
5	38	39	97%	38	39	97%	36	44	82%

Table S2 Growth to gravid adulthood after 96 hours on high Neomycin concentration NGM

Experiment	Bright neo adulthood	Bright neo total	Bright neo % adulthood	Dim neo adulthood	Dim neo total	Dim neo % adulthood	Unselected neo adulthood	Unselected neo total	Unselected neo % adulthood
1	15	48	31%	2	43	5%	1	45	2%
2	26	46	57%	1	42	2%			
3	25	49	51%	15	51	29%	9	47	19%
4	25	48	52%	15	43	35%			
5	6	37	16%	2	36	6%	1	40	3%

Table S3a Number of hypodermal neoplasias

Experiment 1

# Growths	Dim		Unselected		Bright	
	# Animals	% Animals	# Animals	% Animals	# Animals	% Animals
0	10	23%	2	3%	0	0%
1	19	44%	15	21%	7	19%
2	11	26%	25	35%	15	42%
3	3	7%	24	34%	12	33%
4	0	0%	5	7%	2	6%
Total # Animals	43		71		36	

Table S3b Number of hypodermal neoplasias

Experiment 2

# Growths	Dim		Unselected		Bright	
	# Animals	% Animals	# Animals	% Animals	# Animals	% Animals
0	27	29%	31	16%	6	7%
1	30	32%	66	34%	21	26%
2	30	32%	79	40%	41	51%
3	7	7%	16	8%	11	14%
4	0	0%	5	3%	2	2%
Total # Animals	94		197		81	

Table S3c Number of hypodermal neoplasias

Experiment 3

# Growths	Dim		Unselected		Bright	
	# Animals	% Animals	# Animals	% Animals	# Animals	% Animals
0	42	22%	45	18%	18	11%
1	62	32%	79	32%	33	21%
2	68	36%	95	38%	73	46%
3	19	10%	24	10%	31	19%
4	0	0%	5	2%	4	3%
Total # Animals	191		248		159	

Supplemental Material Section 1: Materials and Methods

Genome Engineering and Creation of *C. elegans* Strains

We generated strains for these studies using MosSCI transgenesis[42]. We also used some transgenes from prior studies that were multicopy; RBW2 contains a biolistically integrated transgene (*emr-1::gfp)* and was described previously[14]. TJ2741 contains a multicopy version of $P_{hsp\text{-}16.2}$*::gfp*, *zIS2735* [$P_{hsp\text{-}16.2}$*::egfp::*$T_{unc\text{-}54}$] V, which shows the same physiological worm to worm variation as the several other $P_{hsp\text{-}16.2}$ reporters[14]. We used this particular version because it was genetically compatible with the *let-60(n1046)* IV mutation and enabled our sorter to detect signal in the heatshock-pathway-attenuating *let-60(n1046)* background. We made the DNA constructs fusing promoter sequences to the coding sequence of different fluorescent proteins (XFPs) and the terminator of *unc-54* by using yeast gap repair[43]. All reporter constructs carried the 5'UTR (the upstream regulatory sequences including the promoter) of the selected genes (i.e., $P_{hsp\text{-}16.2}$) fused to an XFP coding sequence (*megfp*, *mcherry*, *mtagbfp2*, *mneptune*) and the 3'UTR of *unc-54*. We integrated fluorescent reporters into Chromosome II in the parent strain RBW6699 (an outcrossed version of EG6699 without an extrachromosomal array; *ttTi5605 II; unc-119(ed9)*), unless otherwise noted. To construct DNA for transgenes, we introduced reporter sequences into a vector that targeted the DNA to be inserted at the chromosome II MosSCI transposon at site *ttTi5605*. We then injected each construct at 50ng/µL into EG6699 animals and recovered single copy insertions that we size and site validated by PCR. Thus, through transformation and/or standard genetic crosses, we created the following strains:

RBW2661 (*hutSi2661[*$P_{daf\text{-}21}$*::megfp::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2642 (*hutSi2642[*$P_{daf\text{-}21}$*::mcherry::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2601 (*hutSi2601[*$P_{hsp\text{-}16.2}$*::megfp::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2561 (*hutSi2561[*$P_{hsp\text{-}16.2}$*::mcherry::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2621 (*hutSi2621[*$P_{vit\text{-}2}$*::megfp::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2581 (*hutSi2581[*$P_{vit\text{-}2}$*::mCherry::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), RBW2531 (*hutSi2531[*$P_{mtl\text{-}2}$*::mcherry::*$T_{unc\text{-}54}$ + *Cbr-unc-119(+)]* II), ARM1

(*let-60(n1046)*) IV; *jjIs699[P_{lmn-1}::emr-1::gfp::T_{unc-54} + unc-119(+)]* ?, not II), ARM6 (*wamSi6[P_{eft-3}::mtagbfp2::T_{unc-54} + Cbr-unc-119(+)]* II), ARM5 (*wamSi5[P_{eft-3p}::mneptune::T_{unc-54} + Cbr-unc-119(+)]* II), RBW3211 (*hutSi3211[P_{hsp-17}::megfp:T_{unc-54} + Cbr-unc-119(+)]* II), TJ2741 ((*zIs2735[$P_{hsp-16.2}$::gfp::T_{unc-54}]*) V; *let-60(n1046)* IV), RBW99 (*oxSi259[P_{eft-3}::megfp::T_{unc-54} + Cbr-unc-119(+)]* I; *oxTi179[P_{rps-27}::Neo^R::T_{unc-54} + unc-18(+)]* II); RBW2 (*zSi3002[$P_{hsp-16.2}$::mcherry::T_{unc-54} + Cbr-unc-119(+)*] II; *jjIs699[P_{lmn-1}::emr-1::gfp::T_{unc-54} + unc-119(+)]* ?)

