

Supplementary File

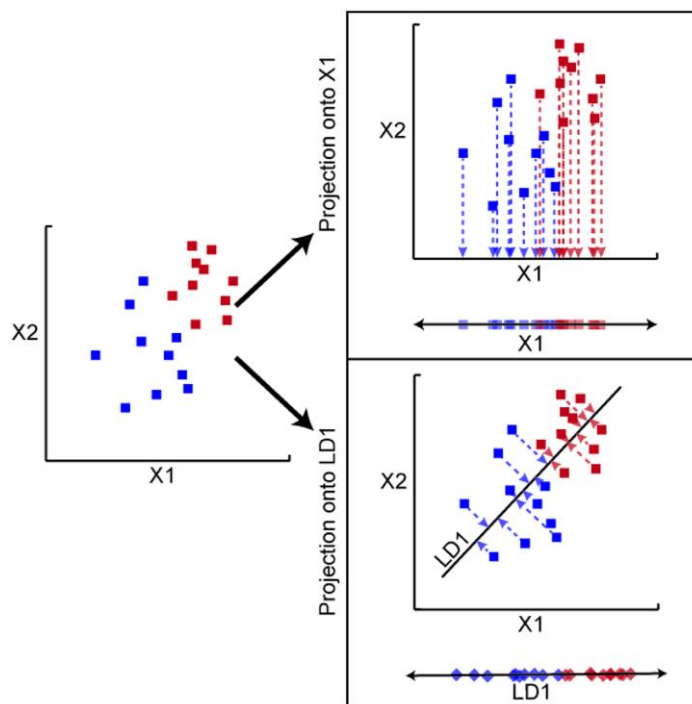
Generation of a composite CeleST score

Locomotion is an important health measure for animals ranging from flies to humans [1–3]. For *Caenorhabditis elegans*, plate-based locomotion measures, typical mean and maximum velocity, are commonly used health measures [4, 5]. The *Caenorhabditis* Intervention Testing Program (CITP) [6] adopted the CeleST platform [7–9], which acquires eight measures of swimming ability, as a more informational deep locomotion assay (see Supplementary Figure 4). While the eight measures provide a broad range of information on the swimming of each of the strains, it was not clear which measures best capture the decline of locomotion with age across our genetic diversity panel [10]. We therefore combined the information from all measures for each strain into a single multivariate composite measure by developing an analysis program pipeline that creates a single composite swimming score from eight original swim test measurements.

The data used to generate the composite score were generated by processing 30 second videos with the published CeleST software [7, 8]. The software provides

eight measurements of mobility, like travel speed, asymmetry, curling, and body waves initiation, for individual worms. We collected the data at three days in adulthood (5, 9, and 12). To generate a composite score, we first analyzed the eight original measurements in each record using a linear discriminant analysis (LDA) approach using the eight original measurements in each record as the predictor variables and the age-dependent decline as the primary discriminator. By projecting the eight predictor variables onto a single axis, LDA creates the first linear discriminant function which maximizes the differences between ages across all measurements, while reducing the dimensionality from eight to one (see Visual Example).

The first linear discriminant function minimizes within-group (age in our case) variance, maximizes between-group (age) variance, and maximizes the separability of the means of the groups (ages), while having a coefficient for each of the original eight predictor variables [11]. While it cannot capture *all* the information provided by the original measurements, it does maximize informational capture. Because this approach maximizes the differences between group levels, linear



Visual example of LDA dimensional reduction. In this example there are two predictor variables (X1 and X2) and two levels to the group (red and blue). All of the information captured is in the leftmost plot of variable X1 x X2. To reduce the number of dimensions down to one, we would project that plot onto a single axis. If projecting to one of the original axes, e.g., the X1 axis presented in the example, all information from the second variable is lost and little benefit in distinguishing between levels is created. By creating a new axis to project onto, LD1 in the example shown, we preserve nearly all of the information from both variables while still reducing the dimension to one, and that new first linear discriminate function has a coefficient for each of the original predictor variables.

discriminant analysis is often used to predict group membership after a training dataset is used. Generating the coefficients for each of the original eight predictor variables in a strain specific manner enabled us to avoid arbitrarily determining how any of the eight measures vary between ages, or if they matter at all in differentiating ages, while avoiding the assumptions that the strains swim and age the same.

