nature communications

Article

https://doi.org/10.1038/s41467-024-52909-y

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

Received: 11 April 2024

Accepted: 18 September 2024

Published online: 29 October 2024

Check for updates

Tyler A. U. Hilsabeck[1,2,3,11], Vikram P. Narayan[1,4,11], Kenneth A. Wilson[1,2], Enrique M. Carrera[1,5], Daniel Raftery[6], Daniel Promislow[7,8,9], Rachel B. Brem[1,2,10], Judith Campisi[1] & Pankaj Kapahi[1,2]

Dietary restriction (DR) is a potent method to enhance lifespan and healthspan, but individual responses are influenced by genetic variations. Understanding how metabolism-related genetic differences impact longevity and healthspan are unclear. To investigate this, we used metabolites as markers to reveal how different genotypes respond to diet to influence longevity and healthspan traits. We analyzed data from *Drosophila* Genetic Reference Panel (DGRP) strains raised under AL and DR conditions, combining metabolomic, phenotypic, and genome-wide information. 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. Notably, orotate was linked to parental age at death in humans and blocked the DR lifespan extension in flies, while threonine supplementation extended lifespan, in a strain- and sex-specific manner. Thus, utilizing natural genetic variation data from flies and humans, we employed a systems biology approach to elucidate potential therapeutic pathways and metabolomic targets for diet-dependent changes in lifespan and healthspan.

Aging is the leading cause of morbidity and mortality in most developed countries. Dietary restriction (DR), restricting specific nutrients or total calories without causing malnutrition, has been demonstrated as a robustly conserved means to extend lifespan and health span. 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, variation 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

[1]Buck Institute for Research on Aging, Novato, CA 94945, USA. [2]Davis School of Gerontology, University of Southern California, University Park, University Park, Los Angeles, CA 90089, USA. [3]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [4]Department of Biology & Chemistry, Embry-Riddle Aeronautical University, Prescott, AZ 86301, USA. [5]Dominican University of California, San Rafael, CA 94901, USA. [6]Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA. [7]Department of Pathology, University of Washington, Seattle, WA 98195, USA. [8]Department of Biology, University of Washington, Seattle, WA 98195, USA. [9]Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, USA. [10]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. [11]These authors contributed equally: Tyler A. U. Hilsabeck, Vikram P. Narayan. e-mail: pkapahi@buckinstitute.org

metabolites that are regulated by genetic variation to influence longevity under DR[4]. Nevertheless, in 2020, Jin and colleagues highlighted the challenge of establishing a correlation between lifespan and metabolite traits. This difficulty arises from the strong association between the DR response in mean lifespan and the metabolite traits observed in flies following the standard diet (AL). To avoid the confounding effects of an AL diet on the DR response, Jin, et al.[4] used the residuals from a simple regression to identify connections between metabolites and lifespan. Meanwhile, another study conducted by Wilson, et al.[2] highlighted a distinct absence of correlation between healthspan traits and lifespan traits, revealing disparate genetic regulators for each facet. 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 exploring the relationship between health and longevity is to incorporate convergent data from multiple species and multiple genotypes within each species. Research conducted using model organisms has identified many candidate genes and pathways responsible for the aging process[5,6]. Some of these genes and processes belong to evolutionarily conserved nutrient-sensing pathways, which suggests they may be relevant in humans as well. However, several large genome-wide association studies (GWAS) of human longevity show a surprising lack of association with lifespan-extending genes identified in model organisms[7]. Despite this overall disconnect, there are genes that have repeatedly shown an association with lifespan, such as FOXO3 and APOE[8,9]. The disconnect might, in part, arise from the effects of natural variation in a lifespan-associated gene, highlighting the importance of studying natural variation in model systems. An effective approach to investigate this variation across multiple species involves utilizing the metabolome as indicators of the pathways utilized in an individual's dietary response, akin to footprints. While utilizing the metabolome has previously been done, as mentioned above, a fresh approach was required to establish a conserved connection between variations in metabolites with health span and lifespan.

Machine learning, which leverages well-established high-dimensional datasets to create more accurate predictive models for specific phenotypes, has proven valuable in discovering novel factors impacting a range of biological traits, including those related to longevity and health[10]. However, it remains to be seen how the combined utilization of metabolomic and phenotypic indicators can transcend species and genotypes in response to DR, enabling the prediction of specific elements contributing to various health and longevity attributes.

In previous work, we utilized the *Drosophila* Genetic Reference Panel (DGRP) to explore the effects of natural genetic variation on DR response in metabolic phenotypes[11], the metabolome[4], and lifespan and healthspan. Across these datasets, we identified that no single metric could reliably predict lifespan, suggesting that multiple factors collectively influence the longevity outcomes of DR. Here, we use this data to probe the interplay between metabolomic, metabolic, and healthspan-related traits and lifespan. Employing random forest modeling, we pinpoint the relevance of metabolites orotate, threonine, and choline in constructing multiple lifespan models. This modeling approach circumvents potential issues arising from standard diet (AL) traits influencing DR response due to its bagging methodology[12–14]. Additionally, to identify genes that regulate multiple traits, we incorporated previously published GWAS for all traits to identify genetic candidates that may impact these associations, including *Src64B*, which demonstrates an association with mean lifespan. This study represents a pioneering effort, being the first to leverage genetic data from 174 strains, encompassing 34 distinct phenotypes and 111 measured metabolites within those strains, while also investigating targets within a human dataset.

To determine 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. The MR analysis is based on three fundamental assumptions: Firstly, the instrumental variables (IVs), i.e., single nucleotide polymorphisms (SNPs), must be valid proxies associated with the metabolites. Secondly, the IVs must not be associated with potential confounders or alternate pathways. Lastly, IVs must not influence the outcome directly but act only through the metabolites. 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. 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. Our findings uncover factors contributing to lifespan determination under DR (Fig. 1A). This approach has led us to uncover pathways that could serve as valuable biomarkers of health and potential therapeutic targets for augmenting both human lifespan and healthspan.

Results

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

We combined and reanalyzed data from previously published DGRP metabolite and phenotype datasets, including Nelson et al.[11], Jin[4], and Wilson et al.[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 Data 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 Data 1 and 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 Data 1 and 2). PCA of this imputed list clustered by diet and was 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 the value on DR, "DR-AL"). In general, DGRP strains responded to a DR diet similar to what has been seen before, with over 93% of strains having decreased body weight and at least 78% having an increase in most lifespan response metrics (Fig. 1B–E and Supplementary Fig 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.

Correlation of dietary response in DGRP strains identifies few metabolites that correlate with lifespan- and healthspan-related phenotypes

We looked for correlations between our predictor traits in each diet and the dietary response (AL, DR, and DR-AL), and found few correlations across data sets, with the majority of these being weak correlations (DR-AL, Fig. 2). Just as observed by Jin et al. in 2020, we also noted a moderate correlation ($r = 0.38$, $P = 2.14 \times 10^{-6}$) in median

