Machine Learning identifies conserved traits that influence lifespan and healthspan responses to dietary restriction

Tyler A.U. Hilsabeck[1,2], Vikram P. Narayan[1], Kenneth A. Wilson[1], Enrique Carrera[1,3], Daniel Raftery[4], Daniel Promislow[5,6], Rachel B. Brem[1,2,7], Judith Campisi[1], Pankaj Kapahi[1,2,*]

Affiliations

[1]Buck Institute for Research on Aging, Novato, California, 94945, United States

[2]Davis School of Gerontology, University of Southern California, University Park, Los Angeles, California, 90089, United States

[3]Dominican University of California, San Rafael, California, 94901, United States

[4]Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, Washington, United States

[5]Department of Pathology, University of Washington, Seattle, WA 98195, United States

[6]Department of Biology, University of Washington, Seattle, WA 98195, United States

[7]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California, 94720, United States

***Correspondence**

pkapahi@buckinstitute.org

Summary

Dietary restriction (DR) is the most robust means to extend lifespan and healthspan across species, but factors such as genetic variation affect how an individual will respond to DR. Additionally, it is unclear how cumulative variations in metabolism and the metabolome influence longevity and health. We utilized metabolomic, phenotypic, and genome-wide data from *Drosophila* Genetic Reference Panel strains raised under *ad libitum* and DR conditions to identify factors which influence longevity and health in response to dietary modulation. We found multiple intra-dataset correlations (e.g., metabolites with metabolites) but few inter-dataset correlations (e.g., metabolites with health-related phenotypes). Through random forest modeling across all traits and Mendelian Randomization, we found key translatable traits that influence lifespan or healthspan determination and validated the role of multiple metabolites in regulating lifespan. Through these approaches, we utilized data from flies and humans to elucidate potential therapeutic pathways and metabolomic targets for diet response, lifespan, and healthspan.

Keywords

aging; dietary restriction; lifespan; metabolome; healthspan; Drosophila; GWAS; Mendelian Randomization; machine learning; random forest modeling

INTRODUCTION

Aging is the leading cause of morbidity and mortality. Dietary restriction (DR), the process of restricting nutrients without causing malnutrition, has been demonstrated as a robustly conserved means to extend lifespan and healthspan. Diet-responsive pathways such as the Target of Rapamycin and Insulin-like signaling have been implicated as driving factors of lifespan extension by DR; however, variations in response to diet between different individuals within the same species indicate that additional mechanisms influence diet response and remain yet to be elucidated[1]. Recently, genetic variation has been implicated as a significant driver of these phenotypic responses to DR. Studies in invertebrates[2] and rodents[3] have utilized systems biology approaches to identify additional DR-responsive factors that influence metabolic health and lifespan. Additionally, multi-omics approaches have identified novel metabolites which are regulated by genetic variation to influence longevity under DR[4]. This study noted the difficulty in correlating lifespan with metabolite traits due to the high correlation of the dietary restriction (DR) response in mean lifespan with metabolite traits of flies on the standard diet (AL). To avoid the confounding effects of an AL diet on the DR response, the study of metabolites and lifespan used the residuals from a simple regression[4]. Despite the findings from these studies, it remains unclear how different phenotypic interactions occurring within an individual influence the relationship between health and longevity.

A powerful approach to explore the relationship between health and longevity is to incorporate convergent data from multiple species. However, aging studies conducted in humans and model organisms such as fruit flies are often separate endeavors. Research conducted using model organisms has identified many candidate genes and pathways responsible for the aging process[5,6]. Often these genes and processes belong to evolutionarily conserved nutrient-sensing pathways which suggests they may be relevant in humans as well. However, several large GWAS of human longevity show a surprising lack of association with lifespan-extending genes identified in model organisms[7].

Recent advances in analytical approaches have allowed for more in-depth investigation of complex traits. Machine learning, which utilizes established datasets to generate more accurate predictive models for given phenotypes, has been useful for identifying new factors which influence a multitude of biological traits, including those involved in longevity and health. It remains to be observed, however, how metabolomic and phenotypic signatures can be utilized across species and genotypes in response to DR to predict specific

factors that contribute to multiple health and longevity traits. Previously, we have used the *Drosophila* Genetic Reference Panel (DGRP) for studies of how DR affects metabolic phenotypes[8], metabolome[4], and lifespan and healthspan[2]. Across these datasets, we found that no single metric was predictive of lifespan, suggesting that multiple factors must influence the longevity effects of DR. Here, we use these data to investigate the relationship between metabolomic, metabolic, and healthspan-related traits and lifespan. Through random forest modeling, we identify metabolites choline, threonine, and orotate to be relevant for building multiple lifespan models. The use of random forest modeling avoids the confounding effects of the AL traits driving DR response due to its bagging method approach. We further identify genetic factors which influence these relationships, including *Src64B* which associated with mean lifespan. To focus on traits that could be used as therapeutic targets in humans, we performed Mendelian Randomization using data from the UK Biobank to test the hypothesis that metabolites identified using machine learning, and their candidate genes influencing longevity in *Drosophila*, are also associated with human lifespan. Together, these results elucidate the complicated nature of DR and how it is influenced by genetic variation and specify factors that contribute to lifespan determination under DR (**Figure 1A**). Through this approach, we have identified pathways that could be used as valuable therapeutic targets for enhancing human lifespan and healthspan.

RESULTS

DGRP strains exhibit a wide variation in metabolite and phenotype responses to diet

We combined data from previously published DGRP metabolite and phenotype datasets, including Nelson et. al. 2016[8], Jin et. al. 2020[4], and Wilson et. al. 2020[2], with flies fed both *ad libitum* (AL, 5.0% yeast extract) and DR (0.5% yeast extract) dietary conditions. From these data, we separated ten primary traits generally used in lifespan and healthspan determination studies as ‘response’ variables for modeling and used the remaining traits as ‘predictors’ (**Supplementary Table 1**). Principal component analysis (PCA) of predictors, after removing strains with incomplete data across all traits, failed to cluster into groups by diet. To determine if trait values from these strains would allow the data to cluster by diet, we imputed averaged values across all strains for these missing values (**Supplementary Table 1&2**). Strains with missing values included four that lacked data for many phenotypes but had data for all metabolites, and 33 additional strains with all metabolite data but missing at least one phenotype value (**Supplementary Table 1&2**). PCA of this imputed list clustered by diet and were used for downstream analysis, similar to what was previously shown[4]. Diet was part of principal

component 1 (PC1), which explained ~23% of the data's variance. Approximately 32 principal components were sufficient to explain 85% of the variance in the data. Since diet explained a large portion of the variance, we determined the dietary response of each DGRP strain for each trait (value on AL subtracted from value on DR, "DR-AL"). In general, DGRP strains benefited from being on a DR diet, with over 93% of strains having decreased body weight and at least 78% having an increase in most lifespan response metrics (**Figure 1B-E** and **Supplementary Figure 1**). These data demonstrate the variation in dietary response across the DGRP, and the few strains that do not have decreased body weight or increased lifespan from DR can give insight into the mechanisms underlying these and other positive effects of dietary intervention.

DGRP **dietary response shows few metabolites correlate with lifespan- and healthspan-related phenotypes**

We looked for correlations between our dietary response predictor traits and found few correlations across data sets, with the majority of these being weak correlations (DR-AL, **Figure 2**). We performed hierarchical clustering using the Ward's minimum variance method to form trait clusters. The strongest correlations on either diet alone (DR or AL) or used in the same data set (DR&AL), were between metabolites in the same pathway or between similar phenotypes (**Supplementary Figure 2A-C**). This trend continued for correlations in dietary response (DR-AL), though the top three correlations were between similar phenotypes rather than between metabolites. Unsurprisingly, the same pairs were strongly correlated in all comparisons. The lack of strong correlations between any of the predictor traits and the response traits points to a need for another method for identifying factors that could predict and/or "build" these response traits.

Common traits used to build random forest models of lifespan response traits.

To determine the predictor traits that contribute to determining the lifespan and healthspan response traits, we built random forest models for each response using all predictor traits as inputs. Predictors were used to build 10,000 initial models per response trait, with a final model being made based on the initial 10,000. Predictors used to build the final models were given an importance score based on the proportion of initial trees/estimators for which they were included. Models were created for each diet alone, DR&AL together, and the DR-AL response, and the important predictor traits were plotted for each response trait (**Figure 3, Supplementary Table 3**). Models for lifespan metrics in any dietary condition tended to be built with climbing-

related traits, except the DR-AL response lifespan models which were also built by metabolite traits (**Figure 3A-F**). Specifically, threonine was used in at least 1% of the trees used for all 7 lifespan response models, peaking at 3% and 4% of the trees in the max lifespan and day 95% of flies dead models, respectively. Arginine was used in all lifespan models except for max lifespan, being in 4% of the late-life α trees, 5% of the late-life β, 4% of the day 95% of flies dead, and 3% of the variance in day of death model trees. Choline was in 1% of late-life β, variance, and max lifespan model trees. Orotate was in 2% and 5% of late-life α and late-life β trees, respectively. Metabolites quinolinate, methylhistidine, and glyceraldehyde were also used to build at least 1% of trees for one or multiple lifespan models. The more traditional physical response traits (glucose levels, triglyceride levels, and body weight), were built predominantly by metabolites and protein levels (**Figure 3G-I**). Particularly, the models for glucose levels that included the AL diet had malondialdehyde as the most important trait. Many of these traits have previously been associated with the model trait found here. The metabolites kynurenine, quinolinate, myristic acid, and threonine were each used to build at least one lifespan random forest model and have been previously implicated in lifespan. Threonine was used to build 8 of the 9 lifespan models and has been shown to increase lifespan in *C. elegans*[9–12]. Kynurenine and quinolinate have both been positively associated with age in humans, with quinolinate also being strongly associated with frailty and mortality[13]. Additionally, myristic acid is associated with maximum lifespan potential in mice[14]. The use of traits already shown to play a role in lifespan validates our modeling and suggests the importance of the other candidates it identified.

We visualized the models and their important traits using a combined kamada kawai and forceatlas force-based network organization, which grouped traits in similar clusters based on the push of common factors. These networks were further augmented by the addition of top GWAS candidate genes, based on p-value. As expected, the lifespan response traits were grouped together and had a few common predictors and genes, specifically latelifealpha and latelifebeta, and mean and median lifespans had significant overlap with each other (hypergeometric density p-values of 1.05E-09 for each overlap, **Figure 4, Supplementary Figure 4A, and Supplementary Table 3**). For example, myristic acid, glycerol phosphate, and the gene *pbl*, were included in the random forest models for mean and median lifespan or was a GWAS candidate, respectively. Similarly, climbing-related response traits clustered meaning they were used in the same random forest models, though there were more common GWAS candidates, particularly in the AL group (**Supplementary**

Figure 3A&4B). Visualizing the connections between these three datasets in this manner gives a better sense of how lifespan and healthspan response traits are built and the genetics that underlie them (**Supplementary Figure 3&4**).

Mendelian Randomization highlights translatable metabolites from Random Forest models.

We next used the networks developed above to identify multiple response models, with traits identified to be relevant to humans via Mendelian Randomization (MR) selected for validation in the fly. We explored the utility of the identified metabolite associations from the random forest model approach by applying them in two-sample MR using the IVW method. We identified 10 metabolites in *Drosophila* that were also detected in the large available GWAS for serum metabolites in humans and influence multiple lifespan and healthspan traits (**Figure 5A**). We found that 2 of the MR-prioritized metabolites, orotate and choline, were implicated in both healthspan and lifespan random forest models. Both metabolites had clear effects upon longevity, Body Mass Index (BMI) and Frailty Index (FI) ($P < 0.05/10 = 0.005$). 4 of the 10 metabolites (urate, inosine, histidine and threonine) were detected in the lifespan-only random forest models. Urate, inosine, and threonine had significant effects upon longevity ($P < 0.05/10 = 0.005$) with histidine having significant effects on BMI and FI ($P < 0.05/10 = 0.005$) but not longevity ($P > 0.05$). Amongst the 4 remaining metabolites (quinolinate, glucose, uridine, and kynurenine) detected in the healthspan random forest model, only glucose was significantly associated with longevity using the inverse-variance weighted method (βIVW, 0.005; 95% CI, 0.002–0.008; $P = 1.12 \times 10^{-3}$). The remaining metabolites exhibited non-significant associations with longevity but suggestive or significant FI-related effects. For quinolinate, glucose, and uridine no evidence was found for BMI-related effects ($P \geq 0.1$). These genetic associations and their effects can also be characterized by their pleiotropy which were observed in our MR analyses (**Supplementary Table 4**)

Overlapping metabolites in random forest lifespan models and Mendelian randomization have genotype- and sex-specific lifespan effects in *Drosophila melanogaster*.