To implement the LDA approach for analyzing compound interventions, LDA is first performed on the control dataset, independently for each strain, to retrieve the coefficients from the first linear discriminant function. The strain-specific coefficients are then used as weighted loadings which are multiplied by the corresponding measurement and divided by the control dataset's measurement standard deviation. This is done for every worm record in the strain dataset, including compound and control records. Those eight weighted measurements are summed up within each record to create the composite swimming score. This composite score can then be used to compare compound interventions and ages within each strain without arbitrarily choosing measurements for comparison. For statistical comparison, we used models with compound treatment and age as an interaction for input into Type-III Analysis of Variance tests for the effect of a factor, given that other factors are present. A significant age by compound interaction is of the most interest, as it implies that the impact size of a compound varies by age. The scripts used for generating the composite

scores and analysis, and additional relevant information, have been made available online [12].

CeleST composite score example

Because each record has eight measurements and belongs to a particular strain, compound, and age, we were presented with question: what do these measures mean in terms of health? Our solution is to combine the information provided by these measurements while considering (i) the relevance of each measure, (ii) the different units and scales, and (iii) how not all measures are independent. By combining the information into a single value using the LDA approach described above, we avoided arbitrary choices. Below we show an example for the processing of sample data.

The example below is from the N2 portion of the dataset included in this manuscript. In this case, Activity Index is the most important measure across the board when it comes to age and will have a greater impact on the final score. Curling and Travel Speed are nearly irrelevant and will not impact the score much. These loadings vary by strain and can be negative or positive. The loadings were negated from the initial output, so the direction of the score goes down over time for any particular strain-age-compound combination. LDA doesn't see each age as older/younger, just as distinct groups, so negating those loadings is the same as flipping the order of the age levels.

$$= \sum_{n=1}^8 \text{Measurement}_n \frac{\text{Loading}_n}{\text{DMSO SD}_n}$$

N2								
	Wave	Brush	Activity	Travel	Curling	Asymmetry	Stretch	Body
Loading	0.355	0.235	1.154	-0.002	-0.046	0.169	0.327	0.336
DMSO SD	23.912	0.077	78.208	2.172	3.321	0.023	0.096	0.557

Now when we look at the composite swimming score for each strain-age-compound combination, much of the relevant information from the eight original measurements is captured, and we can focus our analyses on that single new measurement. Similar to our work with survival/lifespan data [6], we create two mixed effects general linear models (GLM) for each strain:

CompositeSwimmingScore ~ Compound × Age + (1|Lab/Tech/Date/Video)

CompositeSwimmingScore ~ Compound_Age + (1|Lab/Tech/Date/Video)

The score is the response in the same way that death age/time is the response with lifespan data. The fixed effects term has an interaction, and the random effects are nested instead of crossed. This means we break the random effects levels into a tree – with an experiment date at one lab is unique from the same date at a different lab.

To look at the significance of an interaction between compound and age, a Type-III ANOVA is performed on the first GLM. This type of ANOVA tests for the effect of a factor, given that the other factors are present. This approach is best when a significant interaction effect is present, as the effect of that interaction is tested after

considering the individual factors. The presence of a significant interaction effects suggests the score is telling us something about healthspan, not just lifespan. A single test is run for each strain. Any breakdown of the sources of variance uses a similar GLM with the interaction term.

To compare between compound and ages, the second GLM is used. This GLM is similar to the first, but instead of having an interaction term, the response variable combines compound and age. That way the control and the compound can be compared at particular ages. These pairwise comparisons are for the differences between the means of the Composite Swimming Score. All analyses are done within a particular strain.