Table S4: List of crosses to make different F1 animals for Type I & II Experiments
P_{daf-21}*::gfp* (RBW2661) x P_{daf-21}*::mcherry* (RBW2642)
P_{daf-21}*::gfp* (RBW2661) x P_{vit-2}*::mcherry* (RBW2581)
$P_{hsp-16.2}$*::gfp* (RBW2601) x P_{daf-21}*::mcherry* (RBW2642)
$P_{hsp-16.2}$*::gfp* (RBW2601) x $P_{hsp-16.2}$*::mcherry* (RBW2561)
$P_{hsp-16.2}$*::gfp* (RBW2601) x P_{vit-2}*::mcherry* (RBW2581)
P_{vit-2}*::gfp* (RBW2621) x P_{vit-2}*::mcherry* (RBW2581)
P_{mtl-2}*::mcherry* (RBW2531) x P_{daf-21}*::gfp* (RBW2661)
P_{mtl-2}*::mcherry* (RBW2531) x $P_{hsp-16.2}$*::gfp* (RBW2601)
P_{mtl-2}*::mcherry* (RBW2531) x P_{vit-2}*::gfp* (RBW2621)
P_{eft-3}*::mtagbfp2* (ARM6) x P_{daf-21}*::mcherry* (RBW2642)
P_{eft-3}*::mtagbfp2* (ARM6) x $P_{hsp-16.2}$*::mcherry* (RBW2561)
P_{eft-3}*::mtagbfp2* (ARM6) x P_{mtl-2}*::mcherry* (RBW2531)
P_{eft-3}*::mtagbfp2* (ARM6) x P_{vit-2}*::mcherry* (RBW2581)
P_{eft-3}*::mtagbfp2* (ARM6) x P_{eft-3}*::mneptune* (ARM5)
P_{hsp-17}*::gfp* (RBW3211) x P_{daf-21}*::mcherry* (RBW2642)
P_{hsp-17}*::gfp* (RBW3211) x $P_{hsp-16.2}$*::mcherry* (RBW2561)
P_{hsp-17}*::gfp* (RBW3211) x P_{mtl-2}*::mcherry* (RBW2531)
P_{hsp-17}*::gfp* (RBW3211) x P_{vit-2}*::mcherry* (RBW2581)
P_{hsp-17}*::gfp* (RBW3211) x P_{eft-3}*::mneptune* (ARM5)
$P_{hsp-16.2}$*::mcherry* (RBW2561) x *emr-1::gfp* (stable RBW2)

Culture conditions

We maintained animals on NGM plates seeded with live OP50 *E. coli* at 20°[44]. We ensured that the stocks were not starved for at least two generations prior to use for experimentation. Single-color reporter animals (see above) were mated to produce two-color F1 progeny for microscopy analysis. As a result we analyzed expression of reporters driven by identical or different promoters integrated at the identical loci on both copies of chromosome II. The only exception is RBW2 strain that carried a transcriptional reporter ($P_{hsp\text{-}16.2}$*::mCherry*) and *emr-1::gfp* fusion transgene on different chromosomes and was maintained as a stable strain. Animals that carried $P_{hsp\text{-}16.2}$ – driven reporters were heat shocked on the first day of adulthood by exposure to 35°C for one hour on a solid NGM plate. All reporter strains were analyzed on the second day of adulthood. We analyzed crosses listed in Table S4.

Mounting animals for microscopy

We prepared animals for microscopy as previously described in[14]. Briefly, we prepared a 1% agarose pad (about 1-2 cm^2) on a standard glass microscope slide. We used worm anesthesia solution containing 0.2% tricaine and 0.02% tetramisole (Sigma, Inc., St. Louis) in M9. We carefully picked animals off their NGM plates and into the drop of anesthesia on the pad. We allowed the worms to swim off the pick into the drop. After that, we covered animals with a glass coverslip. We waited until worms ceased most active movements (about 15-25 minutes), and placed the prepared slide on the slide holder mounted on the microscope stage.

Microscopic image acquisition

We acquired images of animals expressing multiple fluorescent proteins in a similar fashion to Mendenhall et al, 2015[14]. Briefly, we used a 40X 1.2 NA water immersion objective and acquired a z stack of images of each animal in one field of view. Thus, we were able to capture intestine cells in rings one through four of the intestine. We used minimal laser power to prevent

photobleaching and heat damage to the cells. We adjusted gain and laser power to ensure that the reporter signals fit within the dynamic range of the photomultiplier tubes and did not saturate them. We only imaged animals lying on their left side to prevent signal loss attributable to imaging through the gonad. We digitally rotated the field of view to arrange the animals into diagonal orientation to fit maximal number of intestine cells in a field of view. We then acquired sequential z-slices to span most of the intestine cells' depth. We acquired an optical slice every two micrometers from the proximal starting point at the objective-proximal side of the intestine, and continued imaging until we acquired signal from intestine cell nuclei in the first four rings. We set optical slice thickness to two micrometers (all photons from plus or minus two micrometers of the image plane). Imaging of each animal typically took two minutes at 1024x1024 resolution. Each pixel value was the average of four samples. We imaged approximately ten animals per experiment.

Image Cytometry

We first determined the orientation of the animals in images and then identified individual cells as described previously, in Mendenhall et al, 2015[14]. As the intestine is anchored anteriorly and posteriorly we used the known positions of the anteriormost cells to determine the identity of adjacent cells. Once we identified a cell, we measured signal within an equatorial slice of the cell's nucleus, as a proxy for the whole cell, as previously described; the nuclear signal is nearly perfectly correlated with the cytoplasmic contents. We used the ImageJ software as well as custom built Nuclear Quantification Support Plugin for nucleus segmentation and signal quantification, we call C Entmoot (Alexander Seewald, Seewald Solutions, Inc., Vienna). The algorithm uses a meeting of decision trees (an Ent Moot – a meeting of tree-beings; from J. R. R. Tolkien's *The Lord of the Rings*) used to delineate a nuclear boundary based on the drop in signal intensity, trained on user delineated images (Seewald et al, manuscript in preparation). The values of the nuclei from binucleate cells were averaged.

