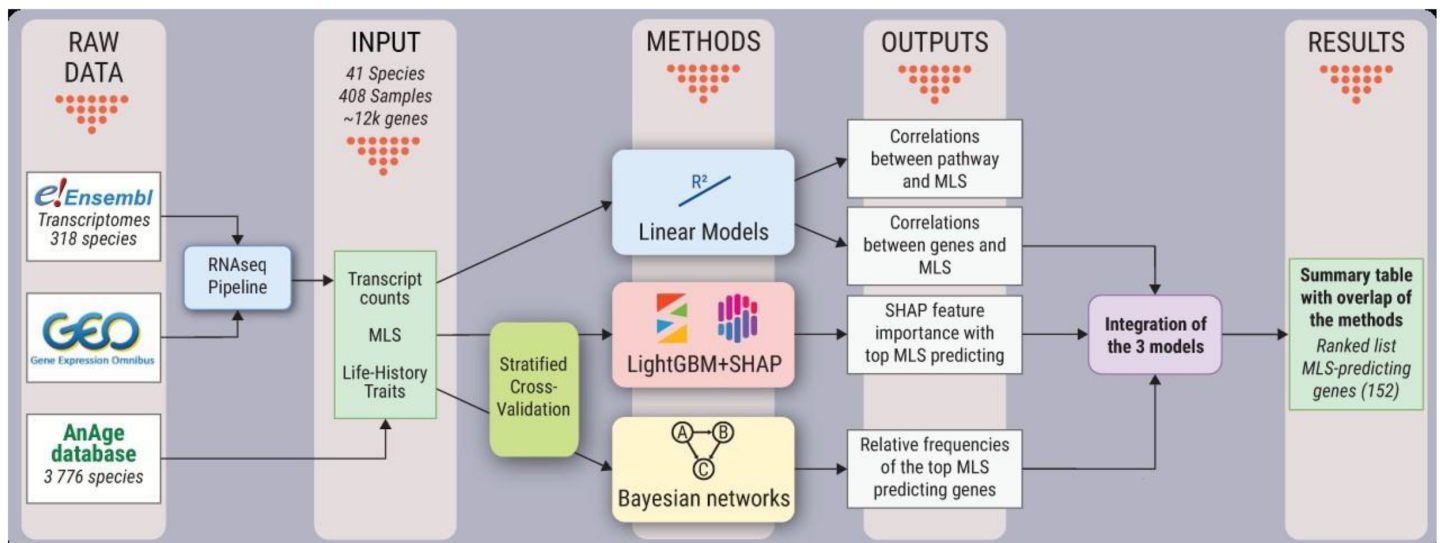


DVC is helping to reveal longevity-associated gene expression landscapes in mammals

DVC has been used in a machine learning pipeline that selects genes connected with maximum lifespan in mammals. Hundreds of genes are involved in the control of longevity in model organisms (mice, rats, etc). Still, until now, only a ~1.5-fold lifespan increase has been achieved through genetic interventions in mammals. In contrast, the maximum lifespan varies 100 times across mammals, where bowhead whale holds the record of living more than 211 years (268 years according to alternative estimates). It hints that the comparative biology of aging has not been exhausted yet and novel genetic interventions might still be discovered by looking at the differences between various species. Interventions that we can potentially use to slow down aging and extend our healthy lives.

In the study called “Open AccessArticle Machine Learning Analysis of Longevity-Associated Gene Expression Landscapes in Mammals” (<https://doi.org/10.3390/ijms22031073>) the team from Systems Biology of Aging Group (<http://aging-research.group>) analyzed gene expression from for five organs (heart, lung, liver, kidney, brain) using linear regression, LightGBM with SHapley Additive exPlanations (SHAP), and Bayesian networks



The machine learning pipeline of the study: (i) combined the results from linear organ-specific with nonlinear machine learning models based on decision trees, (ii) used SHAP explanatory models for gene signature prioritization, and (iii) applied a Bayesian networks-based algorithm to select genes which may have potential causality relationships with MLS.

In the ML pipeline, DVC was used to resolve the datasets and computation results from Google Drive remote and run stages. The ML pipeline's core stages are represented as Jupyter notebooks and can be either investigated manually in Jupyter lab interface or run programmatically as DVC stages with Papermill. All dependencies (incl. DVC) can be resolved from the conda environment.

The research paper authors are grateful to DVC founders Dmitry Petrov and Ivan Shcheklein for their advice on DVC usage and solving problems with Google Drive remotes.

You can read the article at <https://doi.org/10.3390/ijms22031073> (open access)

The code is available at <https://github.com/antonkulaga/yspecies>