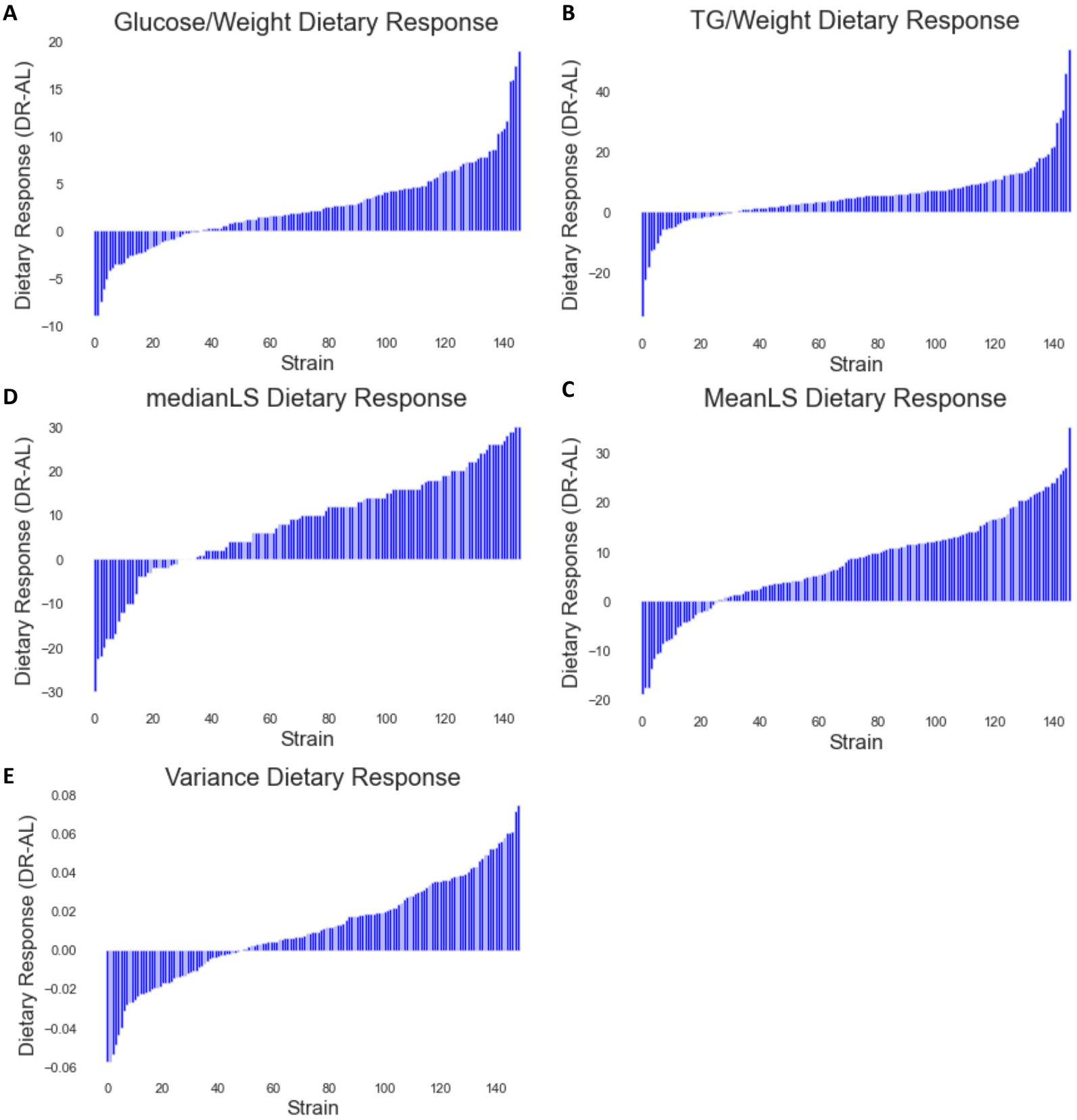
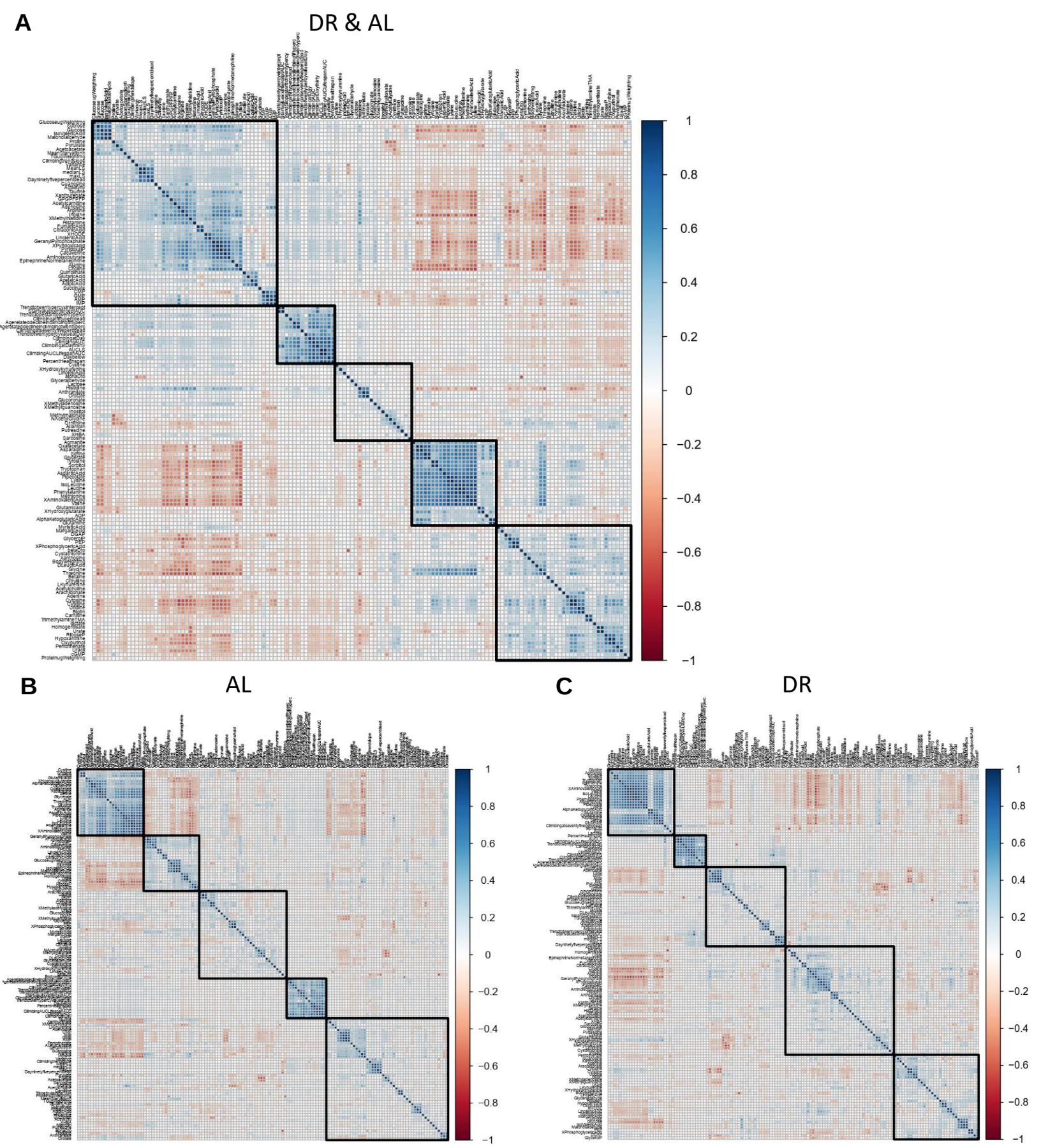
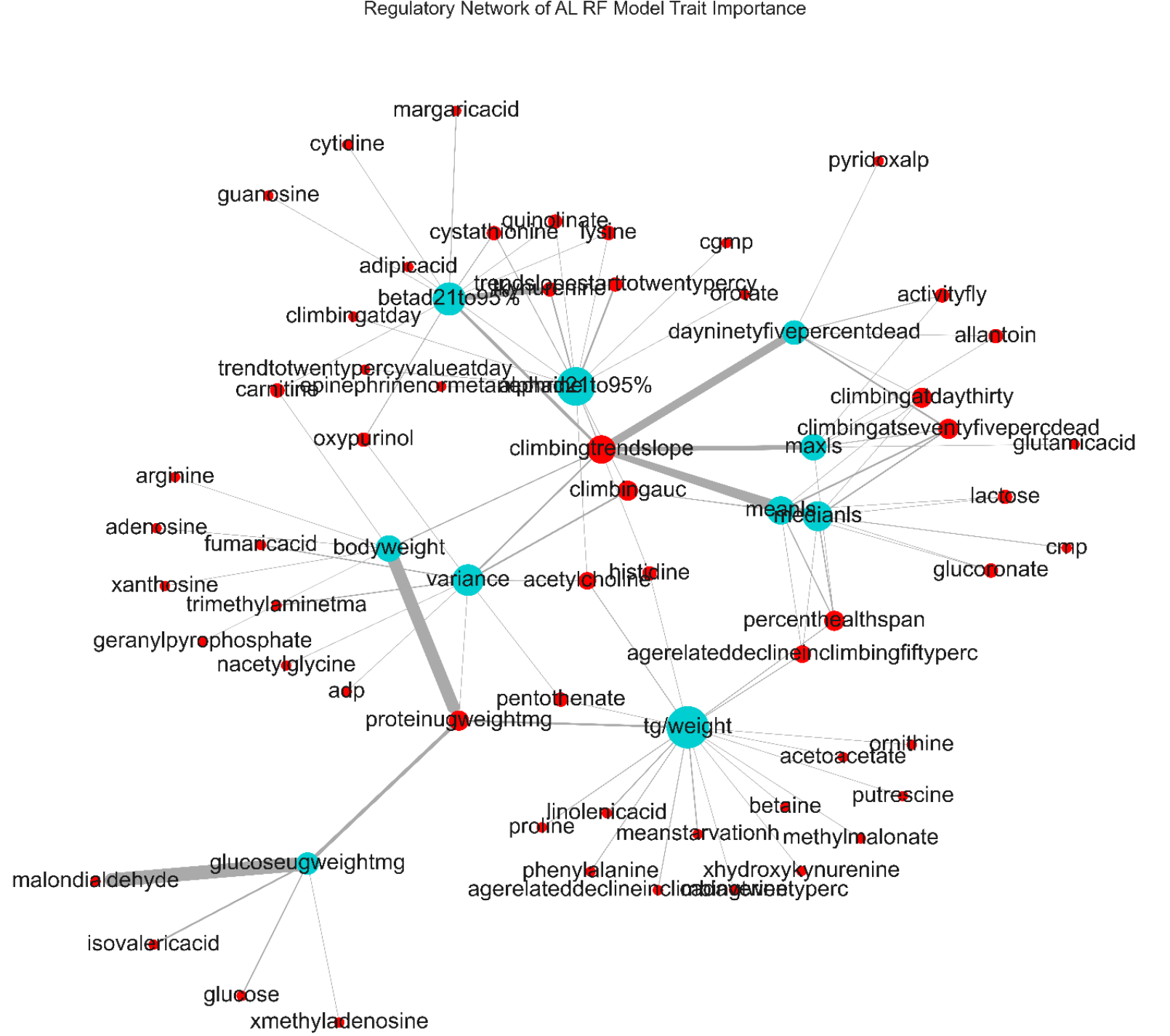
Systems biology approaches identify metabolic signatures of dietary lifespan and healthspan across species