https://doi.org/10.1038/s41467-024-52909-y

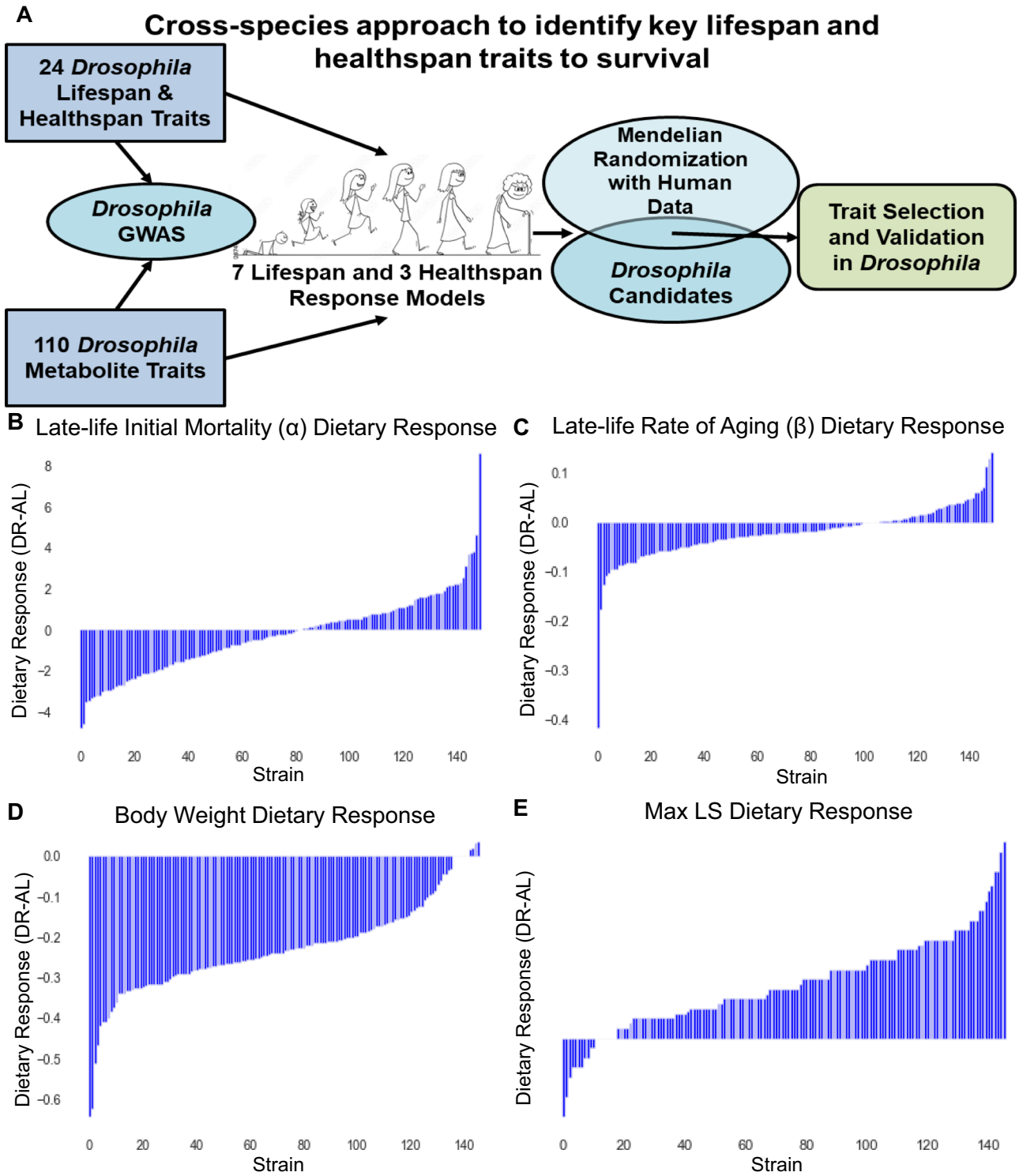


Fig. 1 | Diet influences lifespan and healthspan in a genotype-specific manner. DGRP shows strong strain-specific responses to diet (**A**) Graphical workflow for modeling data from *Drosophila* and filtering traits using Mendelian randomization of human data. **B** DGRP strain traits on a high yeast (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.

lifespan when comparing our "DR-AL" and AL datasets, and similar low to high correlations with the other lifespan metrics. We performed hierarchical clustering using 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 Fig. 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

https://doi.org/10.1038/s41467-024-52909-y

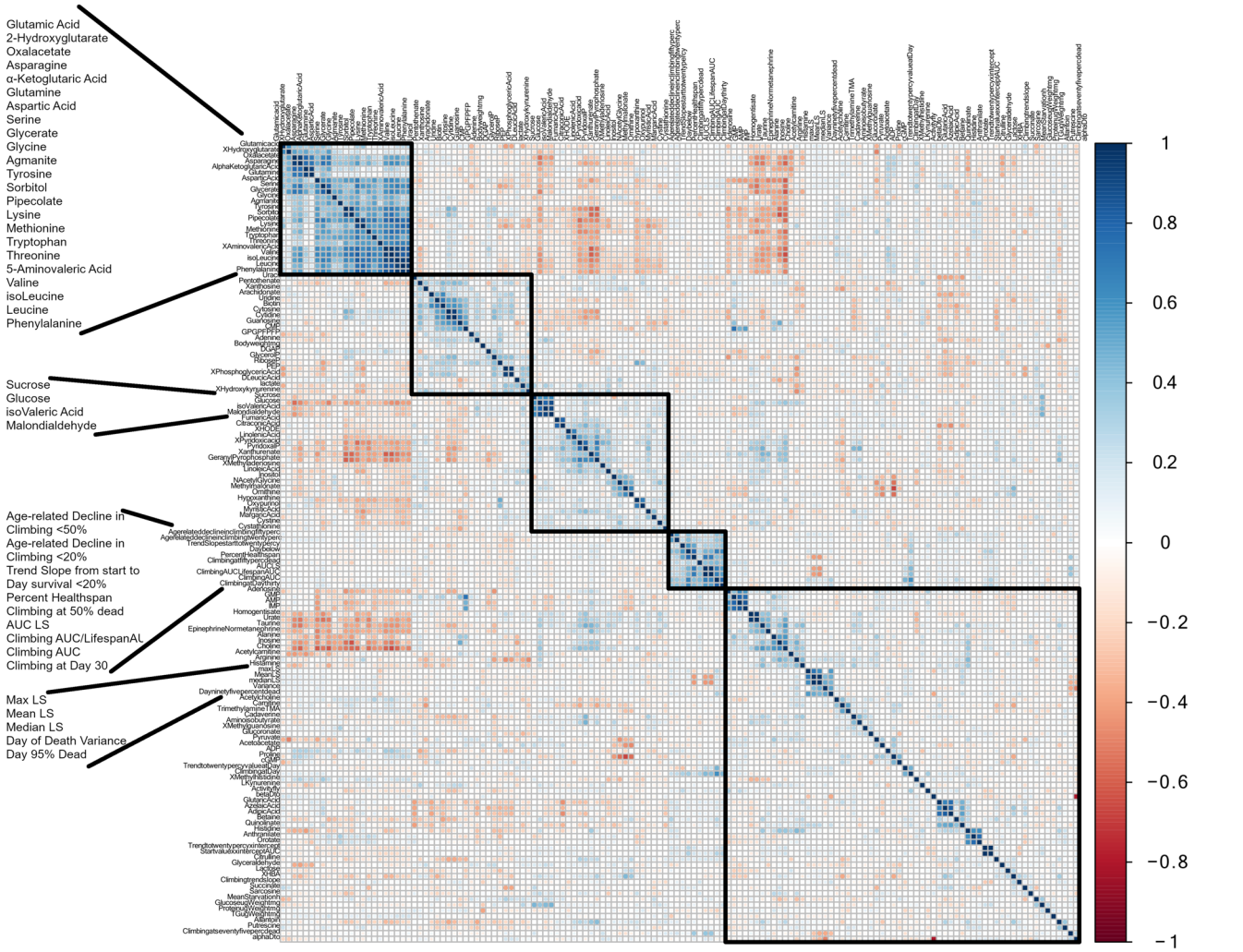


Fig. 2 | DGRP trait dietary responses correlate with similar traits, but few metabolites correlate with phenotypes. Heatmap of metabolite and phenotype dietary response correlations. Traits are clustered via hierarchical clustering using the Ward method, with five clusters highlighted on the diagonal.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

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 a different method for identifying factors that could predict and/or "build" these response traits.

Random forest modeling identified metabolite and phenotype traits used to build multiple 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 (Fig. 3, Supplementary Data 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 (Fig. 3A–F, predictor traits for the day 95% were dead dietary response model are not shown and can be found on the DR-AL sheet of Supplementary Data 3). 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 the 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 (Fig. 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.

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[15,16]. These networks were further augmented by the addition of top candidate genes from previously published GWAS

