**Transcriptome -based prediction of biological age of adult Drosophila**

**Authors**

**Jason Liang, Sidney Blimbaum, Jie Xu, Lei Zhou**

(Names and sequence to be finalized at acceptance).

Corresponding author: [leizhou@ufl.edu](mailto:leizhou@ufl.edu)

Developed for Genome Biology [Short report](https://genomebiology.biomedcentral.com/submission-guidelines/preparing-your-manuscript/short-report).

“ *main text should be around 1000-1500 words, excluding abstract, references, and figure legends, and should contain no headings.”*

**Abstract**

*(The Abstract should not exceed 100 words. Please minimize the use of abbreviations and do not cite references in the abstract. The abstract should be unstructured.)*

Estimating the biological age is of great interest to geriatric research. To develop a global age estimator for *Drosophila melanogaster*, we compiled a transcriptome (RNA-Seq) dataset from public databases. A model developed with elastic net was able to predict the age with an absolute mean error (AME) of about 2.5 days, roughly equivalent to 4 years in human lifespan. Interestingly, a neural network model was able to achieve very similar predictive power. This model-based biological age assessment is applicable for RNAs extracted from whole body, heads, or headless bodies of either sex. Genes identified by the machine learning approach implicated the expression level of innate immune genes as the most significant predictors for biological aging.

**Keywords (3-10)**

Inflammaging,elastic net**,** RNA-Seq**,** inflammatory age, Drosophila

**Background**

*(A short background section should be included of no more than 300 words.)*

Evaluation of the biological age is of significant interest and utility for aging research. In mammalian models, DNA methylation levels in selected CpG loci are considered as the most reliable estimators of biological age [1,2]. However, DNA methylation is absent in the model organisms such as *Drosophila melanogaster* and *C. elegans* [3].

The fruit fly model has been used extensively to study the basic mechanisms of aging and aging -related diseases [4–7]. It has been shown that compared to younger flies, genes involved in immune response have significantly elevated expression in aged flies [8–10]. Lacking adaptive immunity, innate immune system in *Drosophila* is responsible for clearing pathogens as well as tumor-like cells. Similar to mammalian models, age-associated inflammation in the fruit fly is also associated with increased neurodegeneration and shortened lifespan. However, it is not clear as to whether the expression levels of innate immune genes could provide an estimate of biological aging.

The popular lab strain of *Drosophila melanogaster*, w1118, has a mean lifespan of about 45 days in 25 OC [11]. This relatively short lifespan has made it feasible to monitor the whole aging process in experimental settings. However, it nonetheless is a laborious process prone to accidents and disruptions. Moreover, the availability of temporal control of gene expression systems, such as the GeneSwitch system[12,13], led to the demand of monitoring acceleration or deceleration of aging at a given time point. These needs motivated our search of a global estimator for aging.

**Results and Discussion** *(This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures. The discussion should be included in one section with the results.)*

**Compilation of RNA-Seq data sets with age information**

It is possible to obtain transcriptome information from any tissue or cell type from *Drosophila*. However, dissection or cell sorting procedures inevitably produce operation-induced changes in gene expression. Practically, RNA samples with minimal disruption can be reliably obtained from whole body, heads, or headless bodies of adult fruit fly. Heads can be separated from bodies using metal sheeves following snap freeze in liquid nitrogen or dry ice, thus produce minimal, if any, alteration in transcriptome during the separation process.

With the above consideration in mind, we searched for RNA-Seq datasets for these three types of tissues in NCBI/GEO. Samples with clear age information were collected. Wherever possible, age information was verified by checking the publication associated with the dataset. The raw reads files were then downloaded from SRA at NCBI. Transcripts per million (TPM) value for every genes was obtained using Salmon[14].

**Development and verification of a global biological age estimator.**

Flies raised in 25 C with standard food were considered as “normal” samples and were used for generating progression models using the elastic net approach. A total of 471 “normal” samples were presented in this group and we used 70:30 split for training and testing sets. Prior to elastic net regression, we normalized the data with The best elastic net model achieved mean absolute deviation (MAD) of 2.32 days. This is roughly equivalent of 3.7 yeas in human lifespan. This is considerably better than using

**Enrichment of Innate immune genes in the estimator.**

**Conclusions** *(This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study to the field.)*

We have developed an elastic net -based estimator for assessing the biological age of fruit flies based on transcriptomic measurements obtained via RNA-Seq. By focusing on tissues from which total RNA could be obtained reliably without dissection, our method can be applied to assess alteration of aging at any given time point of the adult stage.