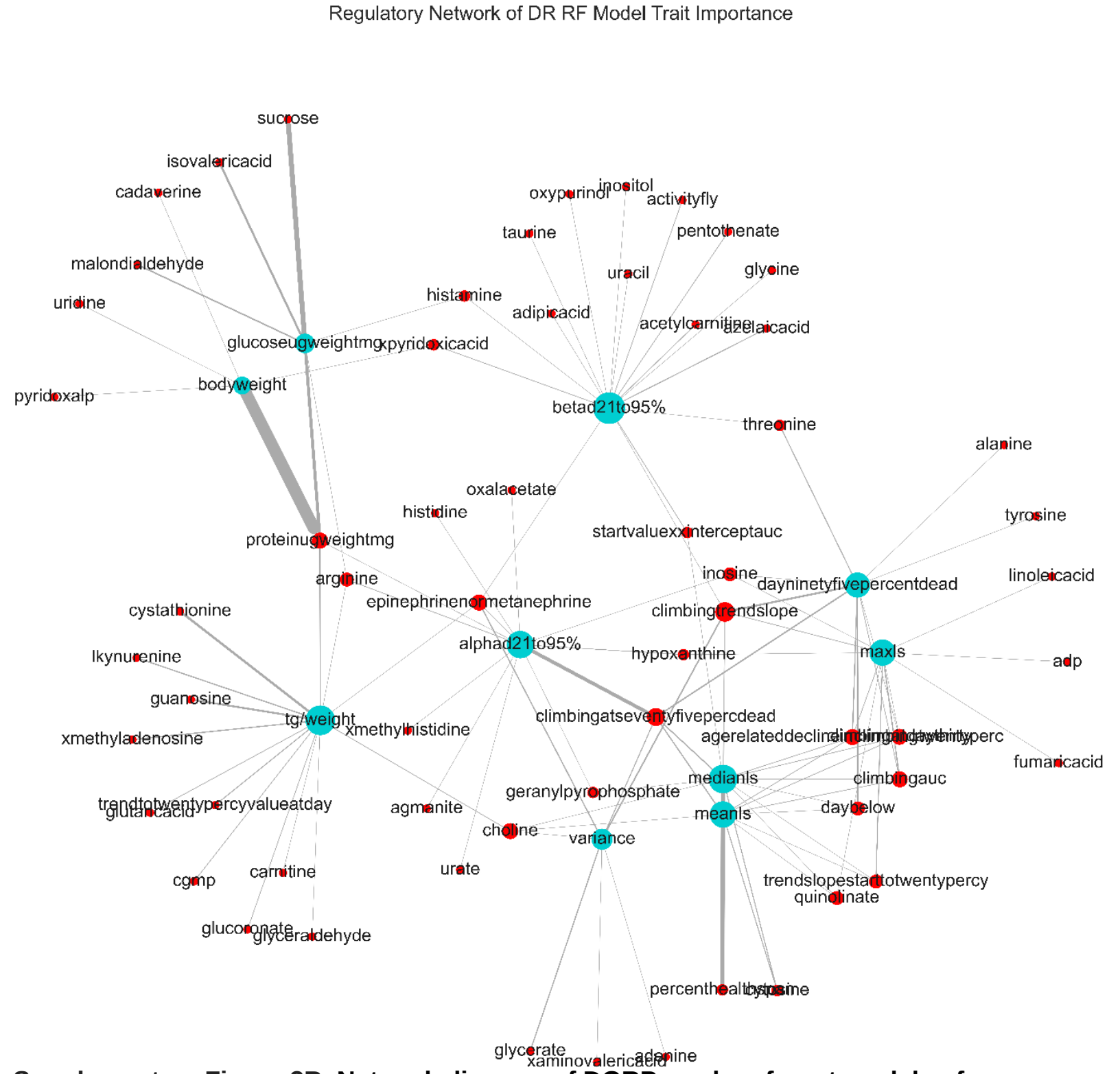
Supplementary Figure 1. Variation of DGRP dietary response in lifespan and healthspan response traits. Dietary response (DR-AL) of 160 DGRP strains in 7 lifespan and healthspan metrics, strains plotted in ascending order.