Sorting Animals in Flow

For an extensive description of our flow cytometry methods, please see Mendenhall et al, 2015[14]. Also bear in mind that these experiments require the growth of large amounts of animals for sorting – more so than for simply measuring the distribution of a population. Briefly, for the animals bearing the *let-60* mutation to be sorted on $P_{hsp\text{-}16.2}$*::EGFP* expression level, we grew a hypochlorite-synchronized population of animals to gravid adulthood, as in [8,14]. For animals we would sort during the L1 diapause on $P_{eft\text{-}3}$*::GFP* expression level, we performed a hypochlorite synchronization and allowed the animals to hatch and enter the L1 diapause in a 3-5mm deep pool of S-basal swirling at approximately 30 rpms on an unseeded 10cm NGM petri dish. At the start of worm flow sorting, we first flowed a sample of the population of animals to determine the distribution of values in order to select the animals at the extremes of the distribution of fluorescent reporter gene signal; we measured between 2-300 animals to get an estimate of the distribution of the population. We then selected animals at the top or bottom 5-10%, or selected all animals, and sorted between 50-250 animals per group (bright dim, unselected). Actual numbers of animals examined were often less due to loss of animals from escape onto the side of the plate. We hid the identity of each group from the person scoring to prevent bias in scoring. Scoring proceeded immediately after sorting for $P_{hsp\text{-}16.2}$*::EGFP* expression levels for adult animals bearing the *let-60(n1046)* mutation. To score, we counted the number of distinct growths emanating from the ventral hypodermis. A distinct growth was considered a protrusion from the soma unconnected to any other apparent protrusion; we verified that these protrusions were caused by the growth of extra cells using strain ARM1, shown in Supplemental Figure 7. The *let-60* gain of function mutation decreases the expression level of the heatshock response, necessitating the use of the multicopy insertion to ensure detection of GFP signal in this genetic background. We showed previously that the single copy and multicopy reporters have identical lifespan prediction capabilities and the same amount of worm-to-worm variation in biomarker expression. For larvae

expressing Neo^R, we used a high magnification stereoscope and scored animals for development to gravid adulthood (at least one fertilized embryo in the uterus) after 96 hours of development on food, after being sorted on $P_{eft\text{-}3}$*::GFP* expression while in the L1 diapause (72 hours for experiments on regular NGM).

Statistical Analysis

We used Sigma Stat (Systat Software, Inc., San Jose) for statistical analyses of the bright and dim animals we sorted. We first determined if measurements comprising each dataset were normally distributed, then, depending on the results of those tests, used appropriate parametric or non-parametric statistics to determine if there was significant difference in any measured parameters. Details of specific tests are shown in figure legends. Additional details regarding grouping of different types of measurements for calculations of η^2(G, P, or γ) and statistical analyses of different variation bins see Supplementary Materials Section 2: Analytical Framework.

Supplemental Material Section 2: Analytical Framework

S2.1 Analytical Framework: Defining Pathway Output and Expression Capacity

For quantitative analysis we adapted the analytical framework developed previously by Coleman-Lerner et al and briefly described below. We considered the production of fluorescent proteins (e.g. GFP and mCherry) in individual cells to be a measure of activities of the promoters that regulate their expression. Furthermore, to understand cell-to-cell and animal-to-animal differences in activities of various promoters we considered the amount of fluorescent protein produced to be the product of two subsystems: “pathway” and “expression”.

“Pathways” is the first subsystem. For instance, for a heat shock inducible promoter $P_{hsp\text{-}16.2}$, the input to pathway is the activation of transcription factors, e.g. HSF-1, that bind to *hsp-16.2* promoter, and the output is the activation of the inducible promoter that drives the fluorescent protein. Pathway output depends on the summed activity of upstream, DNA-bound transcription factors.

For a given cell *i,* pathway output is proportional to P_i, the time-averaged of the level of pathway activation. We separate P_i as $P_i = L_i + \lambda_i$ where L_i is the expectation value of P_i for a given cell type in a cohort of isogenic age-matched animals, and λ_i is the stochastic fluctuation term for that cell. We refer to L_i as the pathway power. The pathway “power” is a function of the activities of upstream signaling molecules and transcription factors that lead to activation of the promoter. We refer to cell-to-cell differences in P_i as “variation in pathway power”, and we describe differences that result from the stochastic fluctuation term λ_i as “transmission noise”. Thus, we can treat the output of the cell as being decomposed into an expectation value, which depends on the number and activity of the molecules that comprise the pathway, and a stochastic fluctuation that occurs because of the inherent randomness of the activity of the pathway, as well as its composition over the time of the experiment. Different cells may have different pathway capacities, and therefore, can have different values of L_i.

The second subsystem consists of the sequence of events from gene transcription through protein translation. We call this subsystem "*expression*". It includes transcriptional initiation, elongation, mRNA maturation, nuclear export, and mRNA translation and degradation (of protein and mRNA). We use our data to measure cell-to-cell variation in expression, but we cannot isolate the contribution of each part of the expression machinery to the overall variation.