https://doi.org/10.1038/s41467-024-52909-y

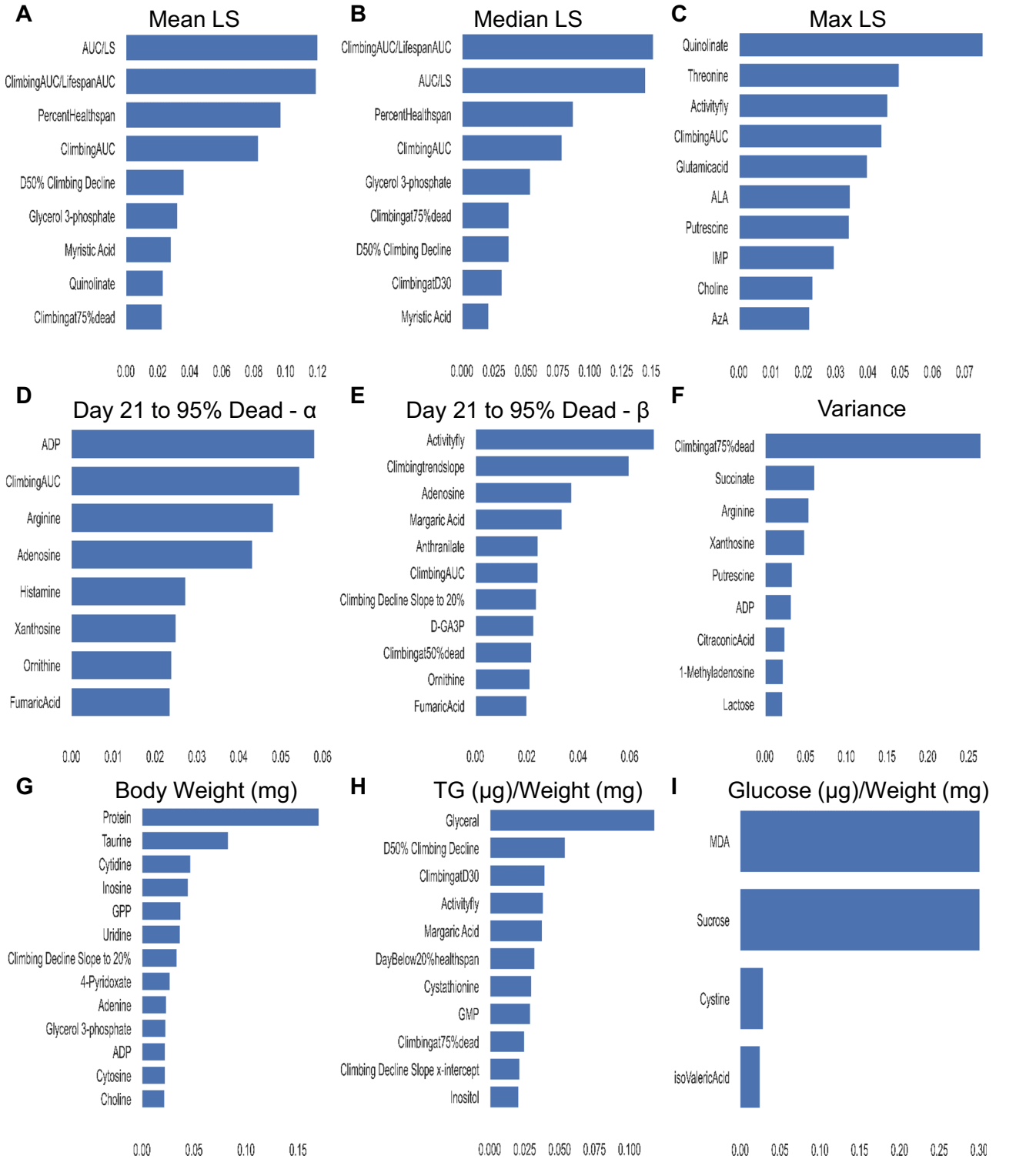


Fig. 3 | Random Forest models utilize common traits to build models for lifespan. Predictor traits with model importance of 0.02 or higher used to build Random Forest models of DGRP trait dietary response of (**A**) mean lifespan, **B** median lifespan, **C** maximum lifespan, **D** initial mortality (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). Predictor traits for the day 95% were dead dietary response model are not shown and can be found on the DR-AL sheet of Supplementary Data 3.

candidate genes of each trait, based on *p*-value[2,4,11]. As expected, the lifespan response traits were grouped and had a few common predictors and genes, specifically late-life α and late-life β, and mean and median lifespans had significant overlap with each other (Fig. 4, Supplementary Fig. 4A, and Supplementary Data 3). For example, myristic acid, glycerol phosphate, and the gene *pbl*, were included in the random forest models for mean and median lifespan or were GWAS candidates, 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

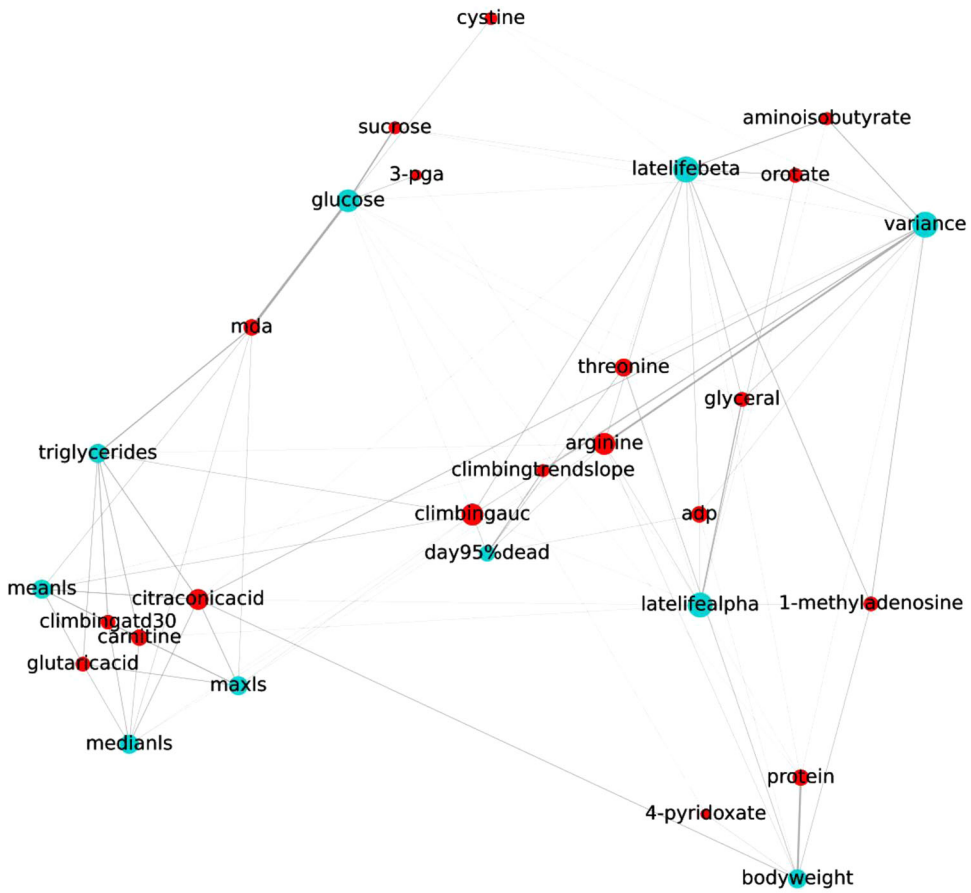


Fig. 4 | Predictor traits used to build RF lifespan models grouped similar lifespan and healthspan response traits in a network diagram of dietary response. A network diagram of lifespan and healthspan response traits (teal nodes) connected to their top 3 predictor traits with model importance of 0.02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

AL group (Supplementary Figs. 3A and 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 Figs. 3 and 4).

Estimates of causal effects of metabolites on health and lifespan-related traits

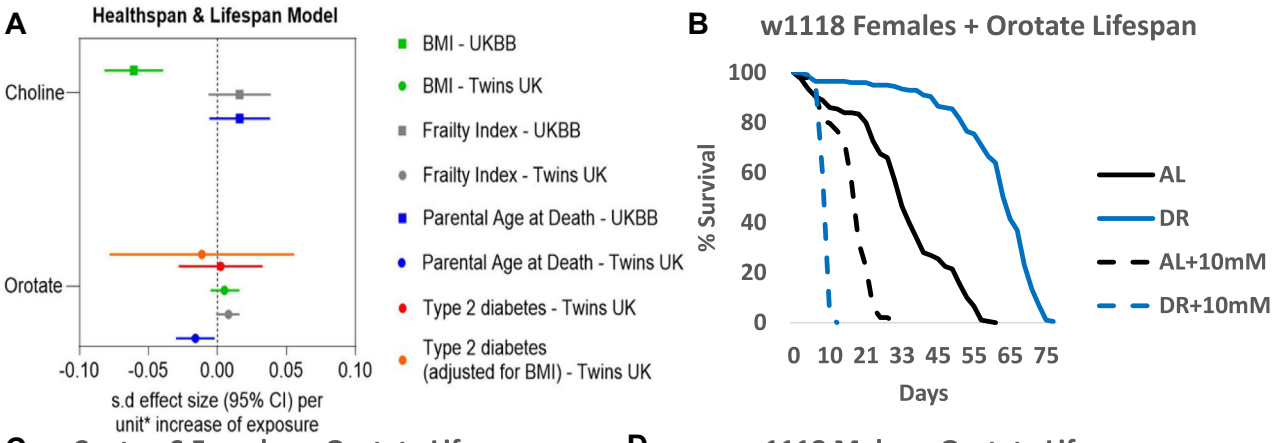
We next used the networks developed above to identify multiple response models, with traits identified to be relevant to humans via MR and selected for validation in *Drosophila*. We explored the utility of the identified metabolite associations from the random forest model approach by applying them in two-sample MR using the IVW method as the primary analysis. We identified 8 common metabolites in *Drosophila* that were also detected in available GWAS for serum metabolites in humans and influence multiple lifespan- and health-span-related traits (Fig. 5A).