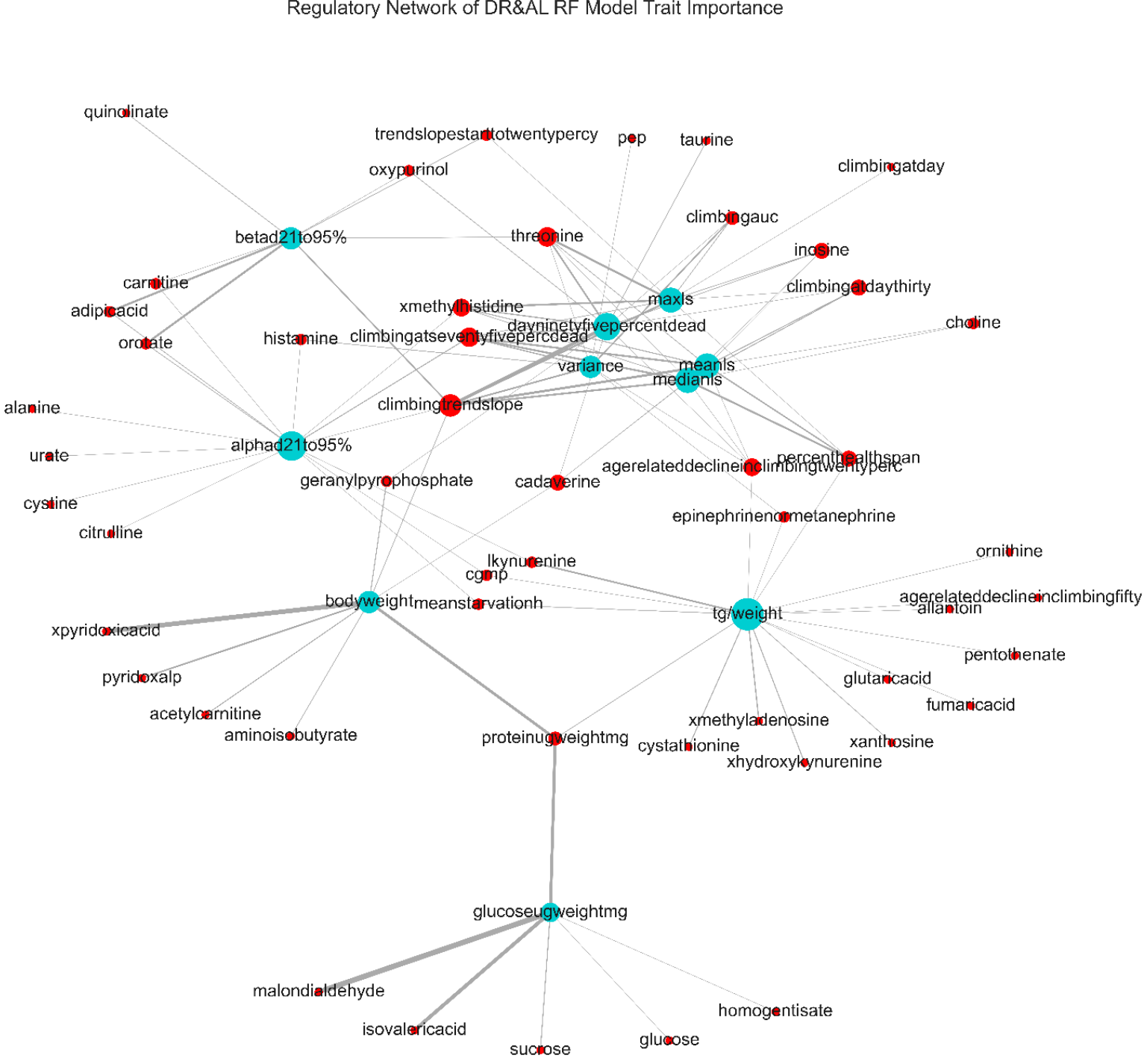
Supplementary Figure 2. DGRP metabolite and phenotype trait correlations on DR and AL, together (A), or separate (B&C). Heatmap of metabolite and phenotype value correlations on DR and AL, together, or separate. Traits are clustered via hierarchical clustering using the Ward method, with 5 clusters highlighted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