The output of the expression subsystem is the total amount of the fluorescent reporter protein, and the input is proportional to the level of promoter activity (the output of the pathway subsystem). For analytical purposes, we assume here that the expression per unit of input is independent of the level of input. For cell *i,* we describe the expression per unit input as the variable E_i, where E_i is given by the sum of G_i and γ_i . The quantity G_i is the expectation value of the expression per unit input for cell *i,* and this quantity may vary from cell to cell. We refer to G_i as the "expression capacity" for cell *i*, and we refer to the cell-to-cell differences in G_i as "variation in expression capacity". The quantity γ_i is the stochastic fluctuation that occurred in cell *i*, and we refer to differences that result from this stochastic fluctuation term as "expression noise."

Using this model, we described the total amount of GFP, y_i, in cell *i* as

$$y_i = (L_i + \lambda_i) \text{ x } (G_i + \gamma_i) \text{ x } \Delta\text{T} \qquad (1)$$

where ΔT is the time of reporter protein production. For heat inducible promoter we can consider ΔT as the time since the heat shock. For constitutive promoters like *vit-2* we cannot precisely measure what fraction of the observed reporter protein pool was produced in the same time frame. Thus, we started this work under "steady state" assumption that reporter protein synthesis and degradation are balanced in young animals. In that case reporter proteins level directly mirror activity of the promoters and expression capacity. Highly correlated expression of the examined promoters suggests that this assumption is applicable and acceptable.

S2.2 Extracting Pathway and Expression Capacity Information from the Data

Variance and Covariance

We define the variance of a given quantity *x* for a population of cells as

$$\sigma^2 = \frac{1}{N}\sum(x_i - \bar{x})^2 \quad (2)$$

$$= \overline{(x_\iota - \bar{x})} \quad (3)$$

where *N* is the number of cells in the population. Since, we examined homologous cells of *C. elegans* nematodes, this number is equal to the number of animals in the population. We use the overbar symbol ($\bar{x}$) to represent the population average. Similarly we define the covariance, *Cov*(*x*,*y*), of two quantities *x* and *y* as

$$Cov(x,y) = \frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y}) \quad (4)$$

$$= \overline{(x_\iota - \bar{x})(y_\iota - \bar{y})} \quad (5)$$

The covariance will be non-zero if x and y are correlated (or anti-correlated).

The correlation coefficient, ρ(x,y), is the covariance scaled by the standard deviations. Specifically

$$\rho(x,y) = \frac{Cov(x,y)}{\sigma(x)\sigma(y)} \quad (6)$$

S2.3 Calculating Variance and Covariance of G, γ, L, and λ

As discussed above we assumed that in standard conditions, the average pathway output (*P*) for a given cell has a pathway power *L* associated with it, along with a stochastic term, *λ*. Thus, we

have $P=L+\lambda$. Similarly, the expression subsystem of the cell (E) contains a capacity term G and a stochastic term γ.

Thus, the amount of fluorescent reporter protein (y_i) produced for a given cell i is the product of the terms

$$y_i = (L_i + \lambda_i) \times (G_i + \gamma_i) \text{ x } \Delta\text{T} \quad (7)$$

$$= (L_i G_i + L_i \gamma_i + \lambda_i G_i + \lambda_i \gamma_i) \times \Delta\text{T} \quad (8)$$

Since the expectations of the stochastic fluctuation terms are zero, and since their fluctuations are uncorrelated to other terms, we expect that the population average of y_i reduces to the population average of the quantity $L_i G_i \Delta T$.

To calculate the average of y, we re-write the product L_i x G_i in terms of population averages and deviations from the average. For these two terms, we use a capital delta (Δ) to represent a deviation from a population average. We get

$$L_i = \bar{L} + \Delta L_i \quad (9)$$

$$G_i = \bar{G} + \Delta G_i \quad (10)$$

We then calculate the population average of y to be

$$\bar{y} = \frac{\Delta}{T} \sum (L_i \times G_i) \quad (11)$$

$$= (\bar{L} \times \bar{G} + Cov(L, G)) \Delta T \quad (12)$$

where we used equation (4) for the definition of covariance and the fact that $\Delta G = \Delta L = 0$. Thus the average number of GFP molecules is the product of L and G plus an extra term to account for their correlation.

To calculate the variance on *y*, we re-write equation (7) in terms of equations (9) and (10). We get

$$y_i = (\bar{L} + \Delta L_i + \lambda_i)\Delta T \times (\bar{G} + \Delta G_i + \gamma_i) = \tag{13}$$

$$= \left(1 \frac{\Delta L_i}{\bar{L}} + \frac{\lambda_i}{\bar{L}}\right) \times \left(1 + \frac{\Delta G_i}{\bar{G}} + \frac{\gamma_i}{\bar{G}}\right) \times \bar{G} \times \bar{L} \times \Delta T \tag{14}$$

$$= \left(1 + \frac{\Delta L_i}{\bar{L}} + \frac{\lambda_i}{\bar{L}}\right) \times \left(1 + \frac{\Delta G_i}{\bar{G}} + \frac{\gamma_i}{\bar{G}}\right) \times \bar{y}(1 - \frac{Cov(L,G)\Delta T}{\bar{y}}) \tag{15}$$

$$\approx y \times (1 + \frac{\Delta L_i}{\bar{L}} + \frac{\lambda_i}{\bar{L}} + \frac{\Delta G_i}{\bar{G}} + \frac{\gamma_i}{\bar{G}}) \tag{16}$$

where we used equation (12) to make the substitution for $\bar{L} \times \bar{G} \times \Delta T$ in going from equation (14) to (15); and where we have dropped, in going from equation (15) to (16) all higher order terms of fractional deviations from the means. These higher order terms are various products of $\frac{\Delta L_i}{\bar{L}}$, $\frac{\lambda_i}{\bar{L}}$, $\frac{\Delta G_i}{\bar{G}}$, and $\frac{\gamma_i}{\bar{G}}$. We assume that each of the higher order terms is small relative to the lower order terms that are retained in equation (16), and therefore we have neglected them. The term $\frac{Cov(L,G)\Delta T}{\bar{y}}$ is second order, and has also been neglected. Discussion of the magnitude of the error introduced by this approximation is provided in Coleman-Lerner et al.