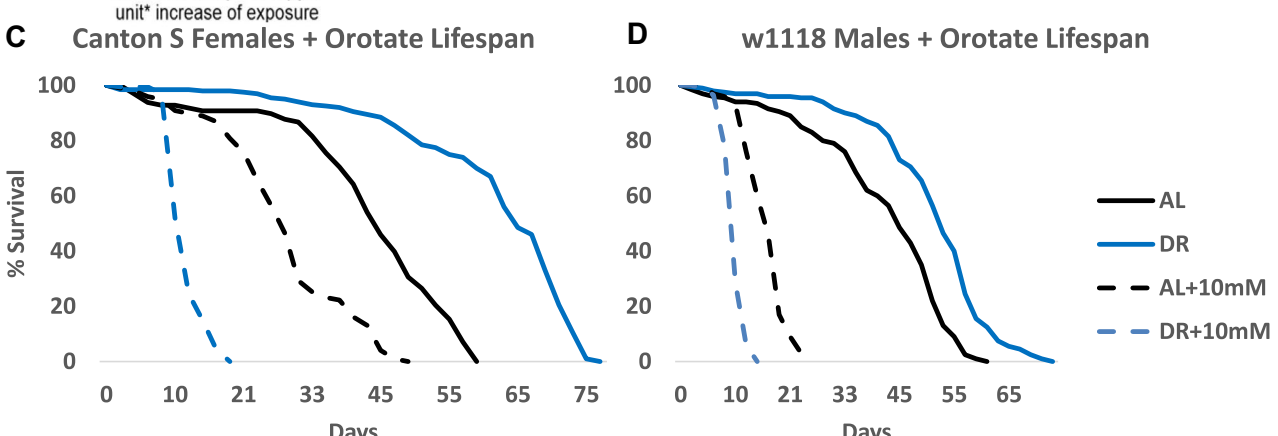
Orotate, one of two metabolites implicated in both healthspan, and lifespan random forest models was also significantly associated with parental longevity (combined parental age at death) using the standard IVW analyses ($\beta_{IVW} = -0.02$; 95% CI, −0.03 to −0.002; $P = 0.03$; Fig. 5A). Though we were unable to carry out sensitivity analyses and test for pleiotropic effects of orotate on longevity due to an insufficient number of SNPs, the Cochran's Q statistics showed no evidence of heterogeneity ($Q = 2.93$, $P = 0.09$). The second metabolite, choline though not significantly associated with parental longevity, was significantly associated with a decrease in BMI ($\beta_{IVW} = -0.06$; 95% CI, −0.08 to −0.03; $P = 2.17 \times 10^{-08}$) which may explain the effect on longevity observed in the random forest model and survival assays in *Drosophila*. Similar findings were observed using the weighted-median estimator ($\beta_{\text{weighted median}} = -0.04$; 95% CI, −0.06 to −0.007; $P = 1.19 \times 10^{-02}$). Cochran's Q statistics showed no evidence of heterogeneity in the IVW analyses ($P = 0.12$). The MR-PRESSO global test ($P = 0.12$) but not the intercept of the MR-Egger method ($P = 0.03$) confirmed the absence of pleiotropy in the IVW estimate. Full MR results can be found in Supplementary Data 4.

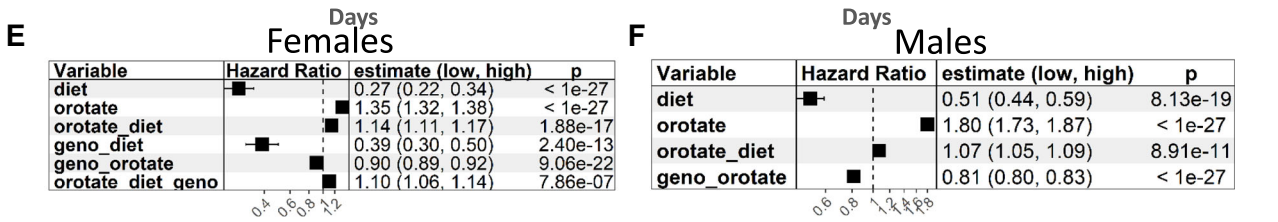
Orotate's lifespan-shortening effect is diet-independent, but the mechanism is strain- and sex-specific

To validate the metabolites that were used to both build Random Forest models of lifespan and healthspan response and were found to be relevant for humans via MR, we supplemented our AL or DR diets with candidate metabolites (Fig. 5A). Validated metabolites were selected based on the number of lifespan models they were incorporated in, the average odds-ratio from MR, and the novelty of the lifespan association (Supplementary Data 5). Orotate supplementation significantly shortened lifespan at 10 mM in both strains and sexes (*w1118* females (Fig. 5B), *Canton S* females (Fig. 5C), *w1118* males (Fig. 5D), and *Canton S* males (Supplementary Figs. 5 and 6). Using Cox proportional hazards analysis, we dissected the effect sizes for diet (AL was reference), strain (*w1118* was reference), and orotate supplementation (0 mM was reference) on mortality in the fly separated by sex (Fig. 5E (females) and F (males), analysis with both sexes in Supplementary Fig. 7A). Being on the DR diet independently decreased the hazard ratio in a sex-specific manner (HR = 0.27, $P < 1 \times 10^{-27}$ in females,

https://doi.org/10.1038/s41467-024-52909-y

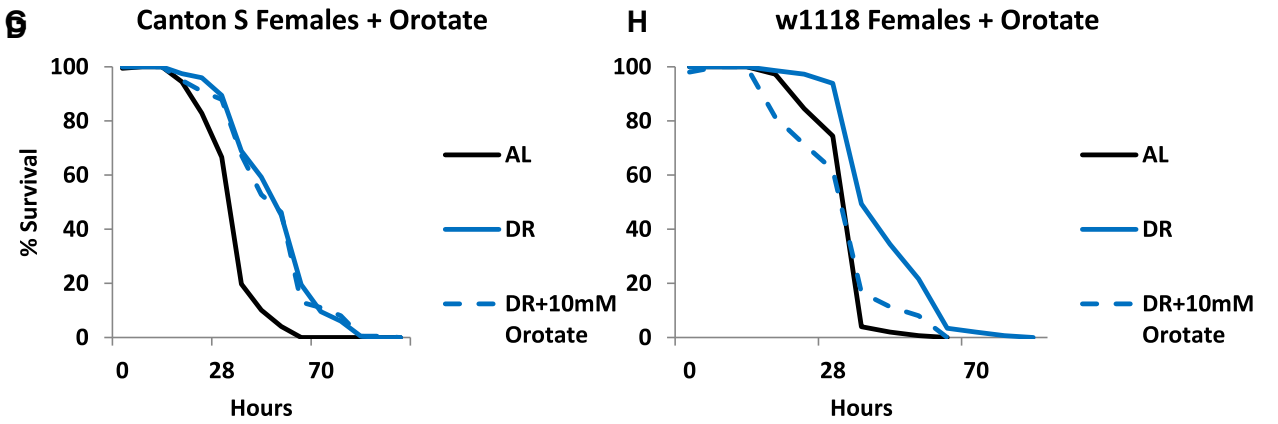






Variable	Hazard Ratio	estimate (low, high)	p
diet		0.27 (0.22, 0.34)	< 1e-27
orotate		1.35 (1.32, 1.38)	< 1e-27
orotate_diet		1.14 (1.11, 1.17)	1.88e-17
geno_diet		0.39 (0.30, 0.50)	2.40e-13
geno_orotate		0.90 (0.89, 0.92)	9.06e-22
orotate_diet_geno		1.10 (1.06, 1.14)	7.86e-07

Variable	Hazard Ratio	estimate (low, high)	p
diet		0.51 (0.44, 0.59)	8.13e-19
orotate		1.80 (1.73, 1.87)	< 1e-27
orotate_diet		1.07 (1.05, 1.09)	8.91e-11
geno_orotate		0.81 (0.80, 0.83)	< 1e-27



HR = 0.51, $P \sim 8.13 \times 10^{-19}$ in males). An AL diet supplemented with 10 mM orotate had a larger increase in the hazard ratio of males (HR = 1.80, $P < 1 \times 10^{-27}$) compared to females (HR = 1.35, $P < 1 \times 10^{-27}$). While the fly strain (genotype/geno) was not independently significant, there were significant sex-specific interaction effects with diet and orotate separately, or with orotate and diet in females. The strongest effect size was the interaction between sex and diet (HR = 5.20, $P < 1 \times 10^{-27}$).