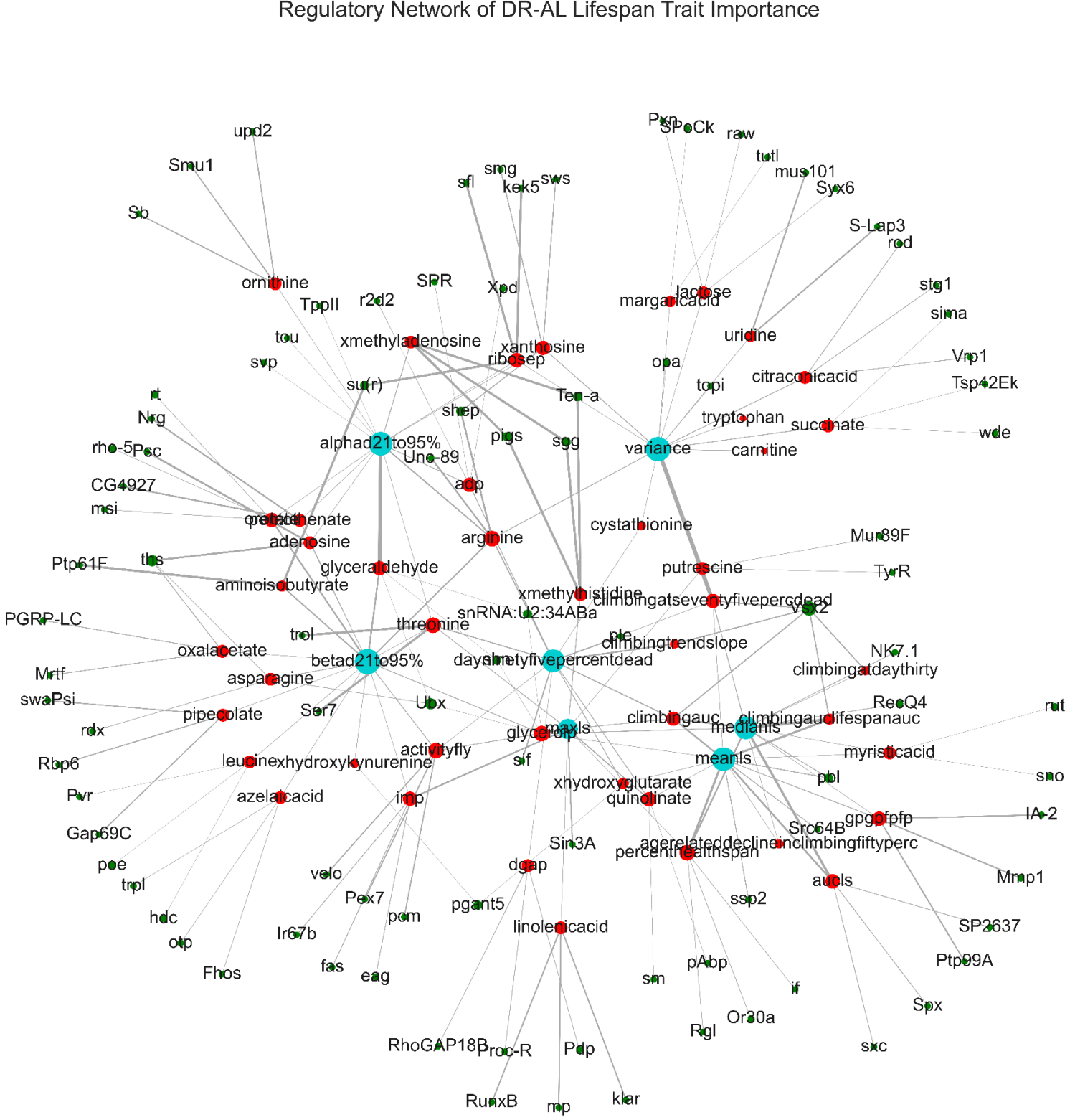
Supplementary Figure 3A. Network diagram of DGRP random forest models of response traits and the predictor traits used to build them; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Network layout is 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, (C) DR & AL, and (D) DR-AL.



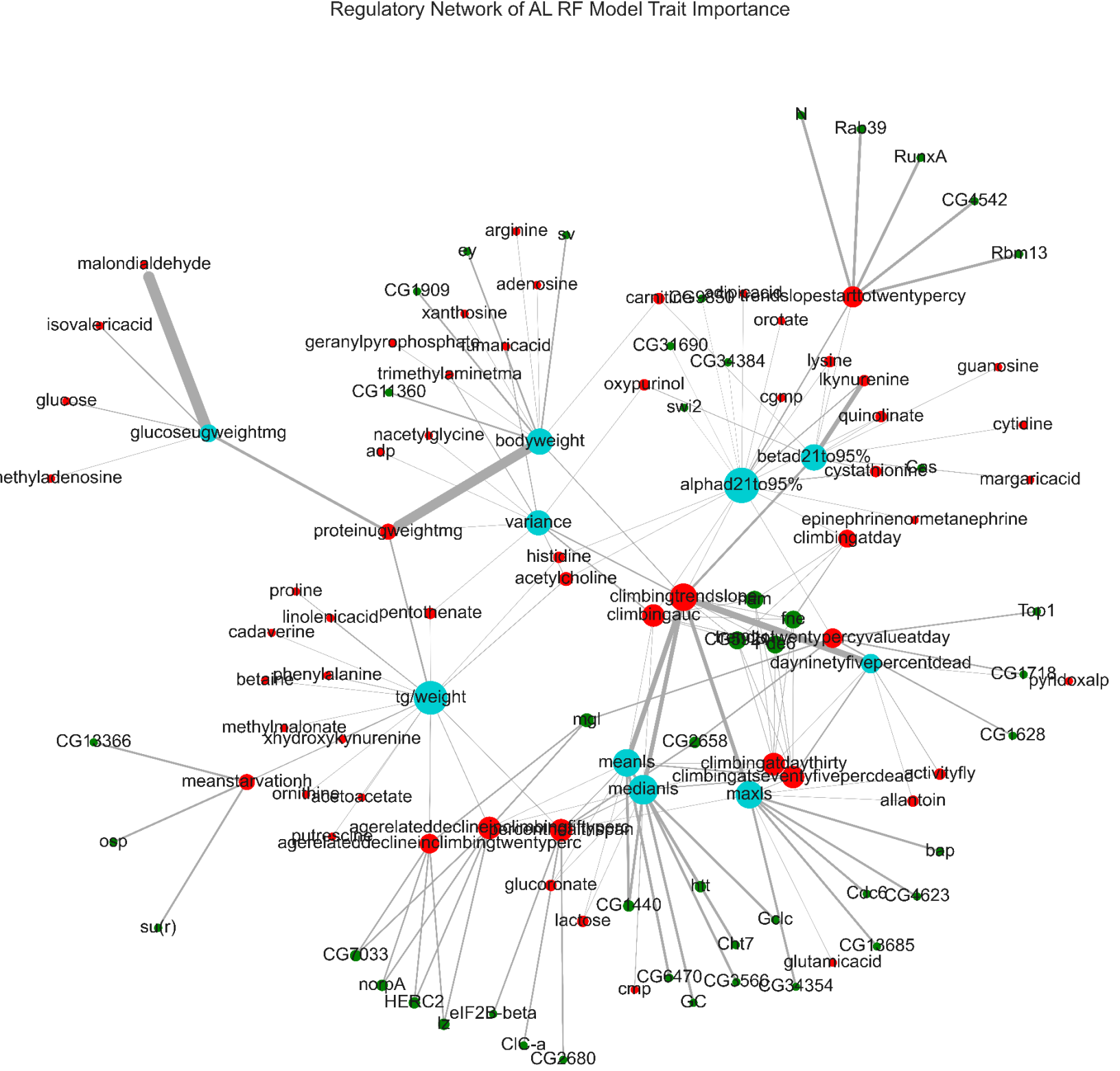
Supplementary Figure 3B. Network diagram of DGRP random forest models of response traits and the predictor traits used to build them; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Network layout is 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, (C) DR & AL, and (D) DR-AL.