We use equation (16) to calculate the variance on the number of GFP molecules using the definitions of variance and covariance above. We get

$$\frac{\sigma^2(y)}{\bar{y}^2} = \frac{\sigma^2(L)}{\bar{L}^2} + \frac{\sigma^2(\lambda)}{\bar{L}^2} + \frac{\sigma^2(G)}{\bar{G}^2} + \frac{\sigma^2(\gamma)}{\bar{G}^2} + 2\rho(L,G)\frac{\sigma(L)}{\bar{L}}\frac{\sigma(G)}{\bar{G}} \tag{17}$$

where *ρ(L,*G*)* is the correlation coefficient between *L* and G.

The correlation coefficient is always between -1 and 1, and a value of 0 corresponds to no correlation. There may be no correlation between *L* and G, but, for example, one possibility that would lead to a correlation is that cells with higher expression capacity may have stronger or weaker pathway output, perhaps because they have higher amounts of positive regulators or negative regulators of the pathway, respectively. This would show up as a positive or negative value for ρ(*L*,G), respectively.

We do not expect that the stochastic fluctuations λ or γ are correlated to any other variables, and thus the terms *ρ(L,λ), ρ(L,γ), ρ(γ,*G*), ρ(λ,*G*), and ρ(λ,γ)* were omitted from equation (17).

All the variances in equation (17) are expressed as fractions of the means squared. We define the variables

$$\eta(L) \equiv \sigma(L)/\bar{L} \qquad (18)$$

$$\eta(\lambda) \equiv \sigma(\lambda)/\bar{L} \qquad (19)$$

$$\eta(G) \equiv \sigma(G)/\bar{G} \qquad (20)$$

$$\eta(\gamma) \equiv \sigma(\gamma)/\bar{G} \qquad (21)$$

The quantity *η* is simply the width of different distributions expressed as a fraction of their means. For the stochastic variables *λ* and *γ*, we will refer to *η(λ)* and *η(γ)* as the “noise” associated with those quantities, and for the non-stochastic variables *L* and G, we will refer to *η(L)* and *η(*G*)* as the “variation.”

We quantified cell-to-cell variation in system output using "normalized variance" ($\eta^2 = \sigma^2/\mu^2$) rather than “noise strength” (σ^2/μ), a measure others have used to describe deviations from purely stochastic Poisson-type biological processes. We used normalized variance for a number of reasons. First, we found that most of the cell-to-cell differences in system behavior we reported are not due to stochastic differences in signal transmission or gene expression, as described in

the main text. Second, use of η^2 allowed examination of different amounts of variation in terms of the fraction of the mean. Third, and most important, because η^2 is unitless, it allowed direct comparison of different measurements, for example, from different fluorescent proteins, and it allowed the definition of total variation as the sum of individual sources of variation plus additional terms to account for correlations.

We re-write equation (17) in terms of η as

$$\eta^2(y) = \eta^2(P) + \eta^2(G) + \eta^2(\gamma) + 2\rho(L,G)\eta(L)\eta(G) \tag{22}$$

We have written $\eta^2(P)$ for $\eta^2(L)+\eta^2(\lambda)$ since the data will not be able to distinguish the cell-to-cell variation in pathway output from the noise or stochastic fluctuations. The term $\eta^2(P)$ is the variation in average pathway output per unit time, which, because pathway output is given by $P\Delta T$ and ΔT is the same for every cell, is identical to variation in pathway output.

To separate $\eta(G)$ from $\eta(P)$ and $\eta(\gamma)$ in the data from the two-promoter, two-color experiments, discussed further below, we introduce the quantity

$$Z(GFP, mCherry) = \frac{\eta^2(GFP)+\eta^2(mCherry)}{2} - \eta(GFP)\eta(mCherry)\rho(GFP, mCherry)$$

$$= \frac{\eta^2(GFP)+\eta^2(mCherry)}{2} - \frac{Cov(GFP, mCherry)}{\overline{GFP} \times \overline{mCherry}} \tag{23}$$

The quantity Z is the average variance divided by the mean square of the two fluorescence signals GFP and mCherry with the correlated part subtracted out. Thus, Z is a measure of the

uncorrelated part of the GFP vs mCherry scatter plot. Depending of the type of the experiment (two identical promoters or two different promoters) Z encompasses distinct sources of variation.

S2.4 Two Color Variants Driven by the Same Promoter and Gene Expression Noise

We calculate the correlation coefficient between GFP and mCherry reporter proteins for the case that both genes are expressed in the same cell and with the same promoter as

$$\rho(GFP, mCherry) \equiv \frac{1}{\sigma(GFP)\sigma(mCherry)} \frac{1}{N} \sum_i \Delta(GFP_i)\Delta(mCherry_i)$$

(24)

$$= \frac{1}{\eta(GFP)\eta(mCherry)} \frac{1}{N} \sum_i \left(\frac{\Delta L_i}{\bar{L}} + \frac{\lambda_i}{\bar{L}} + \frac{\Delta G_i}{\bar{G}} + \frac{\gamma_{GFP,i}}{\bar{G}}\right)\left(\frac{\Delta L_i}{\bar{L}} + \frac{\lambda_i}{\bar{L}} + \frac{\Delta G_i}{\bar{G}} + \frac{\gamma_{mCherry,i}}{\bar{G}}\right)$$

(25)

where we have used equation (16) for the deviations from the mean, $\Delta(GFP_i)$, and an analogous equation for $\Delta(mCherry_i)$. Since the GFP gene has the same promoter as mCherry, ΔL_i is the same for both color variants. The quantity λ_i is the same since stochastic fluctuations in pathway output occur upstream of the promoters (with the exception of the binding of the transcription factors to each individual copy of the reporter genes, see the end of this section), and the quantity ΔG_i is the same since the proteins are being expressed in the same cell. The stochastic fluctuations in gene expression, γ_i, however, are different for the two color variants and we write them as $\gamma_{GFP,i}$ and $\gamma_{mCherry,i}$ for GFP and mCherry respectively.