To validate the metabolites that were both used to build random forest models of lifespan response and found to be relevant for humans via Mendelian Randomization, we supplemented our AL diet, with 3 concentrations of candidate metabolites. Validated metabolites were selected based on the number of lifespan models they were incorporated in, the average odds-ratio from Mendelian randomization, and the novelty of the lifespan

association (**Supplementary Table 5**). The impact of candidate metabolites choline, threonine, histidine, or orotate, on the survival of males and females from two fly strains, w1118 and Canton S, was tracked. Each metabolite showed a genotype-, sex-, and dose-specific effect on lifespan (**Figure 5B-D, Supplementary Figure 5**). Orotate supplementation significantly shortened lifespan at 10 mM in both strains and sexes, while choline and threonine extended or shortened lifespan in a strain-, sex-, and dose-specific manner (Fold changes compared to 0 mM controls and Cox proportional hazards p-values summarized in **Figure 5D, Supplementary Figure 5**). Together, these results validate our modeling approach and identify specific metabolites which are relevant for lifespan and have translatable effects in humans.

DISCUSSION

Genetic variability has been shown to greatly influence a strain's response to various conditions, including changes to diet[4,15,16]. With DR being one of the most robust ways to impact lifespan and other healthspan traits across species, understanding the genetics underlying beneficial responses would help identify the mechanisms by which it functions[17]. While studies have investigated similar connections[4,18–21], none have incorporated data from this many model organism genotypes and utilized Mendelian randomization of human data to identify human-relevant results, nor have they utilized our modeling and visualization approach to select and validate candidates. Here, we have demonstrated the variability in outcome responses on one of two diets and the differing responses of each genotype. We have shown that while trait levels may not correlate across datatypes (i.e. metabolites with phenotypes, and vice versa), they do play a role in different lifespan and healthspan outcomes. This lack of correlation across datasets could indicate non-linear and/or indirect relationships between traits that would not be found in standard correlation analysis. To handle non-linear relationships, we pursued a machine learning approach that could determine non-linear relationships and produce a value for each trait based on the proportion of trees/estimators that trait was included in. In this manner, we could identify relationships between our response and predictor traits that were not found to correlate. Through this unbiased, data-driven approach, we confirmed the importance of traits previously linked with their response variable.

Some examples of traits used to build our response models include the metabolite malondialdehyde, which has previously been linked to glucose and triglyceride levels[22–24]. We found it was important for modeling

glucose levels on AL and DR, separately, as well as the DR-AL response in glucose and triglyceride levels. Another metabolite previously associated with triglyceride levels, glyceraldehyde, was also found to be an important feature in our DR-specific model. Additionally, this metabolite was important for our initial mortality and rate of aging dietary response models, aligning with studies showing glyceraldehyde and the enzyme GAPDH as associated with lifespan[25,26]. The metabolite methylhistidine, previously linked to the RU486 lifespan effect, mortality, and frailty, was important in our initial mortality model on DR, and the dietary response in maximum lifespan[27]. A downstream metabolite in the well-known kynurenine/NAD+ pathway, quinolinate, which has previously been associated with aging, was an important feature in lifespan models from both diets, separately, and the dietary response[28,29]. Metabolites threonine, arginine, and choline were also included in our lifespan and healthspan models and had previously been reported to influence lifespan[9,30–37]. Surprisingly, while climbing and lifespan traits were previously shown to not be correlated, we found climbing traits to be important in our models of lifespan[16].

The identification of these already-known traits lends validates our approach to identify underlying traits implicated in lifespan and healthspan. Using methods such as Mendelian randomization to screen the functional consequences of manipulating expression of candidate human aging genes in *Drosophila* holds the potential to overcome previous shortcomings of both fields. Indeed, we found several traits previously unconnected to lifespan or healthspan traits that make promising candidates as potential regulators or biomarkers. Further, the few GWAS candidates that are shared between traits could be used to influence multiple health outcomes. The diet-specific nature of our models points to differing mechanisms by which the fly responds to diet. While our data shows the diet-specific impact of metabolites, the mechanisms through which these metabolites work in humans may be different. Considering we saw different responses to metabolite supplementation between two *Drosophila* strains, it is expected that there will be similar variation in the response of individual humans to specific metabolites. Ways to optimize potential effects of traits identified by our approach would be to combine treatments that target specific metabolites based on the genotype of an individual. In total, our work demonstrates the importance and value of incorporating multiple data sets to understand how nature “built” systems that influence lifespan and healthspan traits. Our approach also identifies several new potential mechanisms for how DR influences lifespan and healthspan across multiple genotypes. We demonstrate that combining human and *Drosophila* genetics is an effective approach to further

our understanding of the underlying processes regulating longevity and may ultimately contribute to anti-aging strategies in humans.

Acknowledgements

T.A.U.H. was supported by NIH and NIA award F31AG062112 and is currently supported by NIH/NIA training grant T32AG000266-24. K.A.W. was supported by NIH and NIA award F31AG052299 and NIH/NIA training grant T32AG000266-23. This work was funded by grants from the American Federation of Aging Research (R.B.B. and P.K.); NIH grants R56AG038688, R21AG054121, AG045835 (P.K.), and the Larry L. Hillblom Foundation. We thank the Bloomington Drosophila Stock Center, the Vienna Drosophila Stock Center for providing the flies used.

Author Contributions

T.A.U.H., K.A.W., V.P.N., R.B.B., and P.K. designed research; T.A.U.H., V.P.N., E.C., and D.R. performed research; T.A.U.H. and V.P.N. analyzed data; T.A.U.H. and V.P.N. provided R and Python code; T.A.U.H., V.P.N., and K.A.W. wrote the paper; D.P., R.B.B., P.K., and J.C. provided experimental guidance and funding.

Declaration of Interests

The authors declare no competing interests.

Figure Titles and legends

Figure 1. Diet influences Lifespan and Healthspan in a genotype-specific manner. DGRP show strong strain-specific responses to diet (A) Graphical workflow for modeling data from *Drosophila* and filtering traits using Mendelian Randomization of human data. (B) DGRP strain traits on a high yeast (AL, red) or low yeast (DR, blue) diet were used to determine strain-specific dietary response (DR-AL). (B-E) Bar graphs of DGRP strain dietary response in (B) late-life initial mortality α, (C) late-life rate of aging β, (D) body weight (mg) and (E) max lifespan.

Figure 2. DGRP trait dietary responses correlate with similar traits, but few metabolites correlate with phenotypes. Heatmap of metabolite and phenotype dietary response correlations. Traits are clustered via hierarchical clustering using the Ward method, with 5 clusters highlighted on the diagonal. Names of similar

traits within the same datatype that were correlated are highlighted as popouts. Pearson correlation values are shown as a gradient from 1 (dark blue) to -1 (dark red).

Figure 3. Random Forest models utilize common traits to build models for lifespan. Predictor traits with model importance of 0.02 or higher used to build Random Forest models of DGRP trait dietary response of (A) mean lifespan, (B) median lifespan, (C) maximum lifespan, (D) initial mortality (α) and (E) rate of aging (β) for mortality from day 21 to the day 95% were dead, (F) variance in day of death, (G) body weight (mg), (H) triglyceride levels (µg) normalized by weight (mg), and (I) glucose levels (µg) normalized by weight (mg).

Figure 4. Predictor traits used to build RF lifespan models grouped similar lifespan and healthspan response traits in a network diagram of dietary response. A network diagram of lifespan and healthspan response traits (teal nodes) connected to their top 3 predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Nodes repel each other, forcing response traits built with similar predictor traits to be pushed together (cluster).

Figure 5. Supplementation of Mendelian Randomization (MR) and Random Forest (RF) overlapping metabolites threonine and orotate significantly increases and decreases, respectively, fly lifespan in a genotype-, dose-, and sex-specific manner. (A) Mendelian Randomization results for metabolites found important in Random Forest models of age at death, BMI, and frailty phenotypes measured in humans. (B) AL diet supplementation of 1-, 5-, or 10-mM threonine significantly extends lifespan in Canton S male *Drosophila melanogaster* ($p \sim 2.85*10^{-2}$, $1.95*10^{-2}$, and $9.98*10^{-10}$, respectively). (C) AL diet supplementation of 1-, 2-, or 10-mM orotate significantly shortens lifespan in w1118 male *Drosophila melanogaster* ($p \sim 5.00*10^{-6}$, $2.69*10^{-17}$, and $8.90*10^{-30}$, respectively). (D) Heatmap of mean lifespan fold change values in mean lifespan for males and females of two *Drosophila* strains on an AL diet supplemented with one of three concentrations of either orotate, threonine, choline, or histidine, compared to the water-only vehicle control (0 mM). Lifespans began with 200 flies per group, with *Cox proportional hazards* analysis p-values being summarized (ns – not significant, * <= 0.05, ** <= 0.005, *** <= 0.0005).

Materials and Methods

<u>*Fly lines, husbandry, and diet composition.*</u> All fly lines were maintained on standard fly yeast extract medium containing 1.55% yeast, 5% sucrose, 0.46% agar, 8.5% of corn meal, and 1% acid mix (a 1:1 mix of 10%

propionic acid and 83.6% orthophosphoric acid) prepared in distilled water. To prepare the media, cornmeal (85 g), sucrose (50 g), active dry yeast (16 g, "Saf-instant") and agar (4.6 g) were mixed in a liter of water and brought to boil under constant stirring. Once cooled down to 60˚C 10 ml of acid mix was added to the media. The media were then poured in vials (~10 ml/ vial) or bottles (50 ml/bottle) and allowed to cool down before storing at 4˚C for later usage. These vials or bottles were then seeded with some live yeast just before the flies are transferred and used for maintenance of lab stocks, collection of virgins, or setting up crosses.

For each cross, 12-15 virgin females of either *w1118* or *Canton S* strains were mated with 3-5 males of the same genotype in bottles containing an intermediate diet with 1.55% yeast as a protein source. Flies mated for 5 days, then were removed. 9 days later, non-virgin female progeny were sorted onto an AL (standard diet with 5% yeast) or AL diets containing one of three candidate metabolite concentrations. Concentrations for each metabolite were chosen based on a thorough literature review. 4 to 8 vials of 25 mated female flies per vial were collected for each diet, maintained at 25°C and 65% relative humidity, and were on a 12-hour light/dark cycle.

Lifespan analysis. Flies developed on standard fly 1.5% yeast extract medium were transferred to the necessary diet within 72 hours after eclosion. For survivorship analysis, vials with 25 mated females were transferred to fresh food every other day and fly survival was scored by counting the number of dead flies. Each lifespan was repeated at least once to generate independent biological replicates[4,16,38–40]. We used Cox proportional hazards analysis implemented in the python package 'lifelines' to analyze the significance of the metabolite concentration on survival outcomes compared to vehicle controls (0 mM). We report the probability that B1=0, from fitting the formula phenotype=B1*variable1. Fold changes in mean lifespan compared to 0 mM were calculated using the following formula, (test concentration – 0 mM)/0 mM.

Genome-wide association mapping. We used DGRP release 2 genotypes, and FlyBase R5 coordinates for gene models. As in Nelson et al., 2016, we used only homozygous positions and a minor allele frequency of R 25% to ensure that the minor allele was represented by many observations at a given polymorphic locus[15,16]. The collected phenotype and genotype data were used as input into an association test via ordinary least-squares regression using the StatsModels module in Python[41]. The linear model was phenotype = β1 x genotype + β2 x diet + β3 x genotype x diet + intercept. Nominal p-values denoted as ‘‘genotype’’ in Figure 1A report the probability that β1 ≠ 0, and those denoted as ‘‘interaction’’ report the probability that β3 ≠ 0.

Principal Component Analysis (PCA). All DGRP metabolite and phenotype data for strains were first scaled using the StandardScaler function of the sklearn package. Then missing values were imputed based on the mean of all other strains for that trait. PCA was performed on all strains using the PCA and fit_transform functions in the sklearn decomposition package to observe how well the combined metabolome and phenome can separate samples by diet.

Pearson Correlation Analysis. Pearson correlations between all metabolites and phenotypes was performed on all strains for each diet combination using the cor() function in the stats package. Heatmaps were created for each correlation matrix with the corrplot function, using the ward.D2 hierarchical clustering method to group predictor traits.

Random Forest Modeling. DGRP metabolite and phenotype data from each diet combination were split into a response variable set composed of 7 lifespan traits [mean lifespan, median lifespan, max LS, day 95% of flies dead, initial mortality (α), rate of aging (β), and variance in day of death] and 3 healthspan traits [glucose (μg)/weight (mg), triglycerides (μg)/weight (mg), and body weight (mg)], and a predictor set containing the remaining traits. Missing values were imputed based on the mean of all other strains for that trait using the SimpleImputer function in the sklearn.impute package[42]. Predictor and response traits were then split into training (75% of data) and test (25% of data) sets using the train_test_split function in the sklearn.model_selection package[42]. Random Forest models were generated using the RandomForestRegressor function in the sklearn.ensemble package using 10,000 estimators[42]. Mean absolute percentage error and R2 values were determined using the mean_squared_error and r2_score functions in the sklearn.metrics package[42]. Feature importances were extracted from the final model.

Overlapping Trait Analysis. Analysis of the chances of traits overlapping between the Random Forest models was performed in python using the hypergeom function from the scipy.stats package, the number of traits in each group, their overlap, and total number of traits that could have been used to build the model[43].

Network Diagrams. Network diagrams were created for all diet combinations using the networkx package in Python to visualize and identify common traits or GWAS candidate genes among the Random Forest models for each diet combination[44]. Network layouts were determined by first applying the spring_layout with a k parameter value of $5/\sqrt{node\#}$, then applying the kamada_kawai_layout with a scale value of 5.