Considering the sex- and strain-specific effects of orotate on lifespan and given orotate's previously shown role in fatty acid oxidation and lipid metabolism, we used a starvation assay to ascertain whether orotate supplementation was killing the flies by impacting their fat utilization[17,18]. Unlike the survival assays previously mentioned, orotate supplementation of the DR diet prior to starvation conditions did not affect survival compared to the DR control group for *Canton S* females (Fig. 5G, ALvsDR $P = 2.83 \times 10^{-9}$, ALvsDR+10 mM $P = 1.79 \times 10^{-7}$, DRvsDR+10 mM $P = 0.48$), but prevented the benefit of the DR diet for *w1118* females (Fig. 5H, ALvsDR $P = 2.09 \times 10^{-4}$, ALvsDR+10 mM $P = 0.49$, DRvsDR+10 mM $P = 2.09 \times 10^{-5}$). The males of each strain showed similar trends, though *Canton S* males showed no effect of diet

Fig. 5 | Orotate's significant decrease in fly lifespan is independent of diet, but the mechanism is strain- and sex-specific. A Effect estimates (betas per s.d. increase in the exposure) from Mendelian randomization (MR) analysis performed under the inverse variance-weighted (IVW) model are shown for BMI, frailty, parental age at death, and Type 2 diabetes phenotypes in humans for metabolites found to be important in both healthspan and lifespan Random Forest models. Betas and *P* values are shown in Supplementary Data 4. Supplementation of 10-mM orotate significantly shortens lifespan in *w1118* females (**B** AL FC ~ −0.36, $p \sim 4.58 \times 10^{-17}$, DR FC to −0.85, $p \sim 5.66 \times 10^{-32}$), *Canton S* females (**C** AL FC ~ −0.33, $p \sim 2.59 \times 10^{-17}$, DR FC ~ −0.80, $p \sim 3.02 \times 10^{-25}$), and w1118 males (**D** AL FC ~ −0.58, $p \sim 8.90 \times 10^{-30}$, DR FC ~ −0.799, $p \sim 3.64 \times 10^{-30}$) *Drosophila melanogaster* strains on both AL and DR diets. **E**, **F** Forest plots of hazard ratios from Cox proportional hazards analysis significant results testing diet (AL as reference), orotate levels (0 mM as reference), genotype (*w1118* as reference), and their interaction terms in flies separated by sex (Females–E, Males–F). Hazard ratios (HR) and 95% confidence intervals (CI) are shown for each predictor variable. All *P* values are two-sided and have not been corrected for multiple testing. **G** AL diet, but not DR diet supplemented with 10-mM orotate, significantly shortened survival under starvation conditions compared to DR diet in Canton S female *D. melanogaster* ($p \sim 2.83 \times 10^{-9}$, and 0.48, respectively). **H** AL diet alone and DR diet supplemented with 10-mM orotate, significantly shortened survival under starvation conditions compared to DR diet in *Canton S* female *D. melanogaster* ($p \sim 2.09 \times 10^{-4}$, and 2.00×10^{-5}, respectively), but are not significantly different from each other ($p \sim 0.49$). Lifespans began with 200 flies per group, with orotate supplementation lifespans repeated twice, and starvation lifespan once. AL diet is shown with black lines, DR with blue.

on survival under starvation conditions (Supplementary Fig 6, Supplementary Data 6). These data show a strain- and sex-specific response to orotate supplementation and suggest separate mechanisms causing the lifespan decreases.

Shared metabolites between Random Forest lifespan models and MR exhibit sex and strain-specific lifespan-extending effects in *Drosophila melanogaster*

Four serum metabolites (urate, inosine, histidine, and threonine) were detected in the *Drosophila* lifespan-only random forest models (Fig. 6A). In the MR analysis, none of these metabolites were significantly associated with parental longevity (all $P < 0.05$). However, three of the four overlapping metabolites had significant effects (all $P > 0.05$) on BMI, which can account for the effects on lifespan: histidine $\beta_{IVW} = -0.06$; 95% CI, −0.13 to −0.001; inosine $\beta_{Wald\ ratio} = 0.03$; 95% CI, 0.02–0.05; and threonine $\beta_{Wald\ ratio} = -0.05$; 95% CI, −0.06–0.05; Fig. 6A. We were unable to carry out sensitivity analyses and test for pleiotropic effects of all metabolites on BMI due to an insufficient number of SNPs, however the Cochran's Q statistics showed no evidence of heterogeneity (all $P > 0.05$)

Amongst the remaining metabolites (quinolinate, glucose, uridine, and kynurenine) also detected in the healthspan random forest model, only glucose and kynurenine were significantly associated with healthspan-related traits in the MR analyses (all $P > 0.05$). Genetically predicted glucose was significantly associated with both higher BMI ($\beta_{IVW} = 0.05$; 95% CI, 0.009–0.09; $P = 1.70 \times 10^{-02}$) and lower frailty index ($\beta_{IVW} = -0.08$; 95% CI, −0.03 to −0.04; $P = 1.21E{-}04$). The Cochran's Q statistics, MR-PRESSO global test, and the intercept of the MR-Egger method showed no evidence of heterogeneity or pleiotropy in the IVW analyses (all $P > 0.05$). Kynurenine was also significantly associated with increased BMI ($\beta_{wald\ ratio} = -0.02$; 95% CI, −0.03 to −0.001; $P = 4.04 \times 10^{-08}$) and decreased frailty index ($\beta_{wald\ ratio} = 0.04$; 95% CI, 0.007–0.06; $P = 1.59 \times 10^{-02}$). Although Cochran's Q statistics, MR-PRESSO global test, and the intercept of the MR-Egger method showed no evidence of heterogeneity or pleiotropy in the glucose–frailty index estimates (all $P > 0.05$), there was potential heterogeneity ($Q = 25.07$, $P = 1.50 \times 10^{-05}$) and horizontal pleiotropy (global test *p* value = 7.00×10^{-03}) detected in the associations with BMI.

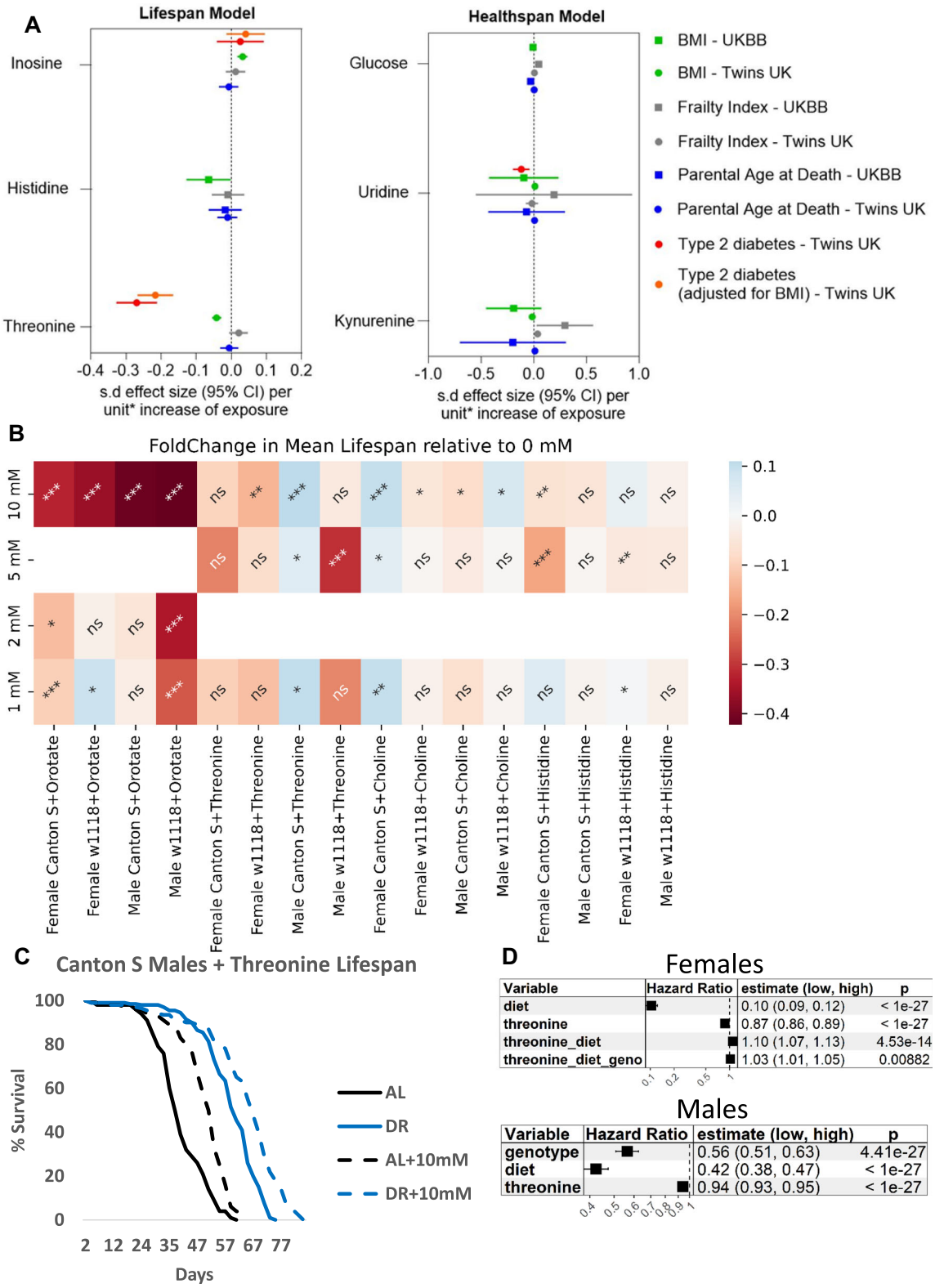
To validate the metabolites that were used to build Random Forest models of lifespan response and found to be relevant for humans via MR, we supplemented our AL diet with 3 concentrations of candidate metabolites (Fig. 6A). Validated metabolites were selected as described above. The impact of candidate metabolites choline, threonine, histidine, or orotate, on the survival of males and females from two fly strains, *w1118* and *Canton S*, was tracked and used to validate their role in lifespan determination. Each metabolite showed a strain-, sex-, and dose-specific effect on lifespan (Fig. 6B, Supplementary Fig. 5). 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 Fig. 6B, Supplementary Fig. 5). Supplementation of a DR diet with threonine, choline, or histidine showed similar strain- and sex-specific lifespan effects compared to AL, with 10 mM threonine supplementation extending female *Canton S* lifespan on both AL and DR diets (0.24- and 0.08-fold changes in mean lifespan, respectively. Figure 6C, Supplementary Fig. 6). Using Cox proportional hazards analysis, we dissected the effect sizes for diet (AL was reference), strain (*w1118* was reference), and threonine supplementation (0 mM was reference) on mortality in the fly separated by sex (Fig. 6D, analysis with both sexes in Supplementary Fig. 7B). Being on the DR diet independently decreased the hazard ratio in a sex-specific manner (HR = 0.10, $P < 1 \times 10^{-27}$ in females, HR = 0.42, $P < 1 \times 10^{-27}$ in males). An AL diet supplemented with 10 mM threonine had a larger decrease in the hazard ratio of females (HR = 0.87, $P < 1 \times 10^{-27}$) compared to males (HR = 0.94, $P < 1 \times 10^{-27}$). There were interaction effects of threonine with diet or with diet and genotype only in female flies. Together, these results validate our modeling approach and identify specific metabolites that are relevant for lifespan and are predicted to have translatable effects in humans. The variable response to metabolites also highlights the need for personalized approaches to lifespan and healthspan interventions.