The uncorrelated terms drop out of equation (25) when we perform the population average. We do not expect the stochastic fluctuations to be correlated with any other terms, and equation (25) becomes

$$\rho(GFP, mCherry) = \frac{\eta^2(P)+\eta^2(G)+2\rho(L,G)\eta(L)\eta(G)}{\eta(GFP)\,\eta(mCherry)} \qquad (26)$$

Where $\eta^2(P) = \eta^2(L) + \eta^2(\lambda)$ as discussed above.

The quantity *Z*(*GFP*, m*Cherry*) as defined in equation (23), gives the contribution of the uncorrelated part of the expression of the two color variants to the average η^2 of the two colors. Only the gene expression noise is uncorrelated, and we get

$$Z(GFP, mCherry) = \frac{\eta^2(\gamma_{GFP})+\eta^2(\gamma_{mCherry})}{2} \qquad (27)$$

Z(*GFP*, m*Cherry*) is the average of gene expression noise of the two promoters. Since the same promoters are driving the GFP and mCherry genes, we expect equal levels of mRNA for the two variants on average, and therefore the gene expression noise is the same for the GFP and mCherry signal. We get $\eta^2(\gamma_{GFP}) = \eta^2(\gamma_{mCherry}) = \eta^2(\gamma)$ and therefore, from equation (27), $Z(GFP, mCherry) = \eta^2(\gamma)$. Thus, equation (23) gives direct measurement of gene expression noise γ when applied to an experiment with two identical promoters:

$$\eta^2(\gamma) = \frac{\eta^2(GFP)+\eta^2(mCherry)}{2} - \frac{Cov(GFP,mCherry)}{\overline{GFP} \times \overline{mCherry}} \qquad (28)$$

<u>We used this equation (28) for practical calculations of gene expression noise of individual promoters.</u>

We note that for the experiment described above, GFP and mCherry were not driven by the same promoter, but rather by identical *copies* of the same promoter. Therefore, stochasticity in the binding of transcription factors to the DNA and in the subsequent activation of transcription contributes to the uncorrelated part of the total variation. In our *ad hoc* model, these molecular

steps are part of the pathway subsystem, and therefore fluctuations in this step should contribute to the transmission noise, $\eta^2(\lambda)$. However, due to experimental limitations of the type of experiment we have just described, the noise caused by these molecular steps is, instead, included in the measure of gene expression noise $\eta^2(\gamma)$.

S2.5 Two Color Variants Driven by Different Promoters and Pathway Variation

For this case, the calculation of *ρ(GFP,mCherry)* is the same as in equation (25) above except that the pathway activity is different for the two color variants since different promoters are driving the GFP and mCherry genes. If we assume that the two terms *L* and *λ* are uncorrelated for the two promoters then we get

$$\rho(GFP, mCherry) = \frac{\eta^2(G)+\rho(L_{GFP},G)\eta(L_{GFP})\eta(G)+\rho(L_{mCherry},G)\eta(L_{mCherry})\eta(G)}{\eta(GFP)\,\eta(mCherry)} \tag{29}$$

where we have written L_{GFP} and $L_{mCherry}$ for capacities of the pathways that lead to the activation of the GFP and mCherry gene respectively. The uncorrelated part of the average η^2 will now include a contribution from the pathway variation; and the quantity *Z*(*gfp*, m*Cherry*) includes this extra contribution. We get

$$Z(GFP, mCherry) = \frac{\eta^2(P_{GFP})+\eta^2(P_{mCherry})}{2} + \frac{\eta^2(\gamma_{GFP})+\eta^2(\gamma_{mCherry})}{2} \tag{30}$$

where $\eta^2(\lambda_{GFP})$ is the transmission noise for the pathway that leads to GFP activation and $\eta^2(P_{GFP}) = \eta^2(L_{GFP}) + \eta^2(\lambda_{GFP})$, and similarly for $\eta^2(\lambda_{mCherry})$ and $\eta^2(P_{mCherry})$.

Thus, when uncorrelated expression variation for two independent promoters includes an average variation in pathways activation and gene expression variations for each promoter. The latter can be calculated from type I experiments. Knowing expression variations $\eta^2(\gamma)$ for each promoter and uncorrelated variation *Z* from equation (23) allows to calculate of pathway variations $\eta^2(P)$. To split contribution of variation of each pathway we need to perform three experiments with pairwise analysis of three promoters and then solve a system linear equations to calculate $\eta^2(P)$ of each pathway.