Mendelian Randomization. Genetic instruments i.e., SNPs strongly ($P<5\times10^{-6}$) predicting metabolites of interest were extracted from an existing GWAS of serum metabolites in 1960 individuals (96.6% female) of European descent[45]. As the primary outcome of the present analysis, we obtained genetic association estimates for the variants selected as metabolite status instruments with combined parental age at death, from the UK Biobank participants of European descent. Parental lifespan is widely used as an outcome in genetic association studies as offspring inherit one-half of their genetic code from their parents resulting in genotypic and phenotypic correlations[46].

Secondary outcomes related to health and lifespan included Body Mass Index (BMI) and frailty index (FI). Briefly, summary statistics for FI were collected from a recent GWAS caried out among participants in the UK Biobank[47]. The FI was constructed based on 49 items ranging from physical to mental well-being and calculated as a proportion of the sum of all deficits. Genetic predictors of BMI were obtained from the largest sex specific meta-analysis of GWAS in the UK Biobank GWAS and the GIANT consortium[48]. More details of the GWAS on each of the outcomes can be found in their respective published articles.

As the main analysis for MR, we used the inverse-variance weighted method (IVW)[49] (Supplementary Table 4). IVW is considered to provide the most accurate estimate but known to be sensitive to pleiotropy[49]. As sensitivity analyses, we also employed the weighted median approach (WM)[50] and MR-Egger regression[51] (Supplementary Table 4). The weighted median method assumes >50% of the weight comes from valid SNPs and can generate consistent causal estimates[50]. The MR-Egger regression detects possible pleiotropic effects and provides corrected estimates for pleiotropy[51]. The p-value for the intercept in MR-Egger was used to detect the directional pleiotropic effect[51]. To gauge evidence of directional pleiotropy, we used Cochrane's Q test[52], with this being the measure of heterogeneity between ratio estimates of variants (Supplementary Table 4). Finally, to identify candidate genes for functional screening in *Drosophila* we set the significance threshold using Bonferroni correction, at $P<5.00\times10^{-3}$ ($=0.05/10$) to declare a causal relationship for the IVW-based MR estimate. Associations with $P<0.05$, but not reaching the Bonferroni-corrected threshold, were reported as suggestive of association. All analyses were performed using the 2-sample MR analyses conducted in R using the TwoSampleMR[53] package.

Supplemental Figure Titles and Legends

Supplemental Figure 1. Variation of DGRP dietary response in lifespan and healthspan response traits. Dietary response (DR-AL) of 160 DGRP strains in 7 lifespan and healthspan metrics, strains plotted in ascending order.

Supplemental Figure 2. DGRP metabolite and phenotype trait correlations on DR and AL, together (A), or separate (B&C). Heatmap of metabolite and phenotype value correlations on DR and AL, together, or separate. Traits are clustered via hierarchical clustering using the Ward method, with 5 clusters highlighted on the diagonal. Pearson correlation values are shown as a gradient from 1 (dark blue) to -1 (dark red).

Supplemental Figure 3. Network diagram of DGRP random forest models of response traits and the predictor traits used to build them; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Network layout a combination of a spring layout followed by a kamada kawai layout, with nodes repelling each other, forcing response traits built with similar predictor traits to be pushed together (cluster). (A) AL, (B) DR, and (C) DR & AL.

Supplemental Figure 4. Network diagram of DGRP random forest models of response traits, the predictor traits used to build them, and GWAS candidate genes; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. GWAS candidates (green nodes) with p-values of magnitude 10^{-5} or lower are connected to their trait by edges with widths inversely weighted by p-value (thicker edges represent lower p-values). Network layout a combination of a spring layout followed by a kamada kawai layout, with nodes repelling each other, forcing response traits built with similar predictor traits to be pushed together (cluster). (A) AL, (B) DR, and (C) DR & AL.

Supplemental Figure 5. Lifespans of Canton S or w1118 females and males on AL food supplemented with one concentration of a metabolite. Survival curves for females and males of *Drosophila melanogaster* strains Canton S or w1118 on an AL diet supplemented with one of three concentrations of either choline (A-C), histidine (D-G), or threonine (H-K), or the water vehicle control. All experiments began with 200 flies per group, with cox proportional hazards analysis used to analyze lifespan differences compared to the vehicle

only (0 mM) control. Fold changes in mean lifespan compared to the 0 mM control and summary p-values are shown in **Figure 5D**.

Supplemental Table Titles and Legends

Supplementary Table 1. Trait list and DGRP strains with missing values. List of metabolite and phenotype traits and their descriptions on the first sheet. The second sheet contains a list of DGRP strains with missing phenotype values that were imputed with the mean of all other strains for that trait.

Supplementary Table 2. DGRP datasets. Raw DGRP data on each diet used for correlation and modeling analysis.

Supplementary Table 3. Random Forest response model traits. Order of traits used to build each random forest response model by the proportion of trees/estimators that trait was included in out of the 10,000 trees/estimators run for each model.

Supplementary Table 4. Mendelian Randomization Results.

Supplementary Table 5. Model trait selection and validation determination.

References

1. Wilson KA, Chamoli M, Hilsabeck TA, et al. Evaluating the beneficial effects of dietary restrictions: A framework for precision nutrigeroscience. *Cell Metab*. 2021;33(11):2142-2173. doi:10.1016/j.cmet.2021.08.018

2. Wilson KA, Beck JN, Nelson CS, et al. GWAS for Lifespan and Decline in Climbing Ability in Flies upon Dietary Restriction Reveal decima as a Mediator of Insulin-like Peptide Production. *Current Biology*. 2020;30(14):2749-2760.e3. doi:10.1016/j.cub.2020.05.020

3. Green CL, Pak HH, Richardson NE, et al. Sex and genetic background define the metabolic, physiologic, and molecular response to protein restriction. *Cell Metab*. 2022;34(2):209-226.e5. doi:10.1016/j.cmet.2021.12.018

4. Jin K, Wilson KA, Beck JN, et al. Genetic and metabolomic architecture of variation in diet restriction-mediated lifespan extension in Drosophila. *PLoS Genet*. 2020;16(7):1-22. doi:10.1371/journal.pgen.1008835

5. Partridge L, Fuentealba M, Kennedy BK. The quest to slow ageing through drug discovery. *Nat Rev Drug Discov*. 2020;19(8):513-532. doi:10.1038/s41573-020-0067-7

6. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023;186(2):243-278. doi:10.1016/j.cell.2022.11.001

7. de Magalhães JP. Why genes extending lifespan in model organisms have not been consistently associated with human longevity and what it means to translation research. *Cell Cycle*. 2014;13(17):2671-2673. doi:10.4161/15384101.2014.950151

8. Nelson CS, Beck JN, Wilson KA, Pilcher ER, Kapahi P, Brem RB. Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. *BMC Genomics*. 2016;17(1). doi:10.1186/s12864-016-3137-9

9. Canfield CA, Bradshaw PC. Amino acids in the regulation of aging and aging-related diseases. *Transl Med Aging*. 2019;3(2019):70-89. doi:10.1016/j.tma.2019.09.001

10. Aon MA, Bernier M, Mitchell SJ, et al. Untangling Determinants of Enhanced Health and Lifespan through a Multi-omics Approach in Mice. *Cell Metab*. 2020;32(1):100-116.e4. doi:10.1016/j.cmet.2020.04.018

11. Yap YW, Rusu PM, Chan AY, et al. Restriction of essential amino acids dictates the systemic metabolic response to dietary protein dilution. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-16568-z

12. Kim J, Jo Y, Cho D, Ryu D. L-threonine promotes healthspan by expediting ferritin-dependent ferroptosis inhibition in C. elegans. *Nat Commun*. 2022;13(1). doi:10.1038/s41467-022-34265-x

13. Solvang SEH, Hodge A, Watne LO, et al. Kynurenine Pathway Metabolites in the Blood and Cerebrospinal Fluid Are Associated with Human Aging. *Oxid Med Cell Longev*. 2022;2022. doi:10.1155/2022/5019752

14. Jové M, Naudí A, Aledo JC, et al. Plasma long-chain free fatty acids predict mammalian longevity. *Sci Rep*. 2013;3. doi:10.1038/srep03346

15. Nelson CS, Beck JN, Wilson KA, Pilcher ER, Kapahi P, Brem RB. Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. *BMC Genomics*. 2016;17(1). doi:10.1186/s12864-016-3137-9

16. Wilson KA, Beck JN, Nelson CS, et al. GWAS for Lifespan and Decline in Climbing Ability in Flies upon Dietary Restriction Reveal decima as a Mediator of Insulin-like Peptide Production. *Current Biology*. 2020;30(14):2749-2760.e3. doi:10.1016/j.cub.2020.05.020

17. Wilson KA, Chamoli M, Hilsabeck TA, et al. Evaluating the beneficial effects of dietary restrictions: A framework for precision nutrigeroscience. *Cell Metab*. 2021;33(11):2142-2173. doi:10.1016/j.cmet.2021.08.018

18. Zhou S, Morgante F, Geisz MS, Ma J, Anholt RRH, Mackay TFC. Systems genetics of the Drosophila metabolome. *Genome Res*. 2020;30(3):392-405. doi:10.1101/gr.243030.118

19. Everett LJ, Huang W, Zhou S, et al. Gene expression networks in the drosophila genetic reference panel. *Genome Res*. 2020;30(3):485-496. doi:10.1101/gr.257592.119

20. Rohde PD, Bøcker A, Jensen CAB, et al. Article genotype and trait specific responses to rapamycin intake in drosophila melanogaster. *Insects*. 2021;12(5):1-11. doi:10.3390/insects12050474

21. Zhao X, Golic FT, Harrison BR, et al. The metabolome as a biomarker of aging in Drosophila melanogaster. *Aging Cell*. 2022;(November 2021):1-14. doi:10.1111/acel.13548

22. Noberasco G, Odetti P, Boeri D, Maiello M, Adezati L. Malondialdehyde (MDA) level in diabetic subjects. Relationship with blood glucose and glycosylated hemoglobin. *Biomedicine and Pharmacotherapy*. 1991;45(4-5):193-196. doi:10.1016/0753-3322(91)90107-5

23. Moreto F, De Oliveira EP, Manda RM, Burini RC. The higher plasma malondialdehyde concentrations are determined by metabolic syndrome-related glucolipotoxicity. *Oxid Med Cell Longev*. 2014;2014. doi:10.1155/2014/505368

24. Jalees SS, Rosaline M. Study of malondialdehyde and estimation of blood glucose levels in patients with diabetes mellitus with cataract. 2017;4(September):319-323. doi:10.18231/2394-6377.2017.0074

25. Hachinohe M, Yamane M, Akazawa D, et al. A Reduction in Age-Enhanced Gluconeogenesis Extends Lifespan. *PLoS One*. 2013;8(1). doi:10.1371/journal.pone.0054011

26. Treaster SB, Chaudhuri AR, Austad SN. Longevity and GAPDH stability in bivalves and mammals: A convenient marker for comparative gerontology and proteostasis. *PLoS One*. 2015;10(11):1-13. doi:10.1371/journal.pone.0143680

27. Landis GN, Doherty D V., Yen CA, et al. Metabolic signatures of life span regulated by mating, sex peptide, and mifepristone/RU486 in female drosophila melanogaster. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. 2021;76(2):195-204. doi:10.1093/gerona/glaa164

28. Lugo-Huitrón R, Ugalde Muñiz P, Pineda B, Pedraza-Chaverrí J, Ríos C, Pérez-De La Cruz V. Quinolinic acid: An endogenous neurotoxin with multiple targets. *Oxid Med Cell Longev*. 2013;2013. doi:10.1155/2013/104024

29. Castro-Portuguez R, Sutphin GL. Kynurenine pathway, NAD+ synthesis, and mitochondrial function: Targeting tryptophan metabolism to promote longevity and healthspan. *Exp Gerontol*. 2020;132:110841. doi:10.1016/j.exger.2020.110841

30. Lombardi B, Pani P, Schlunk FF. Choline-deficiency fatty liver: impaired release of hepatic triglycerides. *J Lipid Res*. 1968;9(4):437-446. doi:10.1016/s0022-2275(20)42721-x

31. Wu G, Zhang L, Li T, Lopaschuk G, Vance DE, Jacobs RL. Choline deficiency attenuates body weight gain and improves glucose tolerance in ob/ob mice. *J Obes*. 2012;2012(Cd):1-7. doi:10.1155/2012/319172

32. Emran S, Yang M, He X, Zandveld J, Piper MDW. Target of rapamycin signalling mediates the lifespan-extending effects of dietary restriction by essential amino acid alteration. *Aging*. 2014;6(5):390-398. doi:10.18632/aging.100665

33. Roe AJ, Zhang S, Bhadelia RA, et al. Choline and its metabolites are differently associated with cardiometabolic risk factors, history of cardiovascular disease, and MRI-documented cerebrovascular disease in older adults. *American Journal of Clinical Nutrition*. 2017;105(6):1283-1290. doi:10.3945/ajcn.116.137158

34. Ravichandran M, Priebe S, Grigolon G, et al. Impairing L-Threonine Catabolism Promotes Healthspan through Methylglyoxal-Mediated Proteohormesis. *Cell Metab*. 2018;27(4):914-925.e5. doi:10.1016/j.cmet.2018.02.004

35. Bayliak MM, Lylyk MP, Maniukh O V., Storey JM, Storey KB, Lushchak VI. Dietary l-arginine accelerates pupation and promotes high protein levels but induces oxidative stress and reduces fecundity and life span in Drosophila melanogaster. *J Comp Physiol B*. 2018;188(1):37-55. doi:10.1007/s00360-017-1113-6

36. Velazquez R, Ferreira E, Knowles S, et al. Lifelong choline supplementation ameliorates Alzheimer's disease pathology and associated cognitive deficits by attenuating microglia activation. *Aging Cell*. 2019;18(6):1-11. doi:10.1111/acel.13037

37. Huang J, Ladeiras D, Yu Y, Ming XF, Yang Z. Detrimental Effects of Chronic L-Arginine Rich Food on Aging Kidney. *Front Pharmacol*. 2021;11(January):1-13. doi:10.3389/fphar.2020.582155

38. Akagi K, Wilson KA, Katewa SD, et al. Dietary restriction improves intestinal cellular fitness to enhance gut barrier function and lifespan in D. melanogaster. *PLoS Genet*. Published online 2018. doi:10.1371/journal.pgen.1007777

39. Sharma A, Akagi K, Pattavina B, et al. Musashi expression in intestinal stem cells attenuates radiation-induced decline in intestinal permeability and survival in Drosophila. *Sci Rep*. 2020;10(1):1-16. doi:10.1038/s41598-020-75867-z

40. Katewa SD, Akagi K, Bose N, et al. Peripheral Circadian Clocks Mediate Dietary Restriction-Dependent Changes in Lifespan and Fat Metabolism in Drosophila. *Cell Metab*. 2016;23(1):143-154. doi:10.1016/j.cmet.2015.10.014

41. Seabold, S., Perktold J. statsmodels: Econometric and statistical modeling with python. Published online 2010.

42. Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, et al. *Scikit-Learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. Vol 12.; 2011. http://scikit-learn.sourceforge.net.

43. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2

44. Hagberg A, Schult D, Swart P. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T, Millman J, eds. ; 2008:11-15.

45. Long T, Hicks M, Yu HC, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet*. 2017;49(4):568-578. doi:10.1038/ng.3809

46. Pilling LC, Kuo CL, Sicinski K, et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging*. 2017;9(12):2504-2520. doi:10.18632/aging.101334

47. Atkins JL, Jylhävä J, Pedersen NL, et al. A genome-wide association study of the frailty index highlights brain pathways in ageing. *Aging Cell*. 2021;20(9):e13459. doi:10.1111/acel.13459

48. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641-3649. doi:10.1093/hmg/ddy271

49. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. doi:10.1002/gepi.21758

50. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-314. doi:10.1002/gepi.21965

51. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017;32(5):377-389. doi:10.1007/s10654-017-0255-x

52. Bowden J, Del Greco M F, Minelli C, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int J Epidemiol*. 2019;48(3):728-742. doi:10.1093/ije/dyy258

53. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7. doi:10.7554/eLife.34408

Figure 5: Supplementation of Mendelian Randomization (MR) and Random Forest (RF) overlapping metabolites threonine and orotate significantly increases and decreases, respectively, fly lifespan in a genotype-, dose-, and sex-specific manner.

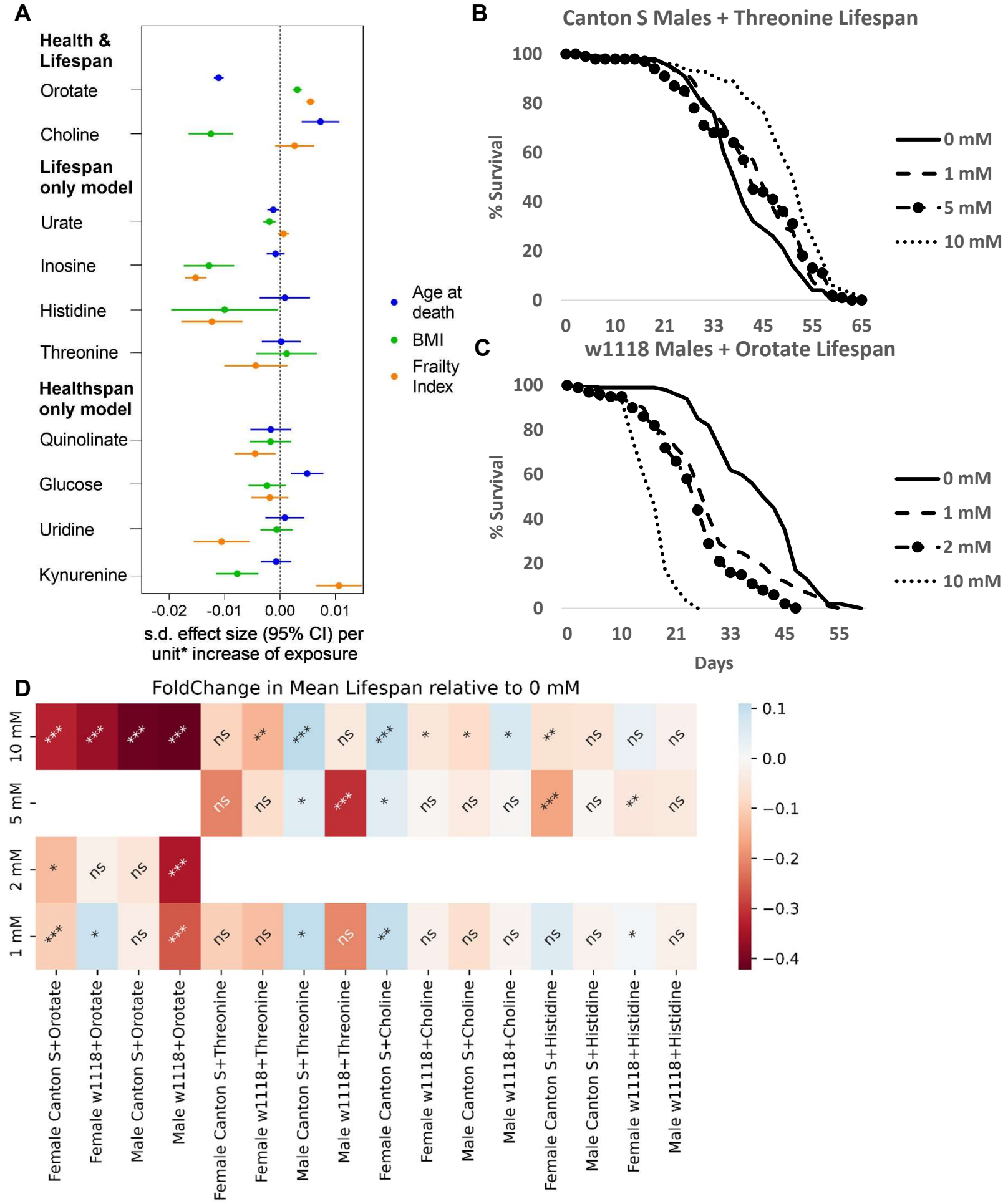
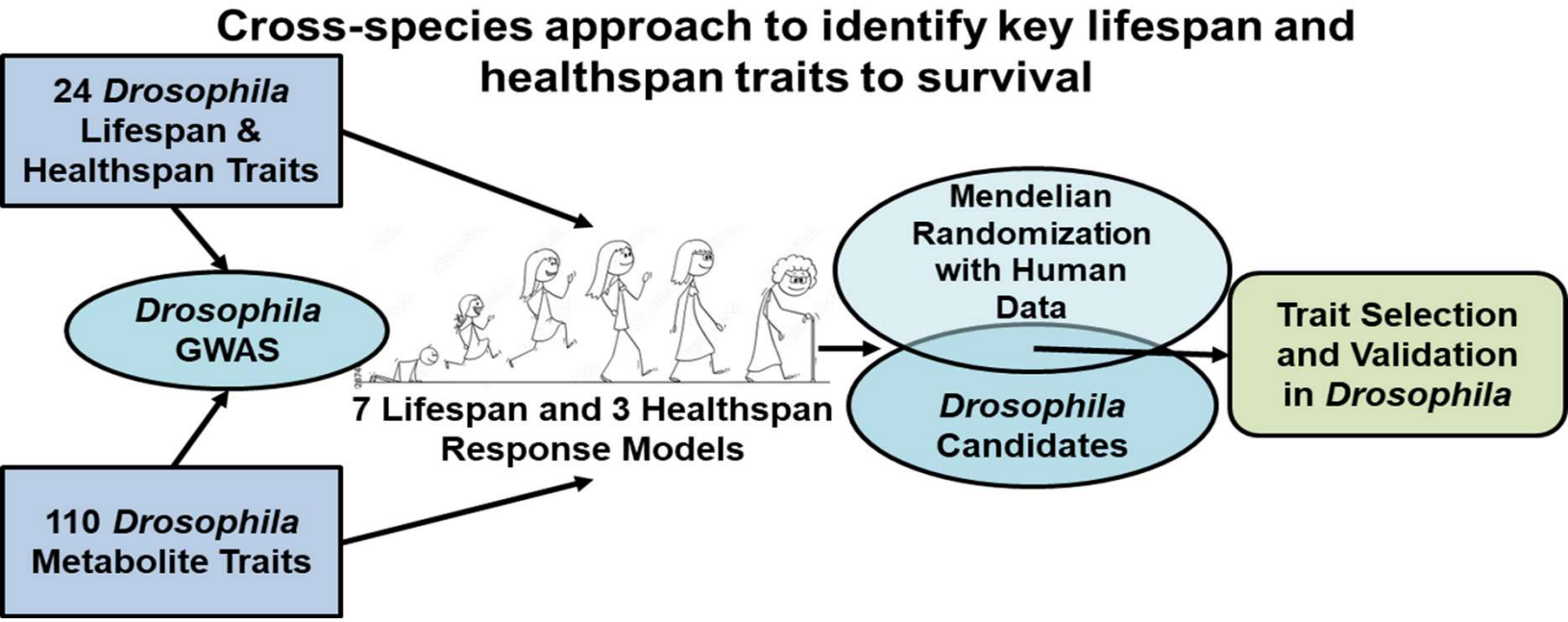
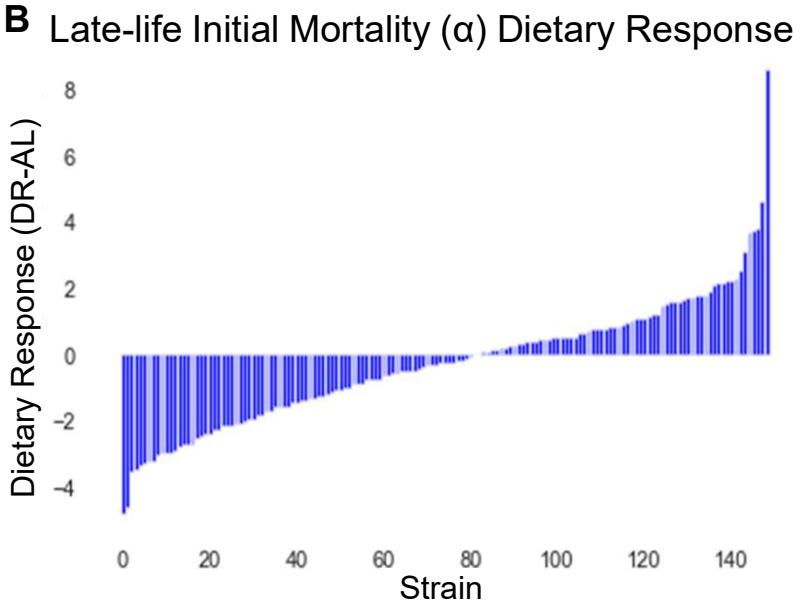
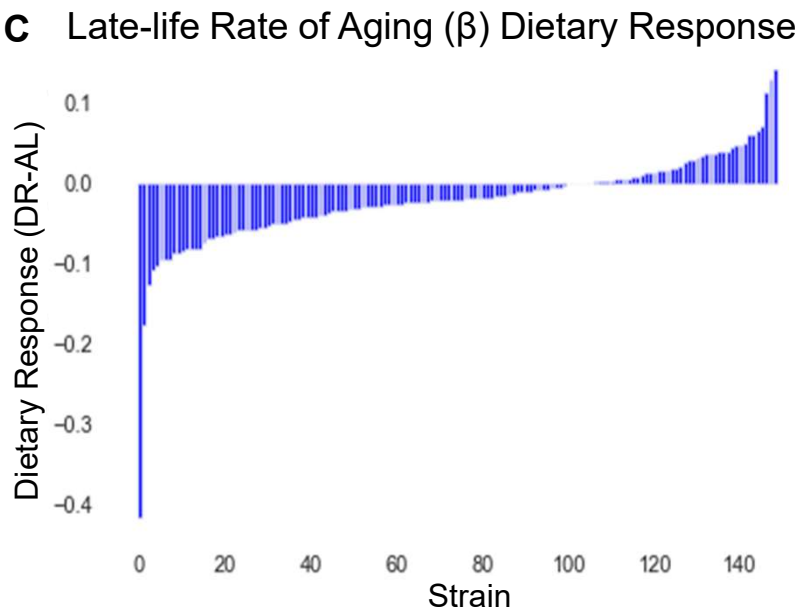


Figure 1: Diet influences Lifespan and Healthspan in a genotype-specific manner.