Discussion

Genetic variability has been shown to greatly influence a strain's response to various conditions, including changes to diet[2,4,11]. Given that DR is one of the most robust ways to impact lifespan and other healthspan traits across species, understanding the underlying genetic variation and its role in diet-specific responses would help identify the mechanisms by which it functions[1]. While studies have investigated similar connections[4,19–22], none have employed a comparable systems biology approach that incorporates data from such a diverse range of model organism genotypes. Additionally, our utilization of MR with human data to uncover pertinent outcomes, along with our distinct modeling and visualization strategy for candidate selection and validation, sets our approach apart. Here, we have demonstrated the variability in outcome responses on one of two diets and the differing responses of each genotype. We have demonstrated that trait levels do play a role in different lifespan and healthspan outcomes, despite the lack of correlation across datasets between metabolites and phenotypes. This lack of correlation could indicate non-linear and/or indirect relationships between traits that would not be found in standard correlation analysis. We used a machine learning approach to identify non-linear relationships between our response and predictor traits. Our approach produced a value for each trait based on the proportion of trees/estimators that the trait was included in. This allowed us to confirm the importance of traits previously linked with their response variable, despite the lack of correlation across datasets. By using an unbiased, data-driven approach, we were able to identify relationships that would not be found in standard correlation analysis.

Some examples of traits used to build our response models include the metabolite malondialdehyde, which has previously been linked to glucose and triglyceride levels[23–25]. We found it was important

https://doi.org/10.1038/s41467-024-52909-y



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[26,27]. The metabolite methylhistidine, renowned for its correlation with the extension of lifespan induced by the drug RU486's transgene activation in flies, as well as its link to mortality and frailty, displayed significance in our initial mortality model under DR conditions[28]. Moreover, it emerged as notably significant in the dietary response influencing maximum lifespan. 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[29,30].

Fig. 6 | Supplementation of Mendelian randomization and RF overlapping metabolite threonine significantly extends fly lifespan and impacts starvation survival in a genotype-, dose-, and sex-specific manner. A Effect estimates (betas per s.d. increase in the exposure) from Mendelian randomization (MR) analysis performed under the inverse variance-weighted (IVW) model are shown for BMI, frailty, parental age at death, and Type 2 diabetes phenotypes in humans for metabolites found important in either lifespan or healthspan Random Forest models. Betas and *P* values are shown in Supplementary Data 4. **B** 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). Cox proportional hazards analysis *p* values were summarized (ns–not significant, *≤0.05, **≤0.005, ***≤0.0005), with exact values found in Supplementary Data 6. **C** Supplementation of 10-mM threonine significantly extends lifespan in *Canton S* males on both diets (AL FC - 0.24, p - 9.98×10^{-10}, DR FC - 0.08, *p* - 3.38×10^{-13}). **D** Forest plots of hazard ratios from Cox proportional hazards analysis significant results testing diet (AL as reference), threonine levels (0 mM as reference), genotype (*w1118* as reference), and their interaction terms in flies separated by sex. Hazard ratios (HR) and 95% confidence intervals (CI) are shown for each predictor variable. All *P* values are two-sided and have not been corrected for multiple testing. Lifespans began with 200 flies per group, with orotate and choline supplementation were repeated twice, threonine and histidine once. AL diet is shown with black lines, DR with blue.

Metabolites threonine, arginine, and choline were also included in our lifespan and healthspan models and had previously been reported to influence lifespan[31–39]. Threonine was used to build 8 of the 9 lifespan models and has been shown to increase lifespan in *C. elegans*[37,40–42]. Long-term supplementation with threonine was also found to improve lipid metabolism and inhibit fat mass in mice in the prevention of diet-induced obesity[43]. While threonine supplementation or preventing its catabolism can extend lifespan, threonine may also affect lifespan by modulating BMI on high-fat diets. We found diet-, strain-, and sex-specific effects of threonine supplementation on lifespan in *Drosophila*. Population studies on associations of threonine with type 2 diabetes mellitus (T2DM) are also mixed, ranging from positive to no association[44–46]. Associations of overweight and obesity with higher all-cause mortality are broadly similar in different populations[47]. Threonine is also currently in Phase 4 Trials for T2DM as part of an amino acid infusion to better understand glucagon and insulin release in comparison to glucose alone in obese individuals with and without T2DM (ClinicalTrials.gov ID NCT05264727). Kynurenine and quinolinate have both been positively associated with age in humans, with quinolinate also being strongly associated with frailty and mortality[48]. Additionally, myristic acid is associated with maximum lifespan potential in mice[49]. 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. 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[2].

We validated our approach to identifying underlying traits implicated in lifespan and healthspan by using traits that have already been shown to play a role in lifespan. While analyses of dietary influences on lifespan can be challenging due to genetic associations characterized by heterogeneity or pleiotropy identified using MR, we also used *Drosophila* to screen the functional consequences of different metabolites on health and lifespan traits. This approach holds the potential to overcome previous shortcomings of both fields and identify potential therapeutic pathways and metabolomic targets for diet response, lifespan, and healthspan. Indeed, we found several traits, including orotic acid, previously unconnected to lifespan or healthspan traits that make promising candidates as potential regulators or biomarkers. Administration of orotic acid is a well-known inducer of fatty liver conditions in rats[50–52]. Although the mechanisms underlying orotic acid-induced fatty liver are not entirely conclusive, various studies have proposed several potential pathways, including impaired fatty acid oxidation, upregulated lipogenesis, and a decrease in hepatic lipid transport[17,18]. In mammalian model organisms, orotic acid-induced fatty liver is species-specific, and thus far, only rats have exhibited susceptibility to its effects[53]. Our results also showed a genotype-specific response to orotic acid supplementation indicative of impaired fatty acid oxidation during starvation in *D. melanogaster*. Our approach using 174 strains and thousands of humans predicts this variable response and suggests that the identified metabolites are likely to be more conserved for most individuals. Nevertheless, the precise molecular mechanism by which orotic acid leads to fat accumulation remains unclear, particularly in the rat liver. It is worth noting that the primary dietary source of orotic acid for the average adult is milk and dairy products (approximately 0.005% of total solids), a concentration insufficient to elicit hepatic changes in rats[54]. However, considering the levels of orotic acid present in numerous dietary and health supplements widely available, as well as recent findings showing increased orotate levels to be detrimental to bone health, there is a need for caution regarding potential health implications for humans[55].

The few GWAS candidates that are shared between the traits used to build our models 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. The validation of overlapping candidates from both our human and *Drosophila* data shows that the candidates are relevant for both species. Threonine, choline, and histidine, in particular, are known to increase with age in humans of both sexes in plasma but not in blood[56]. While our data shows the diet-specific impact of metabolites, the mechanisms through which these metabolites work in humans may be different. Studies conducted using summary statistics for MR analysis preclude further stratification by sex or age. Furthermore, studies conducted in populations of predominantly European ancestry also limit the generalizability of findings to other populations. Despite these limitations, the broader implications of our work would potentially be more generalizable to other populations and highlight the need to study individual-specific responses in lifespan.

Considering we saw different responses to metabolite supplementation between both sexes of two *Drosophila* strains, it is expected that there will be similar variation in the response of individual humans to specific metabolites. The lifespan-extending effects of threonine and choline in flies show the need for more studies to assess whether they will have conserved effects in humans. Ways to optimize the 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.