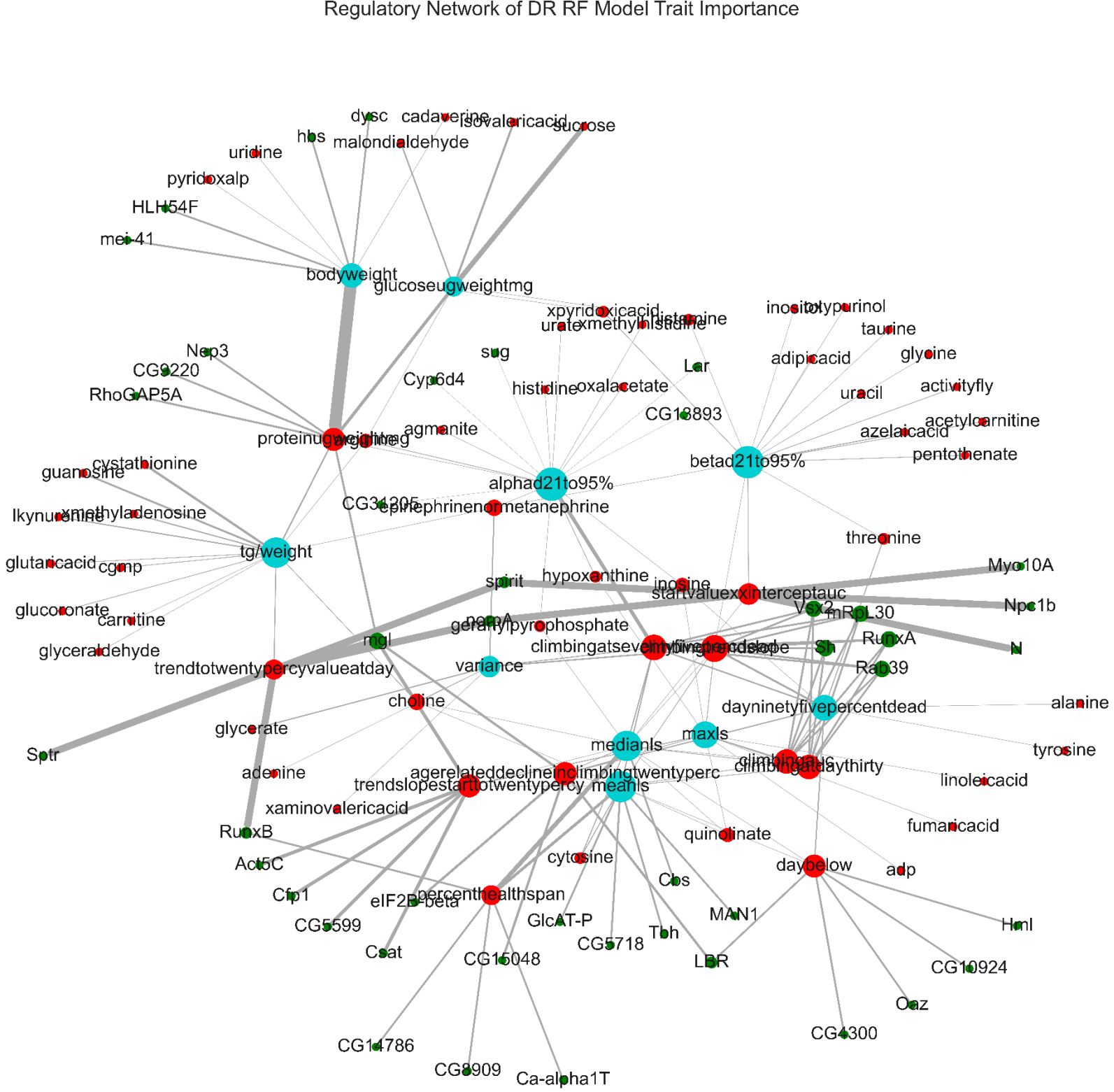
Supplementary Figure 3C. Network diagram of DGRP random forest models of response traits and the predictor traits used to build them; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Network layout is 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, (C) DR & AL, and (D) DR-AL.



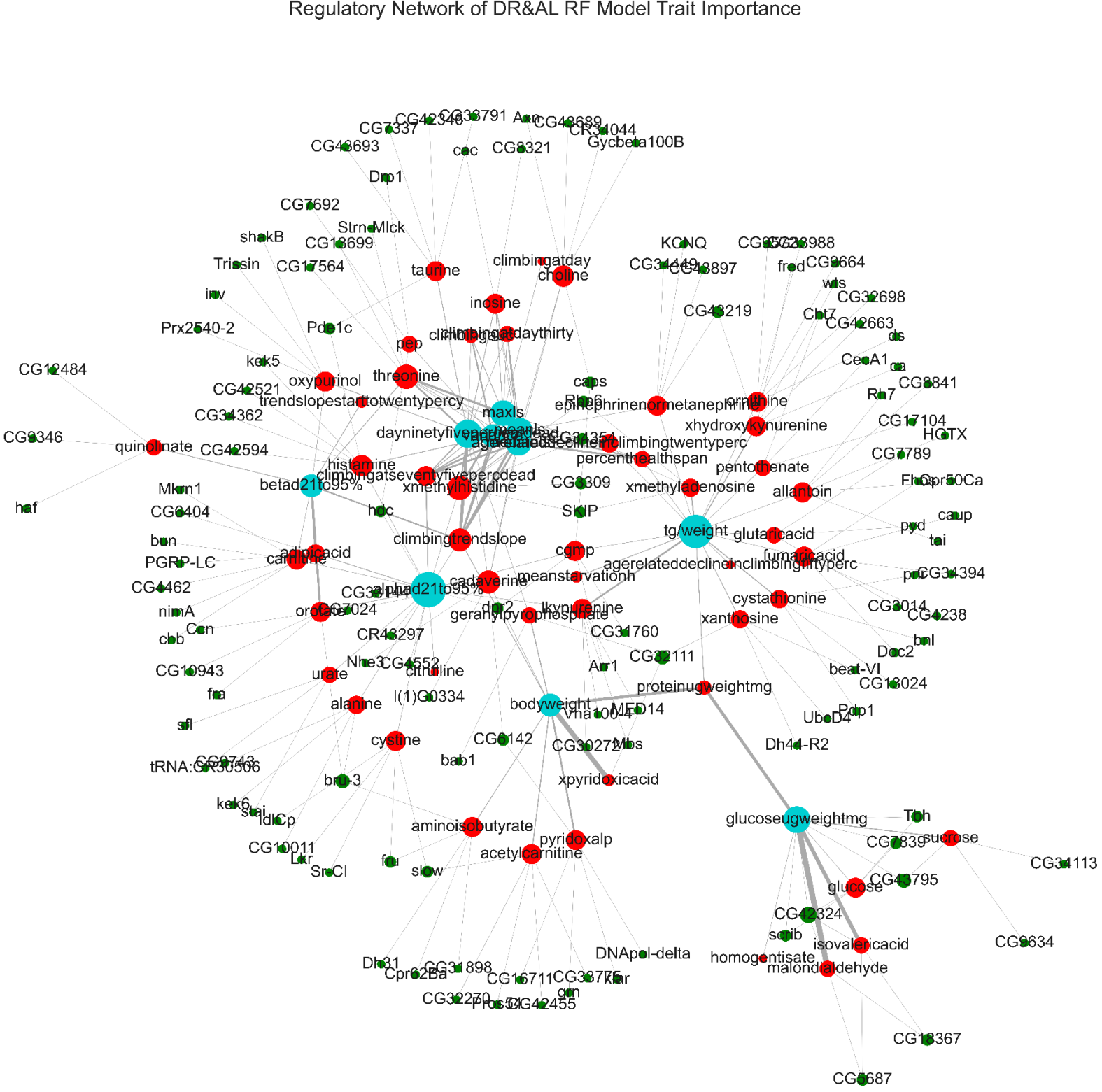
Supplementary Figure 3D. Network diagram of DGRP random forest models of response traits and the predictor traits used to build them; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red nodes) by grey edges whose widths represent their importance. Network layout is 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, (C) DR & AL, and (D) DR-AL.