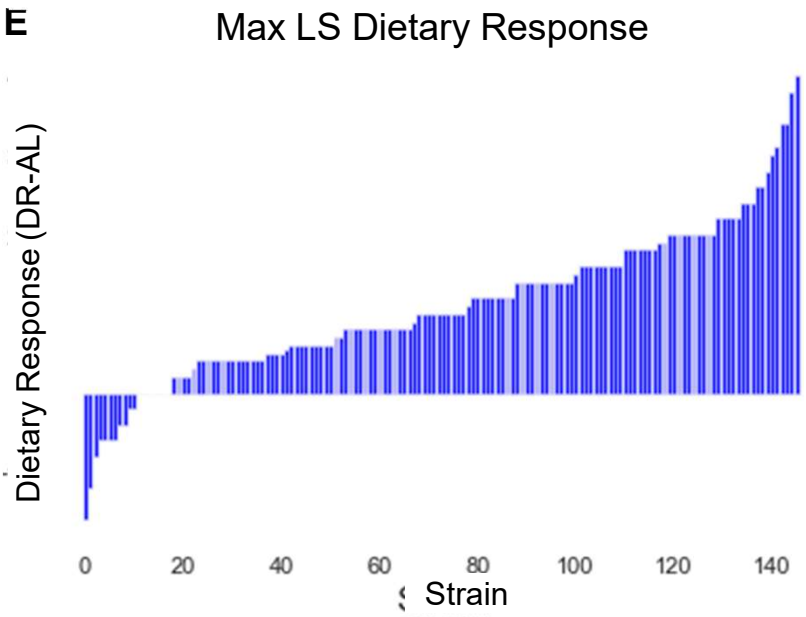
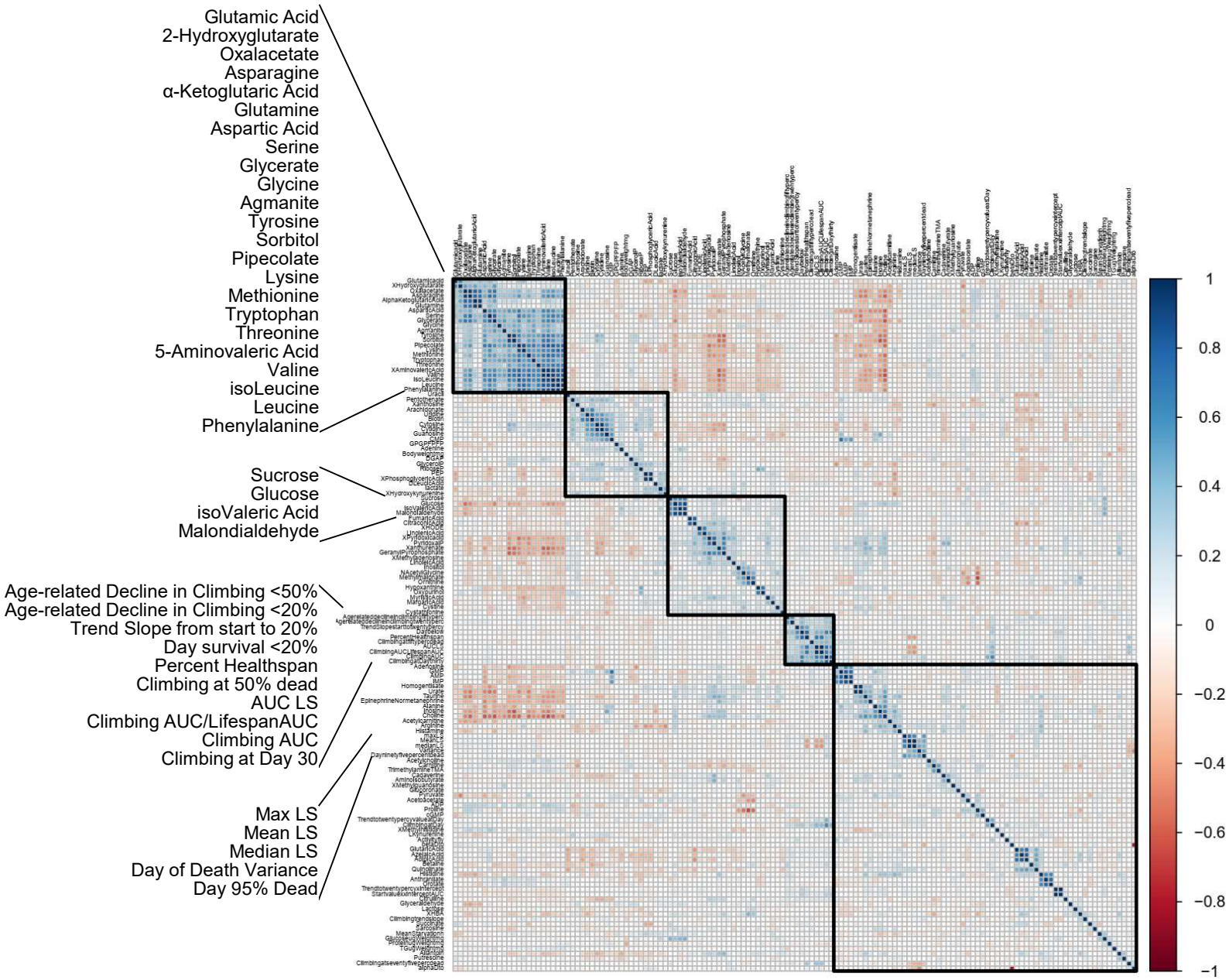
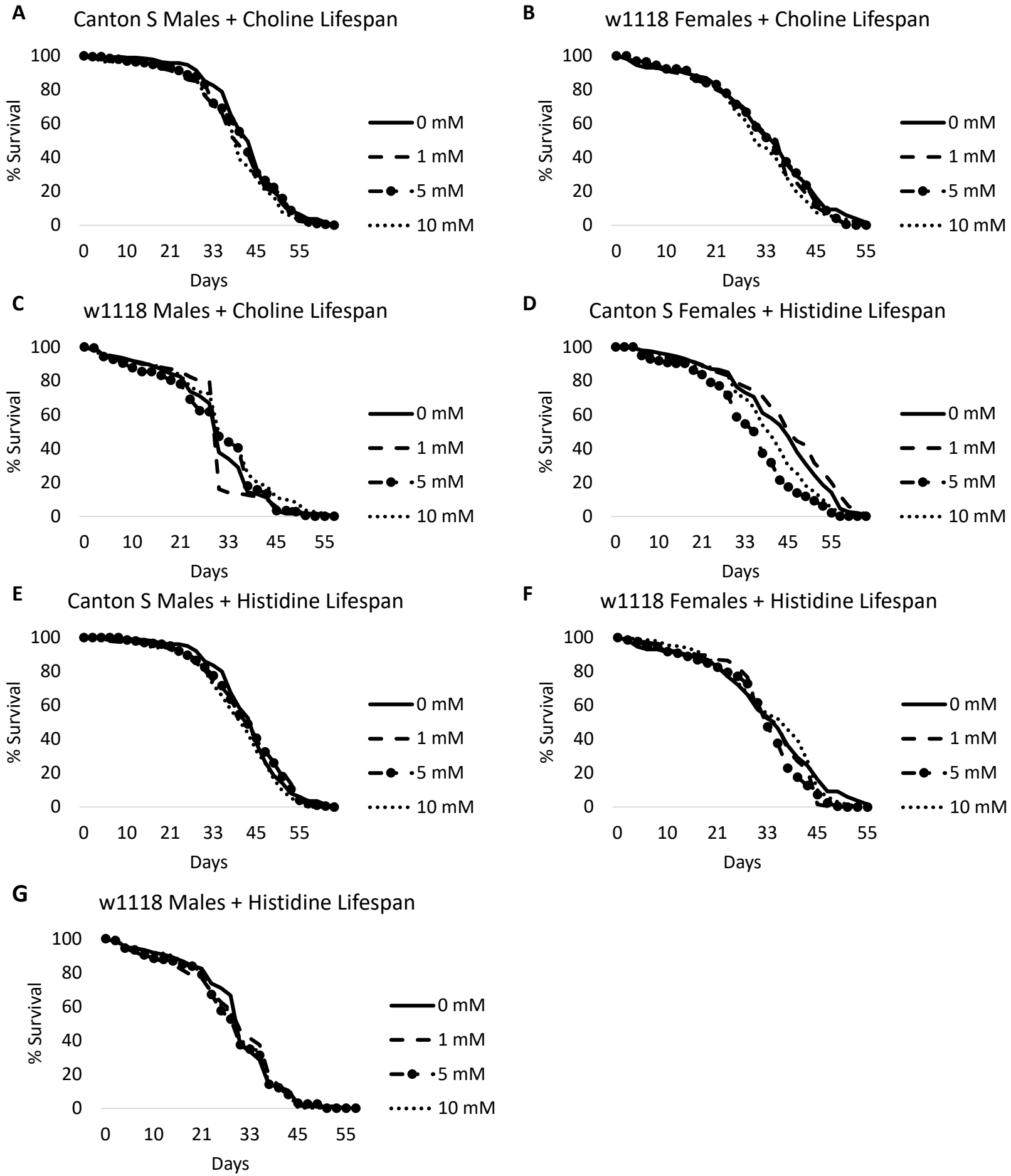


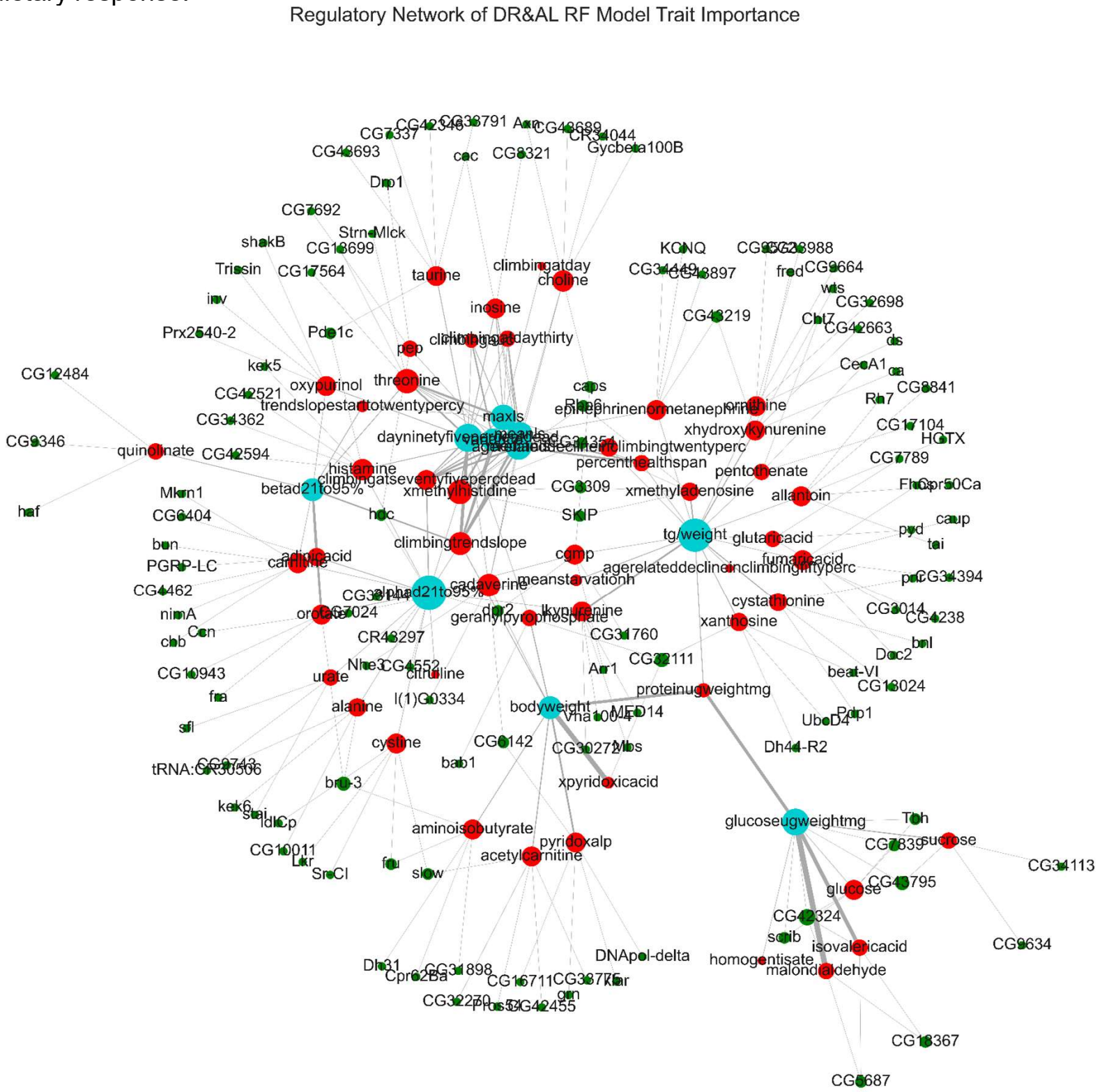
Figure 2: DGRP trait dietary responses correlate with similar traits, but few metabolites correlate with phenotypes.