Methods

Fly lines, husbandry, and diet composition

All fly lines were maintained on a standard fly yeast extract medium containing 1.55% yeast, 5% sucrose, 0.46% agar, 8.5% cornmeal, 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 a 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 into 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 were 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 were mated for 5 days, then were removed. 9 days later, non-virgin female and male progeny were separately sorted into an AL (standard diet with 5% yeast) or AL diet 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 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 were developed on standard fly 1.5% yeast extract medium and were transferred to the necessary diet within 72 h after eclosion. For survivorship analysis, vials with 25 mated flies 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[2,4,57–59]. 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.

Starvation assay

For starvation assays, flies (On AL media until the specified day) were transferred to vials containing 1% agar, and deaths were recorded every 4–6 h, three times a day.

Genome-wide association mapping

We used DGRP release 2 genotypes, and FlyBase R5 coordinates for gene models. As in Nelson et al.[2,11], 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. 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[60]. The linear model was phenotype = $\beta1 \times$ genotype + $\beta2 \times$ diet + $\beta3 \times$ genotype × diet + intercept. Nominal p values denoted as "genotype" in Fig. 1A report the probability that $\beta1 \neq 0$, and those denoted as "interaction" report the probability that $\beta3 \neq 0$.

Principal component analysis

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 were performed on all strains for each diet combination using the cor() function in the stats package in R programming. 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 seven 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[61]. 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[61]. Random Forest models were generated using the RandomForestRegressor function in the sklearn.ensemble package using 10,000 estimators[61]. Mean absolute percentage error and R2 values were determined using the mean_squared_error and r2_score functions in the sklearn.metrics package[61]. Feature importances were extracted from the final model.

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[62]. 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

We selected genetic variants i.e., SNPs strongly associated with different circulating metabolites from three different GWAS available (N up to 115,078). SNPs significantly associated with blood metabolites at the genome-wide significance level ($P < 5 \times 10^{-8}$) were selected and used. We used these genetic variants as IVs to examine the effects of genetically proxied metabolites involved in parental longevity–age at death, frailty index, and body mass index (BMI) in a two-sample summary data MR framework.

Exposures

Blood metabolite-associated SNPs and their genetic effects were obtained from a large-scale whole-genome sequencing study of 644 blood metabolites[63]. Briefly, whole-genome sequencing of 1960 adults of European descent was used to measure the association between genetic variations and blood metabolite levels. The serum samples were gathered over a period of 18 years and three clinical visits and analyzed on a non-targeted metabolomics platform. A linear mixed model was used to conduct association tests. Detailed reports for the methodology, design, sample collection, quality control, and statistical analyses of how the metabolomics analysis was performed are provided in the original study[63].

Wherever possible, we also used GWAS with a larger sample size albeit fewer metabolites overlapping with our random forest model. Additional summary statistics for glucose, histidine, and total cholines were also obtained from Nightingale Health Metabolic Biomarkers Phase 1 release GWAS of 115,078 participants in the UK Biobank (UKBB). SNPs underwent imputation using the HapMap-2 reference panel, and subsequent GWAS analysis was conducted using a linear regression model, with adjustments made for age, sex, and batch effects. Following quality control, a total of over 12.3 million SNPs were retained for subsequent analysis. The summary level data can be retrieved on the MRC-IEU database under the accession ID met-d[64].

Data for kynurenine and uridine were also obtained from a separate GWAS of non-targeted metabolomics. The study included 7824 participants from two European population cohorts: 1768 individuals from the KORA F4 study in Germany and 6056 from the UK Twin

Study[65]. Metabolic profiling was conducted on fasting serum samples using high-performance liquid chromatography and gas chromatography separation combined with tandem mass spectrometry, leading to the identification of 486 human serum metabolites. The final genome-wide association analyses involved ~2.1 million SNPs. The summary level data can be retrieved on the MRC-IEU database under the accession ID met-a[64].

Outcomes

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[66]. Secondary outcomes related to health and lifespan included frailty index, BMI, T2D, and T2D adjusted for BMI. Briefly, summary statistics for the Frailty Index were collected from a GWAS carried out among participants in the UK Biobank[67]. The Frailty Index 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[68]. Data summarizing SNPs linked to T2DM adjusted for BMI were extracted from a GWAS meta-analysis encompassing 48,286 individuals with T2DM and 250,671 control individuals. T2DM diagnosis relied on self-reported medical history and International Classification of Diseases (ICD) codes within electronically linked medical health records[69]. Data summarizing the correlation between genetic variants and physician-diagnosed T2DM were collected from a recent GWAS meta-analysis involving 62,892 T2DM patients and 596,424 control individuals[70]. More details of the GWAS on each of the outcomes can be found in their respective studies. The summary level data for parental longevity, frailty index, BMI, T2DM, and T2DM adjusted for BMI can be retrieved from the MRC-IEU database under the accession ID ebi-a-GCST006702, ebi-a-GCST90020053, ieu-b-40, ebi-a-GCST006867, and ebi-a-GCST007518 respectively.

Statistical analysis

We employed a two-sample MR approach. To choose valid instrumental SNPs, and ensure statistical independence across SNPs instrumenting for metabolite, we pruned the SNPs to eliminate those in linkage disequilibrium (LD) ($r^2 < 0.001$, window size = 10,000 kb) or absent from the LD reference panel (1000 Genomes European reference panel) using the clump function within the TwosampleMR package. We harmonized the SNP-exposure and SNP-outcome association data using the harmonize data function of the TwoSampleMR R package[71]. Data harmonization was carried out to ensure that the IVs had the same corresponding alleles on both exposures and outcomes for their effects. Finally, we calculated the F statistic for each SNP to evaluate its strength. Only SNPs with an F statistic > 10, strong enough to avoid weak instrument bias, were included in the analysis[72]. We then estimated the effect of genetically predicted metabolites on health and lifespan outcomes using the inverse variance-weighted (IVW) method or Wald ratio in case only a single SNP was present at $P < 5 \times 10^{-8}$ as our primary analysis. The MR results are given as effect estimates (β) and corresponding 95% CIs per unit increase in genetically predicted metabolites.

Sensitivity analyses

For our sensitivity analyses, we employed the weighted-median estimator, MR-Egger regression analysis, simple mode method, and weighted mode method to verify the key assumptions of our MR study[73]. The weighted-median method can provide a reliable effect estimate if at least 50% of the genetic variants used are valid instruments. MR-Egger regression, reliant on the InSIDE (instrument strength independent of direct effects) assumption, offers a valid effect estimate even in cases where all SNPs are potentially invalid instruments. The MR-Egger intercept quantifies the overall unbalanced horizontal pleiotropy effect across these genetic variants. We applied Cochran's Q statistic to assess heterogeneity in the genetic variants for the IVW method[74]. MR-PRESSO analysis was utilized to detect horizontal pleiotropy and identify any outlier SNPs[75]. Additionally, we employed the radial-MR method to identify heterogeneity and potential outliers[76].

This study is reported as per the Strengthening the Reporting of Observational Studies in Epidemiology guideline, specific for MR (Supplementary Information)[77]. No specific ethical review approval or informed consent was required as all data included in our study were available in public GWAS datasets, and these original GWAS had previously been approved by the appropriate ethical and institutional review boards.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used and generated in this study have been deposited in the Mendeley Data repository, (https://doi.org/10.17632/67f2sxfwxn). Data used from previously published datasets can be found at the following links: Nelson, 2016 DGRP data: https://doi.org/10.1186/s12864-016-3137-9. Jin, 2020 DGRP data: https://doi.org/10.1371/journal.pgen.1008835. Wilson, 2020 DGRP data: https://doi.org/10.1016/j.cub.2020.05.020. Kynurenine and uridine MR Datasets: https://doi.org/10.1038/ng.2982. Nightingale Health Metabolic Biomarkers Phase 1 (MRC-IEU) MR Datasets:https://gwas.mrcieu.ac.uk/ under the accession IDs met-d and met-a. Blood metabolite and genetic effect MR Dataset: https://doi.org/10.1038/ng.3809. The summary level data for parental longevity, frailty index, BMI, T2DM, and T2DM adjusted for BMI can be retrieved from the MRC-IEU database under the accession ID ebi-a-GCST006702, ebi-a-GCST90020053, ieu-b-40, ebi-a-GCST006867, and ebi-a-GCST007518. Human Longevity Data: https://doi.org/10.18632/aging.101334. Frailty Index Data: https://doi.org/10.1111/acel.13459. Height/Body Mass Index Data: https://doi.org/10.1093/hmg/ddy271. Type 2 diabetes69 Data: https://doi.org/10.1038/s41588-018-0084-1. Type 2 diabetes70 Data: https://doi.org/10.1038/s41467-018-04951-w. Human Phenome Data: https://doi.org/10.7554/eLife.34408. The data generated in this study are provided in the Supplementary Data files and available in the linked database. All statistical analyses were implemented using either R software (version 4.1.2) with the R package TwosampleMR (version 0.5.6), Radial-MR (version 1.1), and MR-PRESSO (version 1.0) or Python (version 3.9) with the packages listed in the methods. Summary GWAS data for MR analyses is publicly available and was obtained from the MRC-IEU OpenGWAS database or their respective studies. Additional specific details about the data, including, where applicable, links for downloading them, are available in the relevant publications referenced in this study.