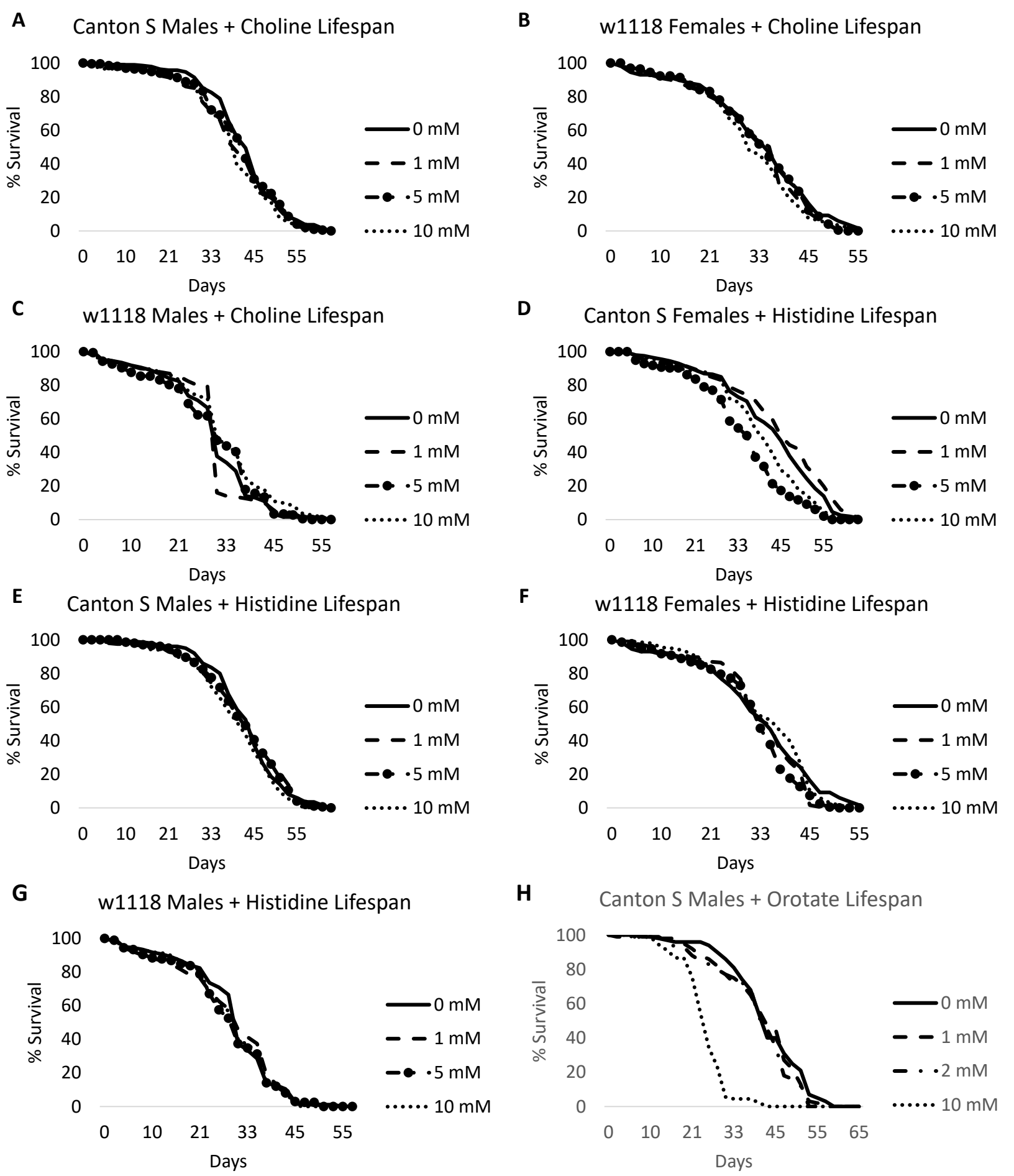
Supplementary Figure 4A. Network diagram of DGRP random forest models of response traits, the predictor traits used to build them, and GWAS candidate genes; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red 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 is 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.



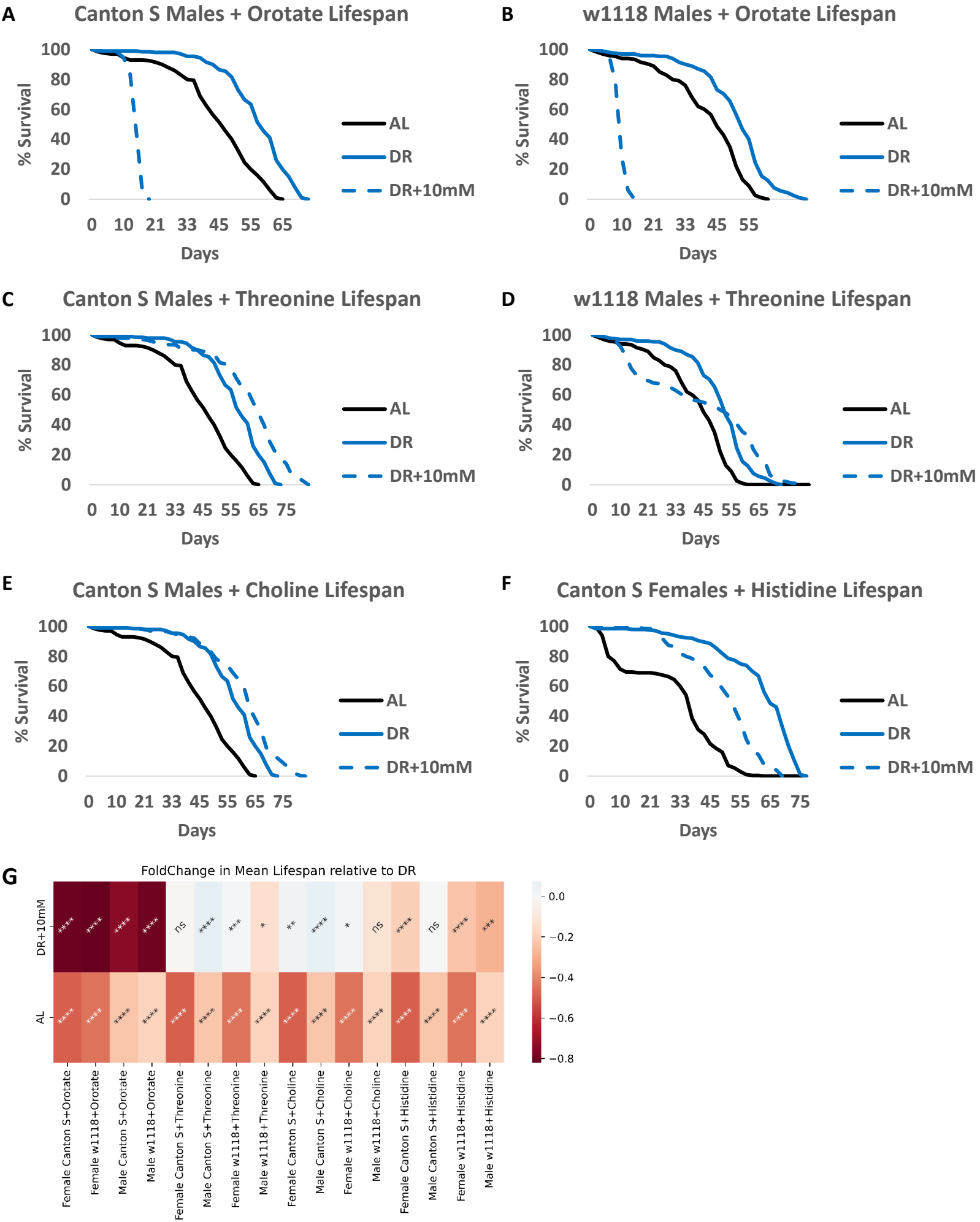
Supplementary Figure 4B. Network diagram of DGRP random forest models of response traits, the predictor traits used to build them, and GWAS candidate genes; separated by diet group. Network diagram of response traits (teal nodes) connected to predictor traits with model importance of 0.02 or higher (red 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 is 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.