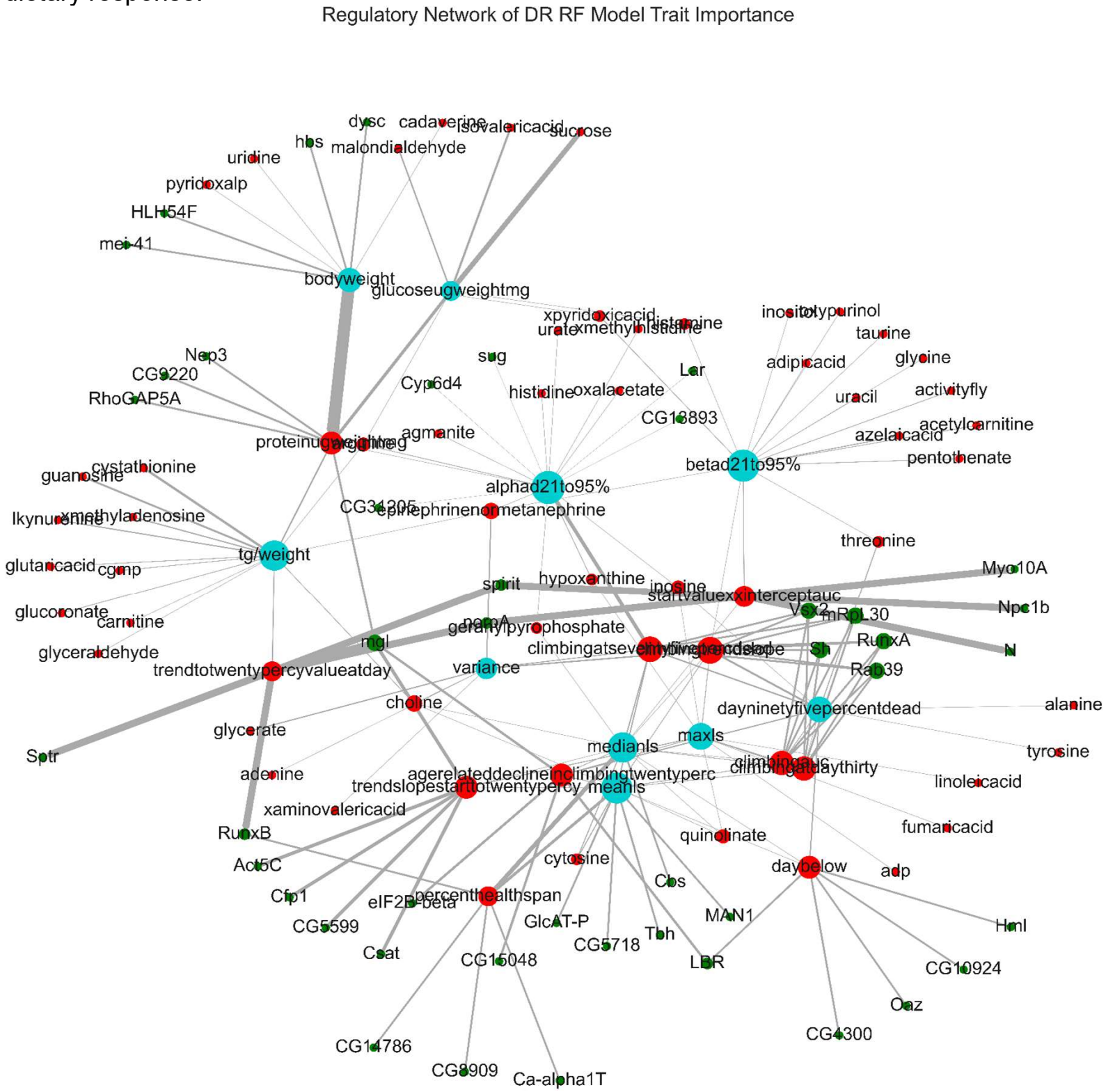
Sup Fig 5: Lifespans of Canton S or w1118 females and males on AL food supplemented with one of three concentrations of a metabolite.



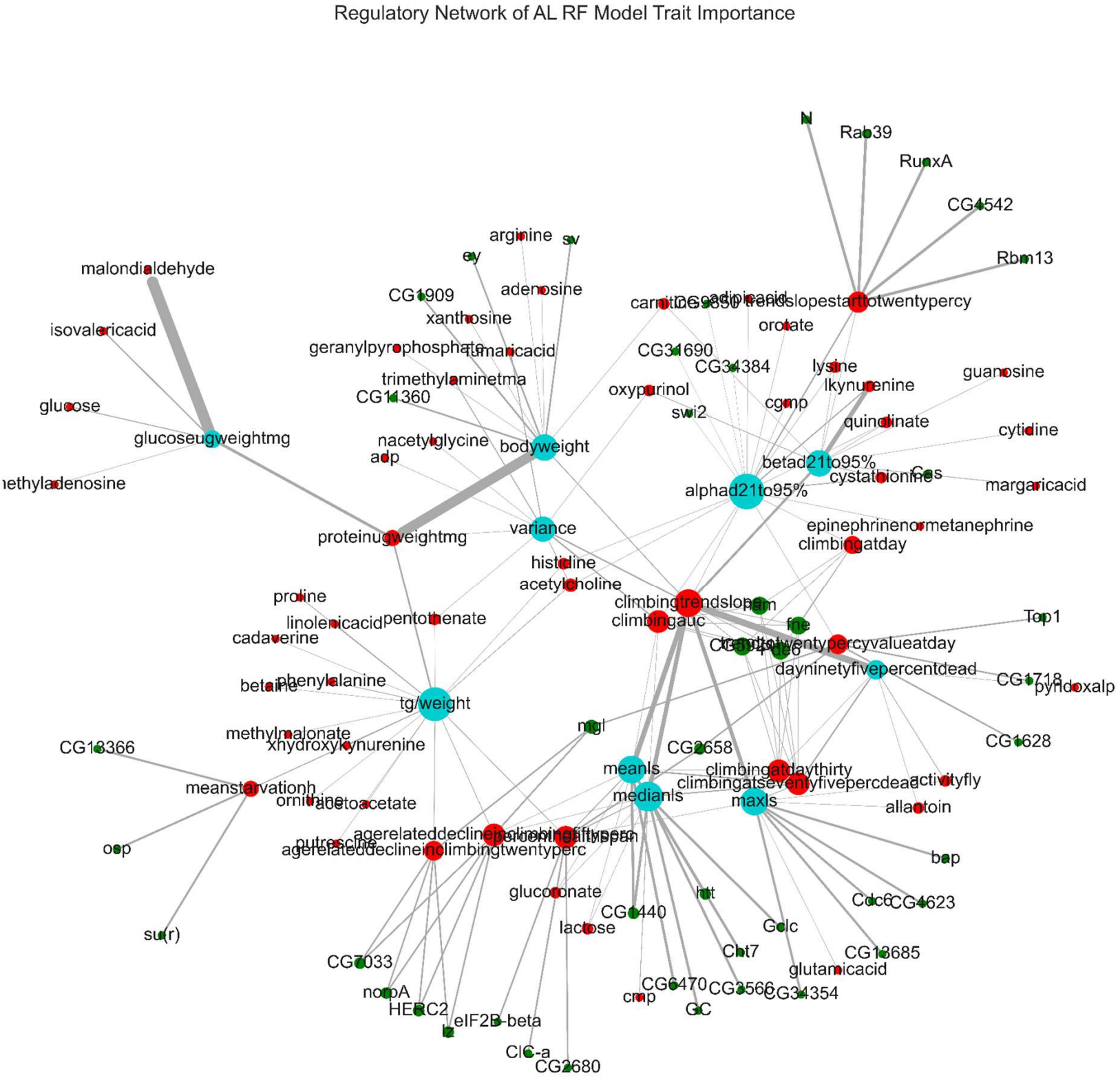
Sup Fig 4D: Predictor traits used to build RF lifespan models, and their GWAS gene candidates, grouped similar lifespan response traits in a network diagram of DGRP dietary response.



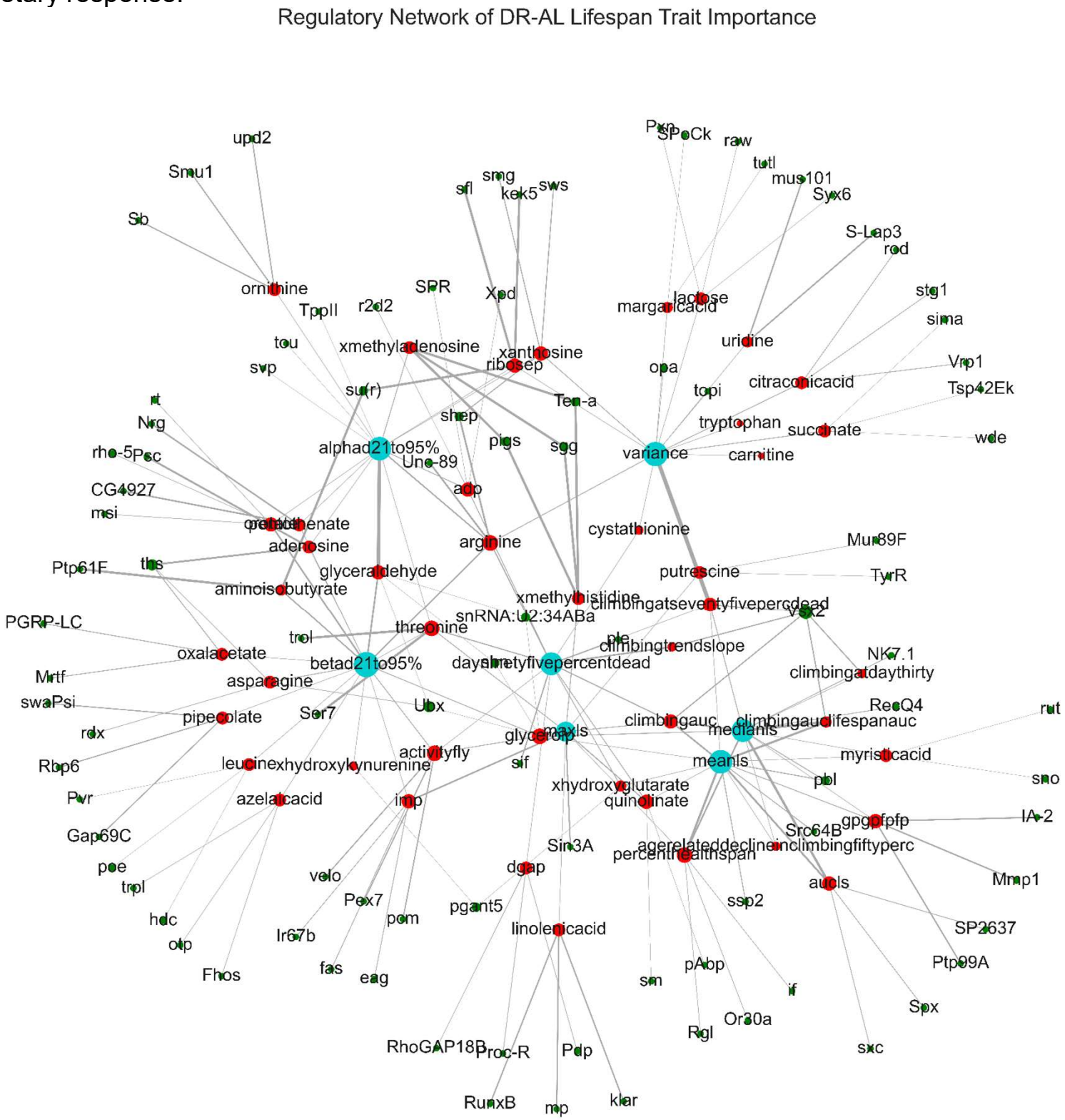
Sup Fig 4C: Predictor traits used to build RF lifespan models, and their GWAS gene candidates, grouped similar lifespan response traits in a network diagram of DGRP dietary response.