* Inflammation plays an important role – recent mammalian transcriptome Drosophila Innate immunity.

**Methods (should be included after Conclusions)**

*The methods section should include:*

* *the aim, design and setting of the study*
* *the characteristics of participants or description of materials*
* *a clear description of all processes, interventions and comparisons. Generic names should generally be used. When proprietary brands are used in research, include the brand names in parentheses*
* *the type of statistical analysis used, including a power calculation if appropriate*
* *software tool requirements*

Datasets.

Obtaining transcript per million (TPM). Transcriptome quantification were obtained with Salmon[14] using the 6.43 Drosophila melanogaster release.

Machine learning.

1day ~1.6 year.

**Abbreviations**

MAD – mean absolute deviation.

Figure Legends.

Figure 1. A global Estimator for Drosophila Inflammatory Age.

1. Performance of Elastic Net mode based on Scale or MinMax. The prediction of the test set (30% of total) were shown.
2. Predicated age vs chronological age of all samples. Sex and tissue origin of the samples are indicated by colors and shapes, respectively.
3. The age-estimation model identified significant difference between flies fed on high fat diet for 7 days vs. the parallel controls fed on standard diet. Flies were 2-3 days old when subject to high fat diet (HFD) or normal diet (ND). Data from GSE123240 [15]. (\*\*, p<0.01; \*\*\*, p<0.001)
4. Application of the model to pro-longevity mutant
5. GO analysis of genes important for the progression model.

References:

1. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. Springer US; 2018;19:371–84.

2. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2015;16.

3. Engelhardt J, Scheer O, Stadler PF, Prohaska SJ. Evolution of DNA Methylation Across Ecdysozoa. J. Mol. Evol. Springer US; 2022;90:56–72.

4. Landis G, Shen J, Tower J. Gene expression changes in response to aging compared to heat stress, oxidative stress and ionizing radiation in Drosophila melanogaster. Aging (Albany. NY). 2012;4.

5. He Y, Jasper H. Studying aging in Drosophila. Methods. 2014;68.

6. Partridge L, Alic N, Bjedov I, Piper MDW. Ageing in Drosophila: the role of the insulin/Igf and TOR signalling network. Exp. Gerontol. Elsevier Inc.; 2011;46:376–81.

7. Piper MDW, Partridge L. Drosophila as a model for ageing. Biochim. Biophys. Acta - Mol. Basis Dis. Elsevier; 2017;0–1.

8. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, et al. Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 2004;101:7663–8.

9. Felix TM, Hughes KA, Stone EA, Drnevich JM, Leips J. Age-specific variation in immune response in Drosophila melanogaster has a genetic basis. Genetics. 2012;191:989–1002.

10. Arora S, Ligoxygakis P. Beyond Host Defense: Deregulation of Drosophila Immunity and Age-Dependent Neurodegeneration. Front. Immunol. 2020;11:1–13.

11. Oxenkrug GF. The extended life span of Drosophila melanogaster eye-color (white and vermilion) mutants with impaired formation of kynurenine. J Neural Transm. 2010;117:23–6.

12. Nicholson L, Singh GK, Osterwalder T, Roman GW, Davis RL, Keshishian H. Spatial and temporal control of gene expression in Drosophila using the inducible GeneSwitch GAL4 system. I. Screen for larval nervous system drivers. Genetics. 2008;178:215–34.

13. Ford D, Hoe N, Landis GN, Tozer K, Luu A, Bhole D, et al. Alteration of Drosophila life span using conditional, tissue-specific expression of transgenes triggered by doxycyline or RU486/Mifepristone. Exp. Gerontol. 2007;42:483–97.

14. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods. Nature Publishing Group; 2017;14:417–9.

15. Stobdan T, Azad P, Heinrichsen E, Sahoo D, Hartley I, Zhou D, et al. High fat diet induces gender-specific differential gene expression in Drosophila melanogaster brain | The FASEB Journal. Faseb. 2017;1–19.