Supplementary Figure 4C. 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 is 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.

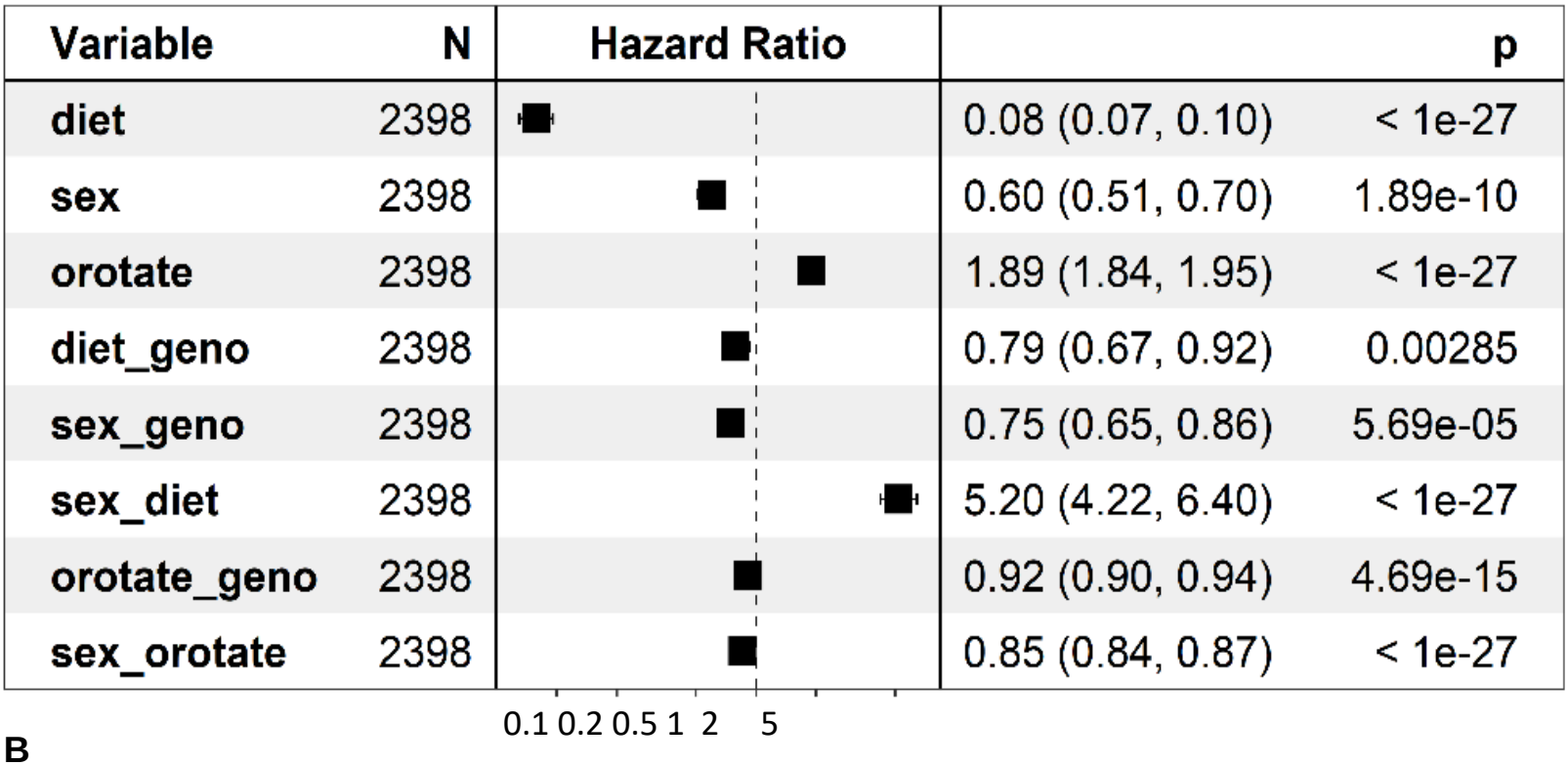

A
Canton S Males + Choline Lifespan
% Survival
Days
0 mM
1 mM
5 mM
10 mM
B
w1118 Females + Choline Lifespan
C
w1118 Males + Choline Lifespan
D
Canton S Females + Histidine Lifespan
E
Canton S Males + Histidine Lifespan
F
w1118 Females + Histidine Lifespan
G
w1118 Males + Histidine Lifespan
H
Canton S Males + Orotate Lifespan
2 mM

Supplementary 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. Orotate and choline supplementation lifespans were repeated twice, and other metabolites once. AL diet is shown with black lines, DR with blue.