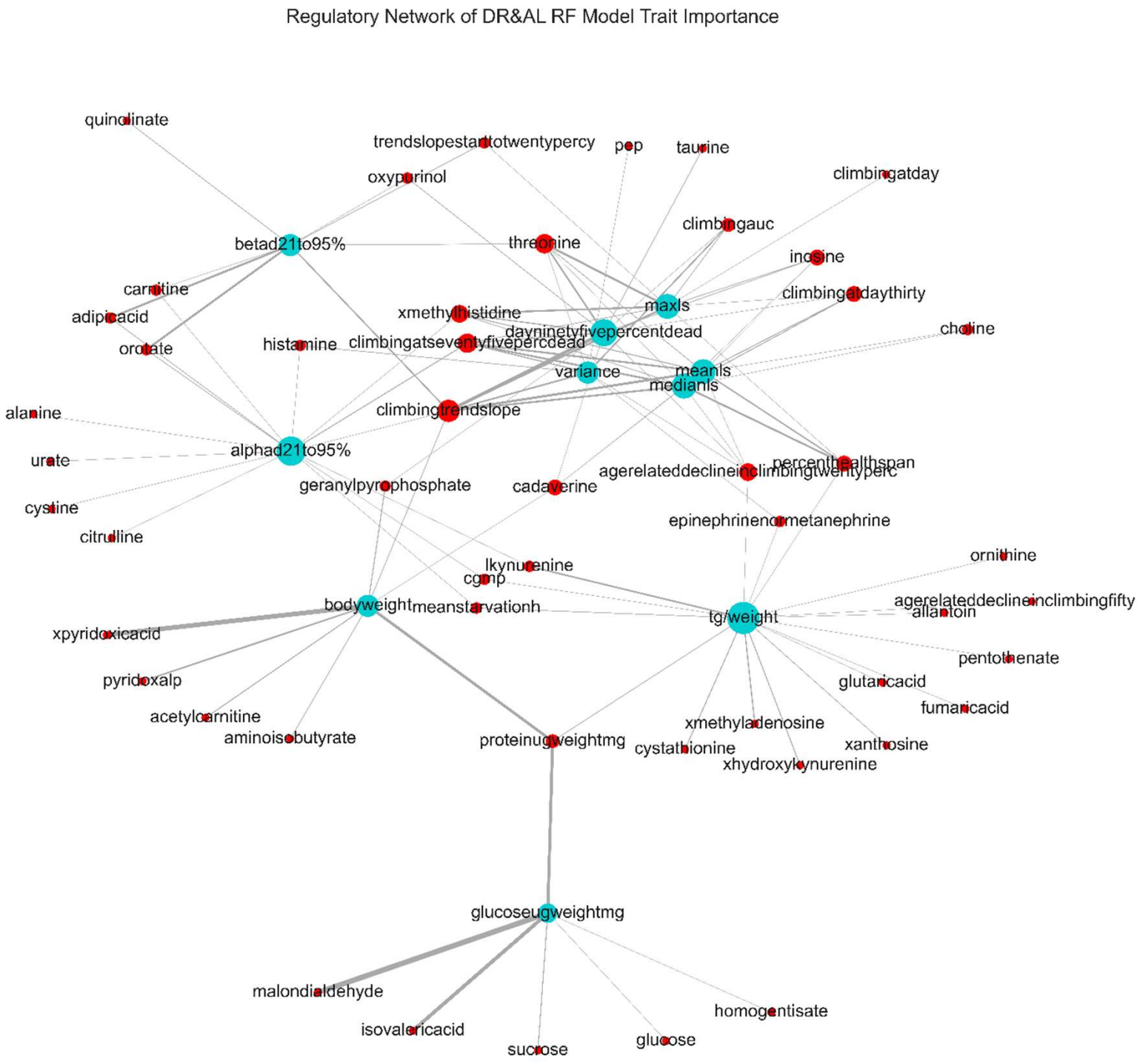
Sup Fig 4B: Predictor traits used to build RF lifespan models, and their GWAS gene candidates, grouped similar lifespan response traits in a network diagram of DGRP dietary response.



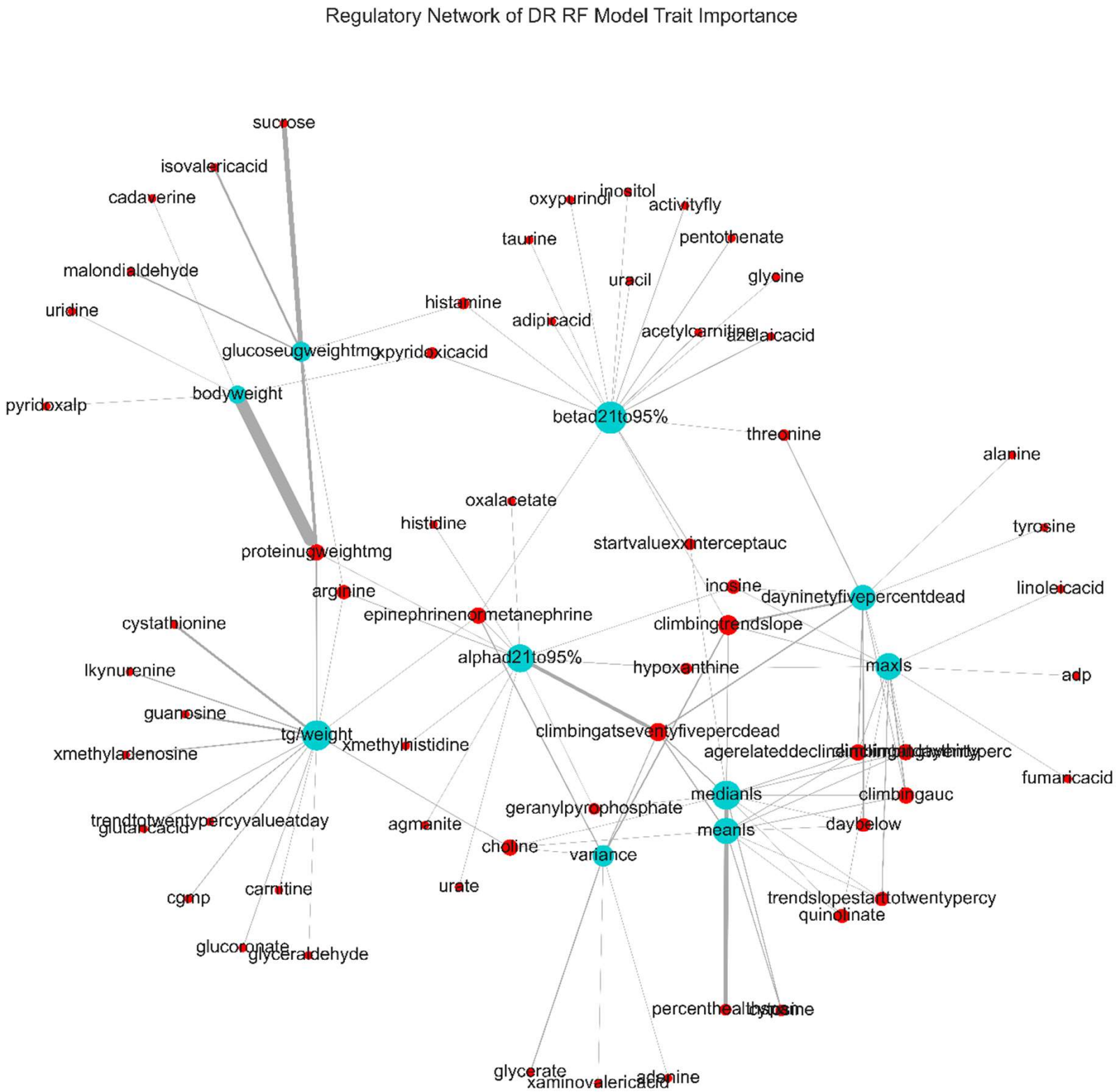
Sup Fig 4A: Predictor traits used to build RF lifespan models, and their GWAS gene candidates, grouped similar lifespan response traits in a network diagram of DGRP dietary response.



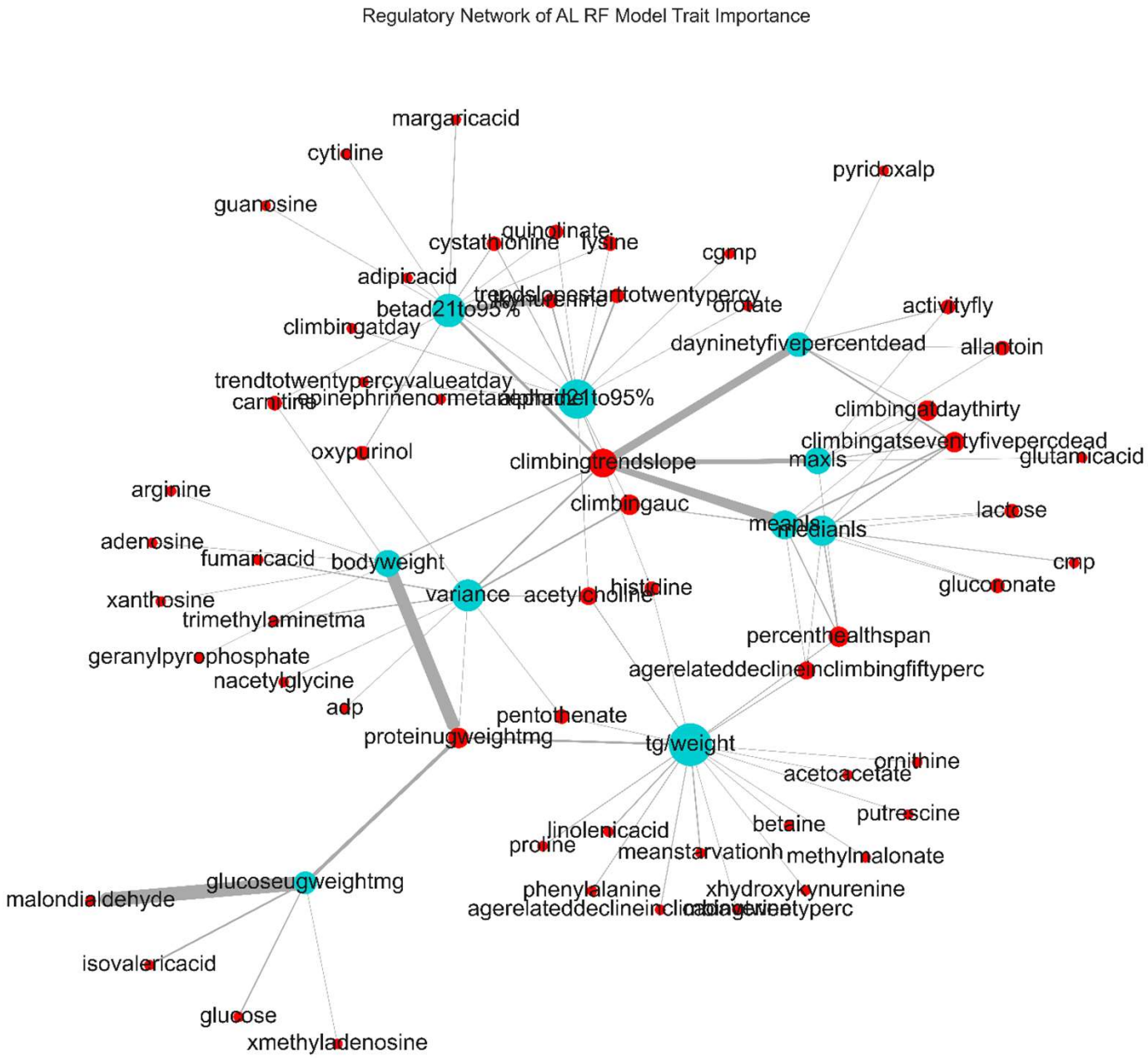
Sup Fig 3C: Pheno/Met/GWAS network – DR&AL



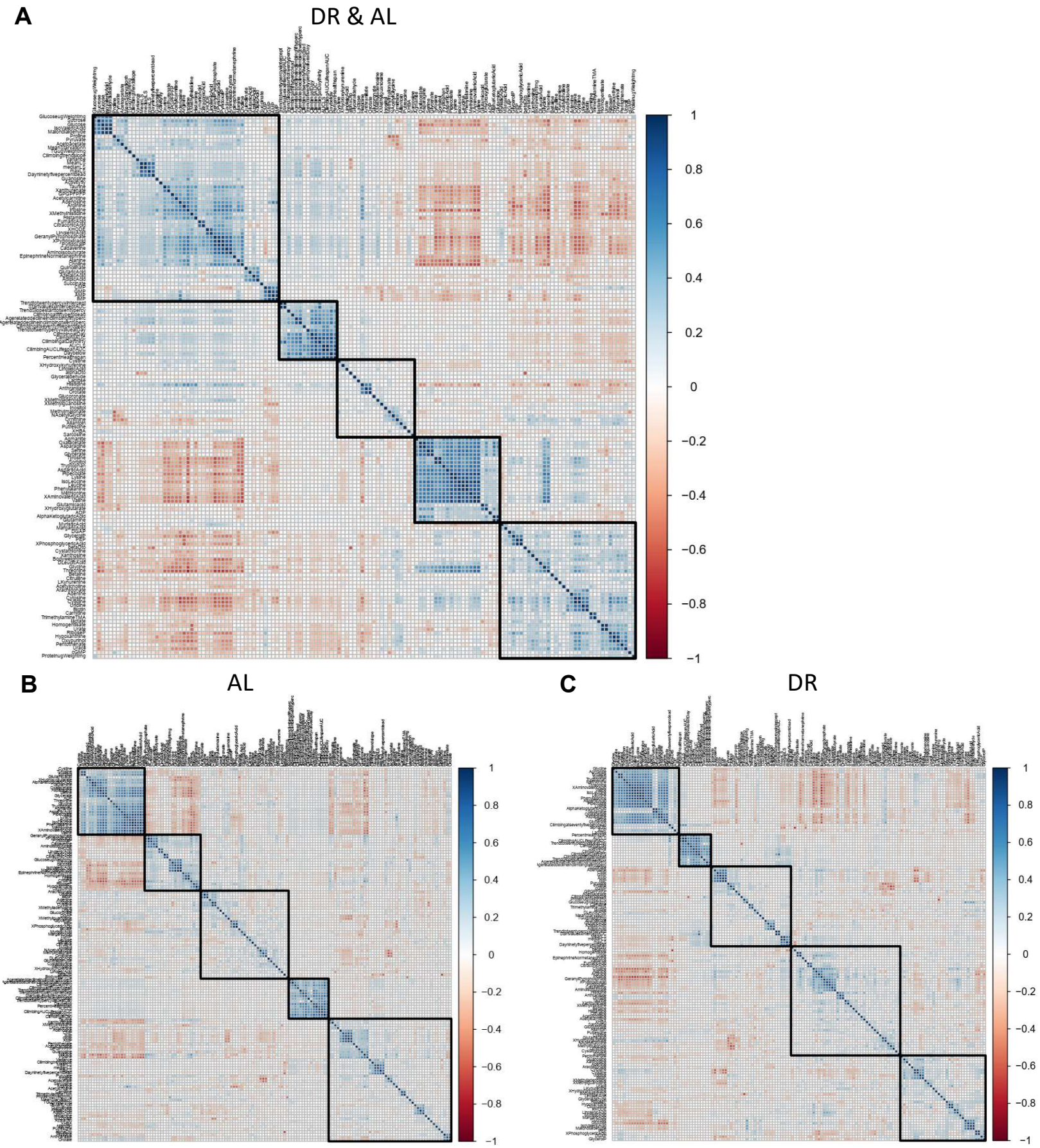
Sup Fig 3B: Pheno/Met/GWAS network - DR



Sup Fig 3A: Pheno/Met/GWAS network - AL



Sup Fig 2: DGRP metabolite and phenotype trait correlations on DR and AL, together (A), or separate (B&C).