Below we give an example with expression of three promoters, *vit-2*, eft-3, daf-21, in cells Int3V and Int3D.

$$Z\left(P_{eft-3} :: BFP, P_{vit-2} :: mCherry\right) = \frac{\eta^2(P_{eft-3})+\eta^2(P_{vit-2})}{2} + \frac{\eta^2\left(\gamma_{P_{eft-3}::BFP}\right)+\eta^2\left(\gamma_{P_{vit-2}::mCherry}\right)}{2} \quad (31)$$

$$Z\left(P_{eft-3} :: BFP, P_{daf-21} :: mCherry\right) = \frac{\eta^2(P_{eft-3})+\eta^2(P_{daf-21})}{2} + \frac{\eta^2\left(\gamma_{P_{eft-3}::BFP}\right)+\eta^2\left(\gamma_{P_{daf-21}::mCherry}\right)}{2} \quad (32)$$

$$Z\left(P_{daf-21} :: GFP, P_{vit-2} :: mCherry\right) = \frac{\eta^2(P_{daf-21})+\eta^2(P_{vit-2})}{2} + \frac{\eta^2\left(\gamma_{P_{daf-21}::GFP}\right)+\eta^2\left(\gamma_{P_{vit-2}::mCherry}\right)}{2} \quad (33)$$

Average gene expression noise of two promoters, e.g. $\frac{\eta^2\left(\gamma_{P_{eft-3}::BFP}\right)+\eta^2\left(\gamma_{P_{vit-2}::mCherry}\right)}{2}$ is calculated from type I experiments and equation (28). It is therefore a known value for each cell type and promoter. We calculate quantity *Z* directly from equation (23). Quantity Z and average expression noise therefore give a particular number of each of the equations above. We can rewrite them in more suitable form to calculate pathways variations

$$\frac{\eta^2(P_{eft-3})+\eta^2(P_{vit-2})}{2} = A_1$$

(34)

$$\frac{\eta^2(P_{eft-3})+\eta^2(P_{daf-21})}{2} = A_2$$

(35)

$$\frac{\eta^2(P_{daf-21})+\eta^2(P_{vit-2})}{2} = A_3$$

(36)

Where A_1, A_2 and A_3 are known values. To calculate pathway variation for a particular promoter, e.g. eft-3, we will sum up (34) and (35) and subtract (36):

$$\frac{\eta^2(P_{eft-3})+\eta^2(P_{vit-2})}{2} + \frac{\eta^2(P_{eft-3})+\eta^2(P_{daf-21})}{2} - \frac{\eta^2(P_{daf-21})+\eta^2(P_{vit-2})}{2} = A_1 + A_2 - A_3$$

(37)

$$\frac{\eta^2(P_{eft-3})+\eta^2(P_{vit-2})+\eta^2(P_{eft-3})+\eta^2(P_{daf-21})-\eta^2(P_{daf-21})-(P_{vit-2})}{2} = A_1 + A_2 - A_3$$

(38)

$$\frac{2\eta^2(P_{eft-3})}{2} = A_1 + A_2 - A_3 \qquad (39)$$

$$\eta^2(P_{eft-3}) = A_1 + A_2 - A_3 \qquad (40)$$

Thus, we have got one estimate of $\eta^2(P_{eft-3})$. To obtain it we used results of the experiments involving P_{eft-3}, P_{vit-2} and P_{daf-21} reporters in pairwise manner. We say that these 3 pairs of experiments form a 'triangle'. However, we could choose another triangle consisting of experiments with P_{eft-3}, P_{vit-2} and $P_{hsp-16.2}$ reporters, or a triangle of P_{eft-3}, $P_{hsp-16.2}$ and P_{daf-21} reporters. To get a better estimate of the true $\eta^2(P_{eft-3})$ value we have calculated it from all three possible triangles and averaged. We did the same for other reporters too. The resulting $\eta^2(P)$ values are shown in Figure 3 bar graphs.

Similarly, we can calculate pathway variation for other promoters for which we have measured their expression noise. In case when we did not measure expression noise, e.g. hsp-17 promoter, we did calculate pathway variation for these promoters as well. In these cases we have limited our analysis with comparison of total correlated variation of two promoters to their uncorrelated variation. In all cases we have observed that correlated variation dominates over uncorrelated term:

$$\frac{Cov(P_{hsp-17}::GFP, P_{eft-3}::mNeptune)}{\overline{P_{hsp-17}::GFP} \times P_{eft-3}::mNeptune} \gg Z(P_{hsp-17} :: GFP, P_{eft-3} :: mNeptune)$$

References:

1 Kirkwood, T. B. & Finch, C. E. *Chance, Development, and Aging.*, (Oxford University Press, 2000).

2 Timofeeff-Ressovsky, N. W. U¨ ber den Einfluss des Genotypus auf das phanotypen Auftreten eines einzelnes. *Gens. J. Psychol. Neurol.* **31**, 305-310 (1925).

3 Romaschoff, D. D. Uber die Variabilita¨t in der Manifestierung eines erblichen Merkmales (Abdomen abnormalis) bei Drosophila funebris. *F. J. Psychol. Neurol.* **31**, 323-325 (1925).

4 Laubichler, M. D. & Sarkar, S. in *Mutaliflg Concepts. Evolving Disciplines: Genetics, Medicine and Society* (eds L.S. Parker & R.A. Ankeny) Ch. 4, 63-85 (Kluwer Academic Publishers, 2002).

5 Mortimer, R. K. & Johnston, J. R. Life span of individual yeast cells. *Nature* **183**, 1751-1752 (1959).

6 Kirkwood, T. B. & Finch, C. E. Ageing: the old worm turns more slowly. *Nature* **419**, 794-795 (2002).

7 Rea, S. L., Wu, D., Cypser, J. R., Vaupel, J. W. & Johnson, T. E. A stress-sensitive reporter predicts longevity in isogenic populations of Caenorhabditis elegans. *Nature genetics* **37**, 894-898 (2005).

8 Mendenhall, A. R. *et al.* Expression of a single-copy hsp-16.2 reporter predicts life span. *The journals of gerontology* **67**, 726-733 (2012).

9 Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. *Science* **335**, 82-85 (2011).

10 Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699-706 (2005).

11 Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-547 (2008).

12 Seewald, A. K., Cypser, J., Mendenhall, A. & Johnson, T. Quantifying phenotypic variation in isogenic Caenorhabditis elegans expressing Phsp-16.2::gfp by clustering 2D expression patterns. *PloS one* **5**, e11426 (2010).