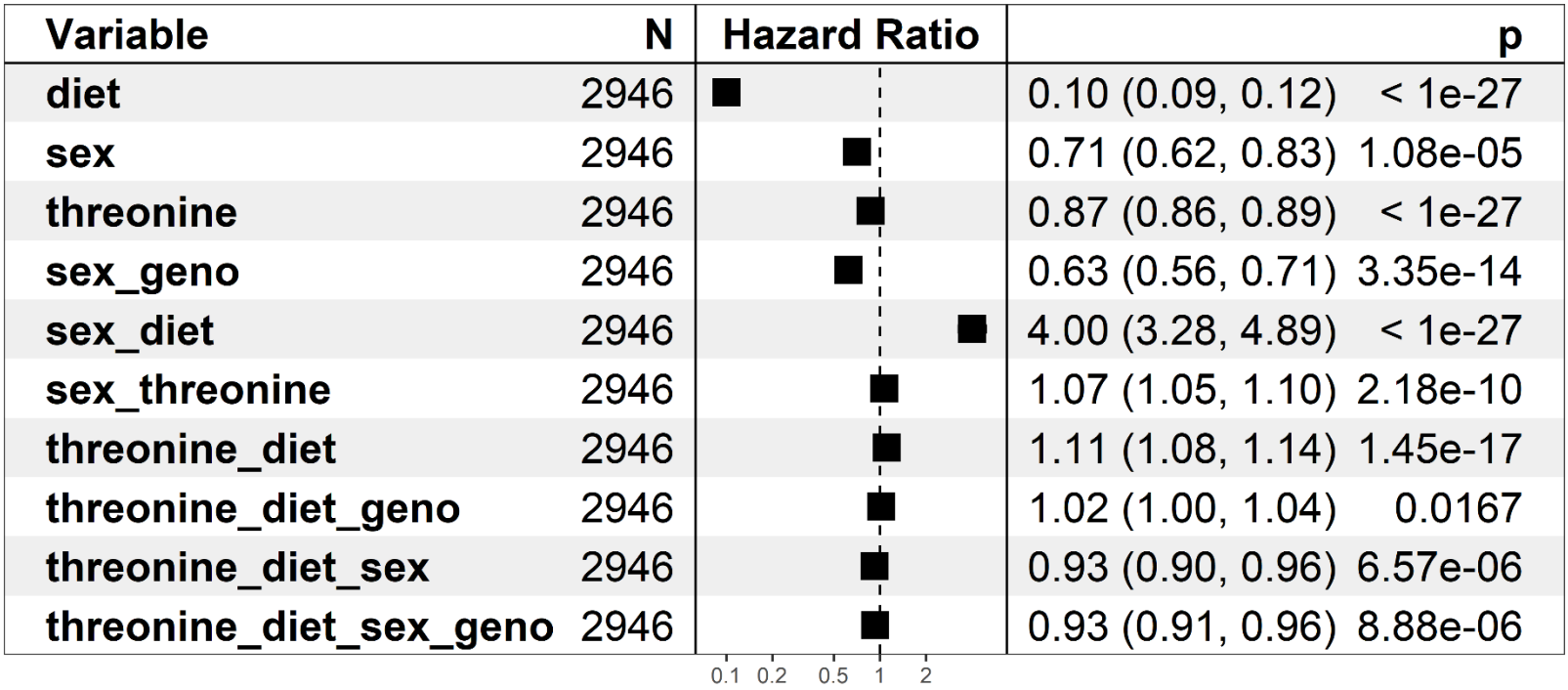

A
Canton S Males + Orotate Lifespan
B
w1118 Males + Orotate Lifespan
C
Canton S Males + Threonine Lifespan
D
w1118 Males + Threonine Lifespan
E
Canton S Males + Choline Lifespan
F
Canton S Females + Histidine Lifespan
% Survival
Days
AL
DR
DR+10mM
G
FoldChange in Mean Lifespan relative to DR
DR+10mM
AL
Female Canton S+Orotate
Female w1118+Orotate
Male Canton S+Orotate
Male w1118+Orotate
Female Canton S+Threonine
Male Canton S+Threonine
Female w1118+Threonine
Male w1118+Threonine
Female Canton S+Choline
Male Canton S+Choline
Female w1118+Choline
Male w1118+Choline
Female Canton S+Histidine
Male Canton S+Histidine
Female w1118+Histidine
Male w1118+Histidine

Supplementary Figure 6. Lifespans of Canton S or w1118 females and males on DR food supplemented with 10mM of orotate, threonine, choline, or histidine. A subset of survival curves for females or males of Drosophila melanogaster strains Canton S or w1118 on a DR diet supplemented with 10 mM of either orotate (A-B), threonine (C-D), choline (E), or histidine (F). (G) Heatmap of mean lifespan fold change values in mean lifespan for males and females of two Drosophila strains on a DR diet supplemented with 10 mM of either orotate, threonine, choline, or histidine, compared to the water-only vehicle control (DR). 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, **** <= 0.00005). Orotate supplementation lifespans repeated twice, and other metabolites once. AL diet is shown with black lines, DR with blue.

A



B



Supplementary Figure 7. Cox proportional hazards analysis including sex as a term. Cox proportional hazards analysis significant results from testing diet (AL as reference), sex (female as reference), orotate (A) or threonine (B) levels (0mM as reference), genotype (w1118 as reference), and their interaction terms in flies. Hazard ratios (HR) and 95% confidence intervals (CI) are shown for each predictor variable. P-values are two-sided and have not been corrected for multiple testing.

STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies[1 2]

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
1	**TITLE and ABSTRACT**	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study	2	We employed two computational and complementary methods across species—random forest modeling within the DGRP as our primary analysis and Mendelian randomization in human cohorts as a secondary analysis. We pinpointed key traits with cross-species relevance as well as underlying heterogeneity and pleiotropy that influence lifespan and healthspan.
	INTRODUCTION			
2	**Background**	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question	4-5	To further validate the therapeutic potential of metabolites derived from machine learning targets in the DGRP, we performed Mendelian randomization (MR) using metabolomic data from individuals in the Twins UK and UK Biobank cohorts. In this situation, MR is more robust to potential confounding and reverse causality which may help estimate the causal effects of metabolites on health and lifespan outcomes. Collectively, evolutionary conserved findings between humans and Drosophila has the potential to shed light on the intricate nature of dietary responses and their modulation by genetic variation.
3	**Objectives**	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects	5	We evaluated the hypothesis that the metabolites identified through a random forest, together with their candidate genes influencing longevity in Drosophila, may also exhibit associations with human health and lifespan. The MR analysis is based on three fundamental assumptions: Firstly, the instrumental variables (IVs) i.e., SNPs must be associated with the metabolites. Secondly, the IVs are not associated with potential confounders. Lastly, IVs must not influence the outcome through metabolites only and not alternate pathways. Collectively, evolutionary conserved findings between humans and Drosophila have the potential to shed light on the intricate nature of dietary responses and their modulation by genetic variation.

	METHODS				
4	**Study design and data sources**		Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:	28-29	Materials and Methods: Supplementary Document 4
		a)	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	28-29	
		b)	Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis	28-29	
		c)	Describe measurement, quality control and selection of genetic variants		
		d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	28-29	
		e)	Provide details of ethics committee approval and participant informed consent, if relevant	29	All summary statistics utilized are from publicly available datasets and did not require ethical review or approval.
5	**Assumptions**		Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis	5	The MR analysis is based on three fundamental assumptions: Firstly, the instrumental variables (IVs) i.e., SNPs must be associated with the metabolites. Secondly, the IVs are not associated with potential confounders. Lastly, IVs must not influence the outcome through metabolites only and not alternate pathways.
6	**Statistical methods: main analysis**		Describe statistical methods and statistics used	30-31	
		a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)	30	
		b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	30	
		c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples	30	
		d)	Explain how missing data were addressed		NA
		e)	If applicable, indicate how multiple testing was addressed		NA

7	**Assessment of assumptions**	Describe any methods or prior knowledge used to assess the assumptions or justify their validity		NA
8	**Sensitivity analyses and additional analyses**	Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)	31	
9	**Software and pre-registration**			
	a)	Name statistical software and package(s), including version and settings used	31	
	b)	State whether the study protocol and details were pre-registered (as well as when and where)	NA	
	RESULTS			
10	**Descriptive data**		28-29	
	a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram		
	b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)		
	c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	NA	
	d)	For two-sample MR: i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples ii. Provide information on the number of individuals who overlap between the exposure and outcome studies		
11	**Main results**		13-14	
	a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale		
	b)	Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference		
	c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		

		d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	
12	**Assessment of assumptions**		13-14	
		a)	Report the assessment of the validity of the assumptions	
		b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value)	
13	**Sensitivity analyses and additional analyses**		13-14	
		a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	
		b)	Report results from other sensitivity analyses or additional analyses	
		c)	Report any assessment of direction of causal relationship (e.g., bidirectional MR)	
		d)	When relevant, report and compare with estimates from non-MR analyses	
		e)	Consider additional plots to visualize results (e.g., leave-one-out analyses)	
	DISCUSSION			
14	**Key results**		Summarize key results with reference to study objectives	19-20
15	**Limitations**		Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them	21
16	**Interpretation**			19-21
		a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	
		b)	Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions	
		c)	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	

17	**Generalizability**	Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure	21
	OTHER INFORMATION		
18	**Funding**	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based	22
19	**Data and data sharing**	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where	30
20	**Conflicts of Interest**	All authors should declare all potential conflicts of interest	22

1. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) Statement. JAMA. 2021;under review.

2. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration. BMJ. 2021;375:n2233.