Sup Fig 1: Variation of DGRP dietary response in lifespan and healthspan response traits.

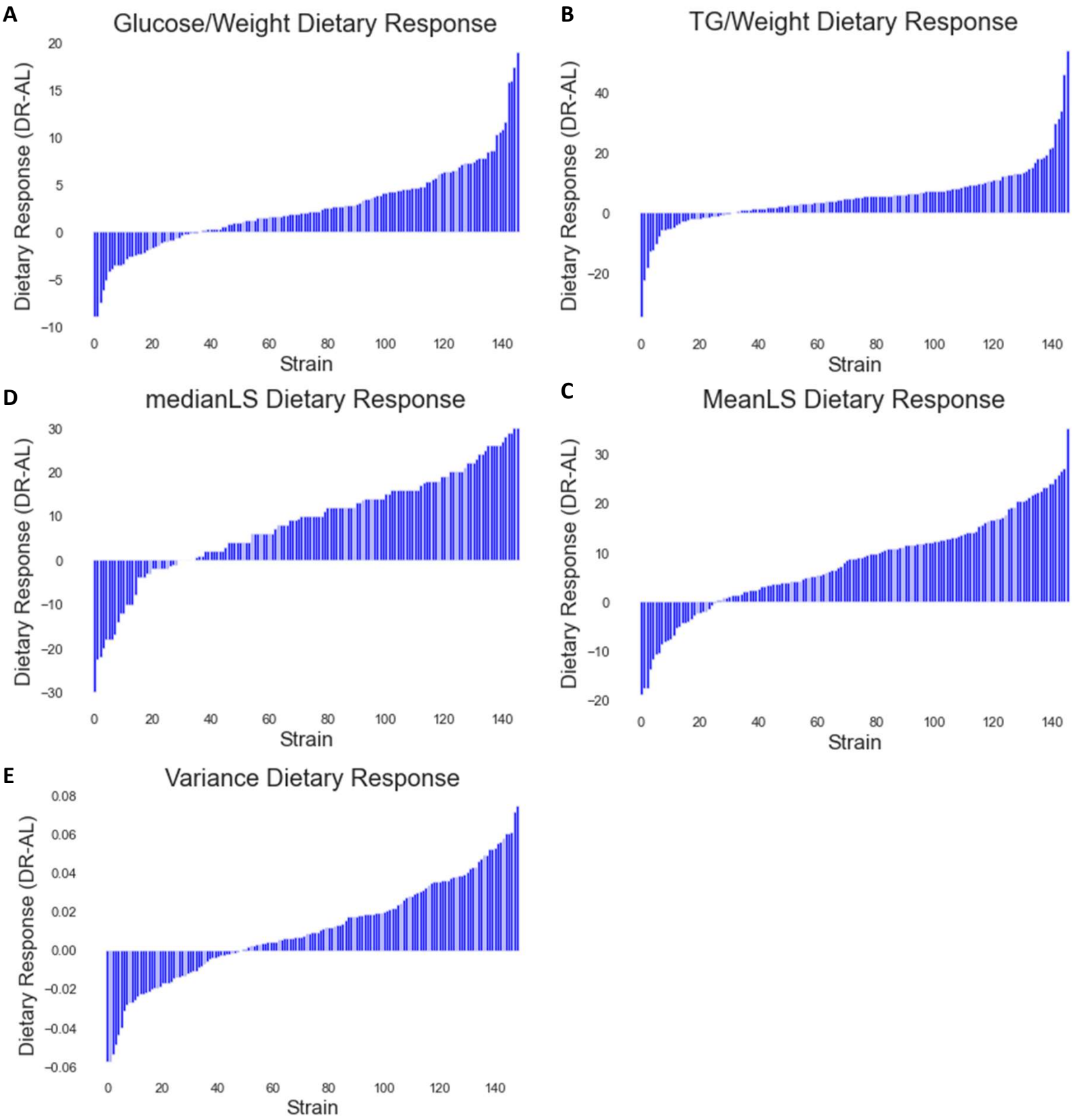


Figure 4: Predictor traits used to build RF lifespan models, and their GWAS gene candidates, grouped similar lifespan response traits in a network diagram of DGRP dietary response.

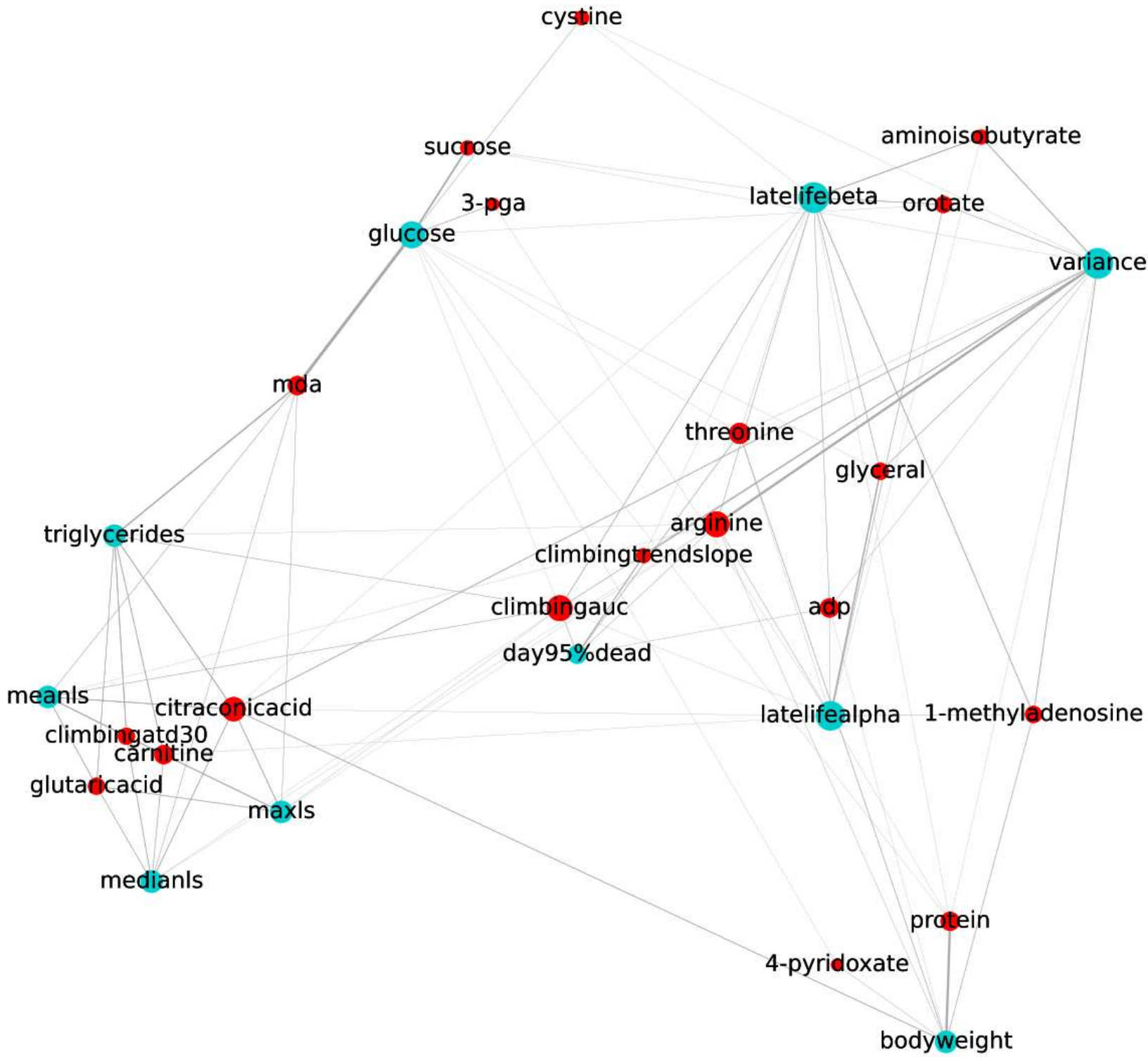
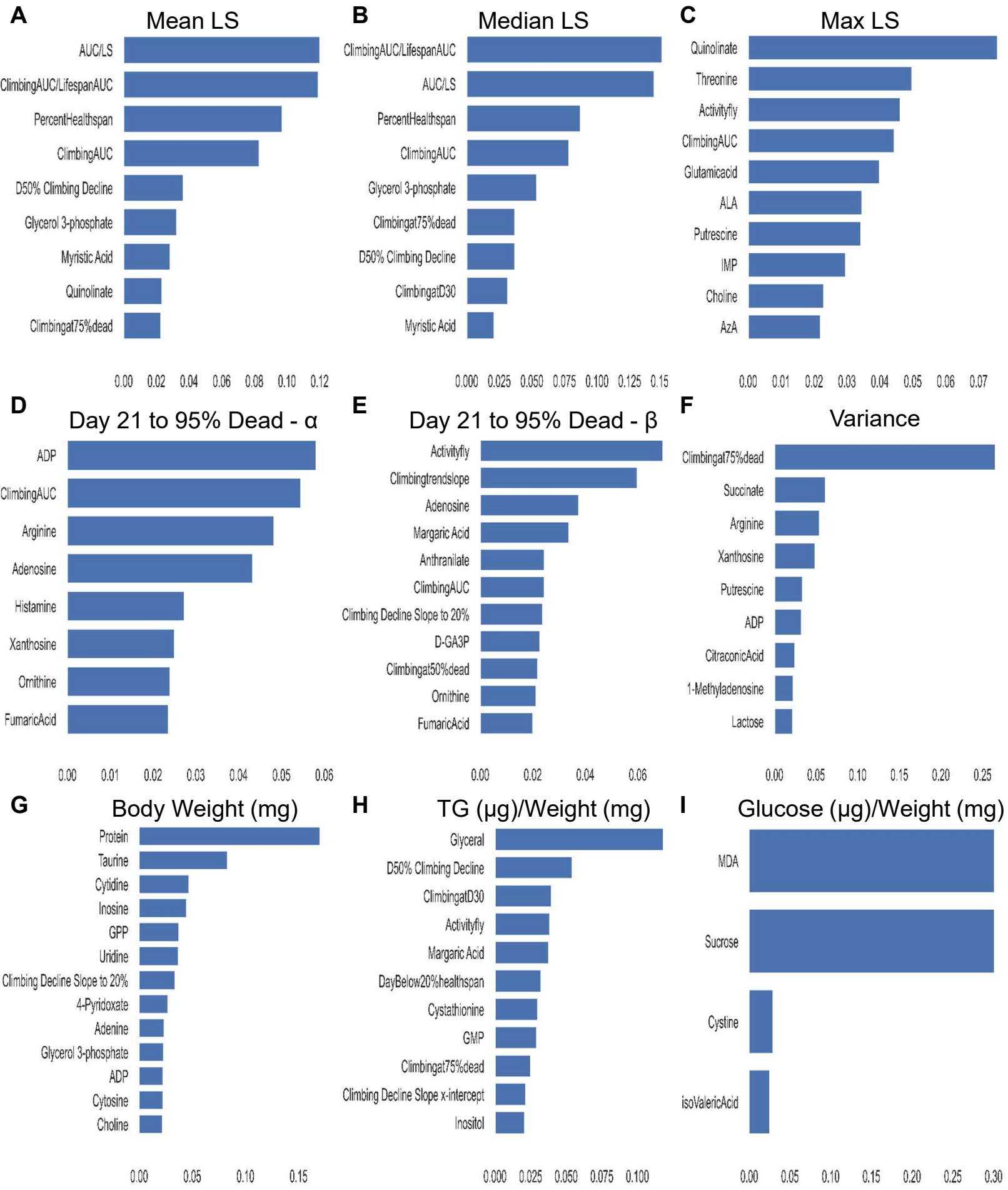



Figure 3: Random Forest models utilize common traits to build models for lifespan.