13 Cypser, J. R. *et al.* Predicting longevity in C. elegans: fertility, mobility and gene expression. *Mechanisms of ageing and development* **134**, 291-297, doi:10.1016/j.mad.2013.02.003 (2013).

14 Mendenhall, A. R., Tedesco, P. M., Sands, B., Johnson, T. E. & Brent, R. Single Cell Quantification of Reporter Gene Expression in Live Adult Caenorhabditis elegans Reveals Reproducible Cell-Specific Expression Patterns and Underlying Biological Variation. *PloS one* **10**, e0124289, doi:10.1371/journal.pone.0124289 (2015).

15 Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. *Cell* **115**, 489-502 (2003).

16 Sanchez-Blanco, A. & Kim, S. K. Variable pathogenicity determines individual lifespan in Caenorhabditis elegans. *PLoS genetics* **7**, e1002047 (2011).

17 Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186, doi:10.1126/science.1070919 (2002).

18 Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* **99**, 12795-12800, doi:10.1073/pnas.162041399 (2002).

19 Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814 (2004).

20 Han, M. & Sternberg, P. W. let-60, a gene that specifies cell fates during C. elegans vulval induction, encodes a ras protein. *Cell* **63**, 921-931 (1990).

21 Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011-1014 (2006).

955 22 Prahlad, V. & Morimoto, R. I. Neuronal circuitry regulates the response of Caenorhabditis
956 elegans to misfolded proteins. *Proc Natl Acad Sci U S A* **108**, 14204-14209,
957 doi:10.1073/pnas.1106557108 (2011).
958 23 Prahlad, V., Cornelius, T. & Morimoto, R. I. Regulation of the cellular heat shock response in
959 Caenorhabditis elegans by thermosensory neurons. *Science* **320**, 811-814 (2008).
960 24 Mendenhall, A., Crane, M. M., Tedesco, P. M., Johnson, T. E. & Brent, R. Caenorhabditis elegans
961 Genes Affecting Interindividual Variation in Life-span Biomarker Gene Expression. *The journals*
962 *of gerontology*, doi:10.1093/gerona/glw349 (2017).
963 25 Walker, G. A. & Lithgow, G. J. Lifespan extension in C. elegans by a molecular chaperone
964 dependent upon insulin-like signals. *Aging cell* **2**, 131-139 (2003).
965 26 Baird, N. A. *et al.* HSF-1-mediated cytoskeletal integrity determines thermotolerance and life
966 span. *Science* **346**, 360-363, doi:10.1126/science.1253168 (2014).
967 27 Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic
968 variation in genetic interaction partners. *Nature* **480**, 250-253 (2011).
969 28 Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56-67,
970 doi:10.1016/j.cell.2012.08.026 (2012).
971 29 Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic
972 stem cells. *Cell* **151**, 68-79, doi:10.1016/j.cell.2012.08.033 (2012).
973 30 Pickett, C. L., Breen, K. T. & Ayer, D. E. A C. elegans Myc-like network cooperates with
974 semaphorin and Wnt signaling pathways to control cell migration. *Developmental biology* **310**,
975 226-239, doi:10.1016/j.ydbio.2007.07.034 (2007).
976 31 Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies
977 incomplete penetrance. *Nature* **463**, 913-918 (2010).
978 32 Kirstein-Miles, J., Scior, A., Deuerling, E. & Morimoto, R. I. The nascent polypeptide-associated
979 complex is a key regulator of proteostasis. *The EMBO journal* **32**, 1451-1468,
980 doi:10.1038/emboj.2013.87 (2013).
981 33 Dhondt, I. *et al.* FOXO/DAF-16 Activation Slows Down Turnover of the Majority of Proteins in C.
982 elegans. *Cell reports* **16**, 3028-3040, doi:10.1016/j.celrep.2016.07.088 (2016).
983 34 Depuydt, G., Shanmugam, N., Rasulova, M., Dhondt, I. & Braeckman, B. P. Increased Protein
984 Stability and Decreased Protein Turnover in the Caenorhabditis elegans Ins/IGF-1 daf-2 Mutant.
985 *The journals of gerontology* **71**, 1553-1559, doi:10.1093/gerona/glv221 (2016).
986 35 Visscher, M. *et al.* Proteome-wide Changes in Protein Turnover Rates in C. elegans Models of
987 Longevity and Age-Related Disease. *Cell reports* **16**, 3041-3051,
988 doi:10.1016/j.celrep.2016.08.025 (2016).
989 36 Schmalhausen, I. *Factors of evolution; the theory of stabilizing selection*. (Blakiston, 1949).
990 37 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and
991 complex traits. *Nat Rev Genet* **6**, 95-108, doi:10.1038/nrg1521 (2005).
992 38 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753,
993 doi:10.1038/nature08494 (2009).
994 39 Baryshnikova, A., Costanzo, M., Myers, C. L., Andrews, B. & Boone, C. Genetic interaction
995 networks: toward an understanding of heritability. *Annu Rev Genomics Hum Genet* **14**, 111-133,
996 doi:10.1146/annurev-genom-082509-141730 (2013).
997 40 Mendenhall, A., Driscoll, M. & Brent, R. Using measures of single-cell physiology and
998 physiological state to understand organismic aging. *Aging cell* **15**, 4-13, doi:10.1111/acel.12424
999 (2016).
1000 41 Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat Rev*
1001 *Genet* **14**, 168-178, doi:10.1038/nrg3404 (2013).

1002 42 Frokjaer-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated
1003 transgenesis in C. elegans. *Nature methods* **9**, 117-118 (2012).
1004 43 Sands, B. *et al.* A toolkit for DNA assembly, genome engineering and multicolor imaging for C.
1005 elegans. *Translational Medicine of Aging* **X**, x-xx In Press, doi:In Press (2018).
1006 44 Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).

1007