SUPPLEMENTARY REFERENCES

1. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, Fox M, Guralnik JM. Physical performance measures in the clinical setting. *J Am Geriatr Soc.* 2003; 51:314–22.
<https://doi.org/10.1046/j.1532-5415.2003.51104.x>
PMID:12588574
2. Volpato S, Cavalieri M, Sioulis F, Guerra G, Maraldi C, Zuliani G, Fellin R, Guralnik JM. Predictive value of the Short Physical Performance Battery following hospitalization in older patients. *J Gerontol A Biol Sci Med Sci.* 2011; 66:89–96.
<https://doi.org/10.1093/gerona/glq167>
PMID:20861145
3. Jones MA, Grotewiel M. *Drosophila* as a model for age-related impairment in locomotor and other behaviors. *Exp Gerontol.* 2011; 46:320–5.
<https://doi.org/10.1016/j.exger.2010.08.012>
PMID:20800672
4. Glenn CF, Chow DK, David L, Cooke CA, Gami MS, Iser WB, Hanselman KB, Goldberg IG, Wolkow CA. Behavioral deficits during early stages of aging in *Caenorhabditis elegans* result from locomotory deficits possibly linked to muscle frailty. *J Gerontol A Biol Sci Med Sci.* 2004; 59:1251–60.
<https://doi.org/10.1093/gerona/59.12.1251>
PMID:15699524
5. Hahm JH, Kim S, DiLoreto R, Shi C, Lee SJ, Murphy CT, Nam HG. *C. elegans* maximum velocity correlates with healthspan and is maintained in worms with an insulin receptor mutation. *Nat Commun.* 2015; 6:8919.
<https://doi.org/10.1038/ncomms9919>
PMID:26586186
6. Lucanic M, Plummer WT, Chen E, Harke J, Foulger AC, Onken B, Coleman-Hulbert AL, Dumas KJ, Guo S, Johnson E, Bhaumik D, Xue J, Crist AB, et al. Impact of genetic background and experimental reproducibility on identifying chemical compounds with robust longevity effects. *Nat Commun.* 2017; 8:14256.
<https://doi.org/10.1038/ncomms14256>
PMID:28220799
7. Restif C, Ibáñez-Ventoso C, Vora MM, Guo S, Metaxas D, Driscoll M. CeleST: computer vision software for quantitative analysis of *C. elegans* swim behavior reveals novel features of locomotion. *PLoS Comput Biol.* 2014; 10:e1003702.
<https://doi.org/10.1371/journal.pcbi.1003702>
PMID:25033081
8. Ibáñez-Ventoso C, Herrera C, Chen E, Motto D, Driscoll M. Automated Analysis of *C. elegans* Swim Behavior Using CeleST Software. *J Vis Exp.* 2016; 54359.
<https://doi.org/10.3791/54359>
PMID:28060275
9. Onken B, Sedore CA, Coleman-Hulbert AL, Hall D, Johnson E, Jones EG, Banse SA, Huynh P, Guo S, Xue J, Chen E, Harinath G, Foulger AC, et al. Metformin treatment of diverse *Caenorhabditis* species reveals the importance of genetic background in longevity and healthspan extension outcomes. *Aging Cell.* 2022; 21:e13488.
<https://doi.org/10.1111/acer.13488>
PMID:34837316
10. Teterina AA, Coleman-Hulbert AL, Banse SA, Willis JH, Perez VI, Lithgow GJ, Driscoll M, Phillips PC. Genetic diversity estimates for the *Caenorhabditis* Intervention Testing Program screening panel. *MicroPubl Biol.* 2022; 2022.
<https://doi.org/10.17912/micropub.biology.000518>
PMID:35098051
11. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 1936; 7:179–88.
<https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
12. *Caenorhabditis* Intervention Testing Program. *Caenorhabditis* Intervention Testing Program: CeleST Data and Composite Score. figshare. 2021.
<https://doi.org/10.6084/m9.figshare.c.5126579.v1>