Code availability

The custom code generated in this study has been deposited in the Mendeley Data repository (https://doi.org/10.17632/67f2sxfwxn).

References

1. Wilson, K. A. et al. Evaluating the beneficial effects of dietary restrictions: a framework for precision nutrigeroscience. *Cell Metab.* **33**, 2142–2173 (2021).
2. Wilson, K. A. 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. *Curr. Biol.* **30**, 2749–2760.e3 (2020).

3. Green, C. L. et al. Sex and genetic background define the metabolic, physiologic, and molecular response to protein restriction. *Cell Metab.* **34**, 209–226.e5 (2022).
4. Jin, K. et al. Genetic and metabolomic architecture of variation in diet restriction-mediated lifespan extension in Drosophila. *PLoS Genet.* **16**, 1–22 (2020).
5. Partridge, L., Fuentealba, M. & Kennedy, B. K. The quest to slow ageing through drug discovery. *Nat. Rev. Drug Discov.* **19**, 513–532 (2020).
6. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: an expanding universe. *Cell* **186**, 243–278 (2023).
7. de Magalhães, J. P. Why genes extending lifespan in model organisms have not been consistently associated with human longevity and what it means to translation research. *Cell Cycle* **13**, 2671–2673 (2014).
8. Broer, L. et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. Ser. A* **70**, 110–118 (2015).
9. Flachsbart, F. et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat. Commun.* **8**, 2063 (2017).
10. Xu, C. & Jackson, S. A. Machine learning and complex biological data. *Genome Biol.* **20**, 76 (2019).
11. Nelson, C. S. et al. Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. *BMC Genomics* **17**, 867 (2016).
12. Matsuki, K., Kuperman, V. & Van Dyke, J. A. The Random Forests statistical technique: an examination of its value for the study of reading. *Sci. Stud. Read.* **20**, 20–33 (2016).
13. Tomaschek, F., Hendrix, P. & Baayen, R. H. Strategies for addressing collinearity in multivariate linguistic data. *J. Phon.* **71**, 249–267 (2018).
14. Kulesa, A., Krzywinski, M., Blainey, P., Altman, N. & Lever, J. Ensemble methods.: bagging random. *Mach. Learn* **14**, 5–32 (2017).
15. Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **31**, 7–15 (1989).
16. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* **9**, e98679 (2014).
17. HEBBACHI, A.-M., SEELAENDER, M. C. L., BAKER, P. W. & GIBBONS, G. F. Decreased secretion of very-low-density lipoprotein triacylglycerol and apolipoprotein B is associated with decreased intracellular triacylglycerol lipolysis in hepatocytes derived from rats fed orotic acid or n–3 fatty acids. *Biochem. J.* **325**, 711–719 (1997).
18. Miyazawa, S., Furuta, S. & Hashimoto, T. Reduction of beta-oxidation capacity of rat liver mitochondria by feeding orotic acid. *Biochim. Biophys. Acta* **711**, 494–502 (1982).
19. Zhou, S. et al. Systems genetics of the Drosophila metabolome. *Genome Res.* **30**, 392–405 (2020).
20. Everett, L. J. et al. Gene expression networks in the drosophila genetic reference panel. *Genome Res.* **30**, 485–496 (2020).
21. Rohde, P. D. et al. Article genotype and trait specific responses to rapamycin intake in drosophila melanogaster. *Insects* **12**, 1–11 (2021).
22. Zhao, X. et al. The metabolome as a biomarker of aging in Drosophila melanogaster. *Aging Cell* **21**, e13548 (2022).
23. Noberasco, G., Odetti, P., Boeri, D., Maiello, M. & Adezati, L. Malondialdehyde (MDA) level in diabetic subjects. Relationship with blood glucose and glycosylated hemoglobin. *Biomed. Pharmacother.* **45**, 193–196 (1991).
24. Moreto, F., De Oliveira, E. P., Manda, R. M. & Burini, R. C. The higher plasma malondialdehyde concentrations are determined by metabolic syndrome-related glucolipotoxicity. *Oxid. Med. Cell. Longev.* **2014**, 505368 (2014).
25. Jalees, S. S. & Rosaline, M. Study of malondialdehyde and estimation of blood glucose levels in patients with diabetes mellitus with cataract. **4**, 319–323 (2017).
26. Hachinohe, M. et al. A reduction in age-enhanced gluconeogenesis extends lifespan. *PLoS One* **8**, e54011 (2013).
27. Treaster, S. B., Chaudhuri, A. R. & Austad, S. N. Longevity and GAPDH stability in bivalves and mammals: a convenient marker for comparative gerontology and proteostasis. *PLoS One* **10**, 1–13 (2015).
28. Landis, G. N. et al. Metabolic signatures of life span regulated by mating, sex peptide, and mifepristone/RU486 in female drosophila melanogaster. *J. Gerontol. Ser. Biol. Sci. Med. Sci.* **76**, 195–204 (2021).
29. Lugo-Huitrón, R. et al. Quinolinic acid: an endogenous neurotoxin with multiple targets. *Oxid. Med. Cell. Longev.* **2013**, 104024 (2013).
30. Castro-Portuguez, R. & Sutphin, G. L. Kynurenine pathway, NAD+ synthesis, and mitochondrial function: Targeting tryptophan metabolism to promote longevity and healthspan. *Exp. Gerontol.* **132**, 110841 (2020).
31. Lombardi, B., Pani, P. & Schlunk, F. F. Choline-deficiency fatty liver: impaired release of hepatic triglycerides. *J. Lipid Res.* **9**, 437–446 (1968).
32. Wu, G. et al. Choline deficiency attenuates body weight gain and improves glucose tolerance in ob/ob mice. *J. Obes.* **2012**, 1–7 (2012).
33. Emran, S., Yang, M., He, X., Zandveld, J. & Piper, M. D. W. Target of rapamycin signalling mediates the lifespan-extending effects of dietary restriction by essential amino acid alteration. *Aging* **6**, 390–398 (2014).
34. Roe, A. J. 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. *Am. J. Clin. Nutr.* **105**, 1283–1290 (2017).
35. Ravichandran, M. et al. Impairing L-threonine catabolism promotes healthspan through methylglyoxal-mediated proteohormesis. *Cell Metab.* **27**, 914–925.e5 (2018).
36. Bayliak, M. M. et al. 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.* **188**, 37–55 (2018).
37. Canfield, C. A. & Bradshaw, P. C. Amino acids in the regulation of aging and aging-related diseases. *Transl. Med. Aging* **3**, 70–89 (2019).
38. Velazquez, R. et al. Lifelong choline supplementation ameliorates Alzheimer's disease pathology and associated cognitive deficits by attenuating microglia activation. *Aging Cell* **18**, 1–11 (2019).
39. Huang, J., Ladeiras, D., Yu, Y., Ming, X. F. & Yang, Z. Detrimental effects of chronic L-arginine rich food on aging kidney. *Front. Pharmacol.* **11**, 1–13 (2021).
40. Aon, M. A. et al. Untangling determinants of enhanced health and lifespan through a multi-omics approach in mice. *Cell Metab.* **32**, 100–116.e4 (2020).
41. Yap, Y. W. et al. Restriction of essential amino acids dictates the systemic metabolic response to dietary protein dilution. *Nat. Commun.* **11**, 2894 (2020).
42. Kim, J., Jo, Y., Cho, D. & Ryu, D. L-threonine promotes healthspan by expediting ferritin-dependent ferroptosis inhibition in C. elegans. *Nat. Commun.* **13**, 6554 (2022).
43. Chen, J. et al. Threonine supplementation prevents the development of fat deposition in mice fed a high-fat diet. *Food Funct.* **13**, 7772–7780 (2022).
44. Yu, L. et al. The dietary branched-chain amino acids transition and risk of type 2 diabetes among chinese adults from 1997 to 2015: based on seven cross-sectional studies and a prospective cohort study. *Front. Nutr.* **9**, 881847 (2022).

45. Morze, J. et al. Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies. *Diabetes Care* **45**, 1013–024 (2022).
46. Vangipurapu, J., Stancáková, A., Smith, U., Kuusisto, J. & Laakso, M. Nine amino acids are associated with decreased insulin secretion and elevated glucose levels in a 7.4-year follow-up study of 5,181 Finnish men. *Diabetes* **68**, 1353–1358 (2019).
47. Nichols, E. et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 88 (2019).
48. Solvang, S. E. H. et al. Kynurenine pathway metabolites in the blood and cerebrospinal fluid are associated with human aging. *Oxid. Med. Cell. Longev.* **2022**, 5019752 (2022).
49. Jové, M. et al. Plasma long-chain free fatty acids predict mammalian longevity. *Sci. Rep.* **3**, 3346 (2013).
50. Standerfer, S. B. & Handler, P. Fatty liver induced by orotic acid feeding. *Exp. Biol. Med.* **90**, 270–271 (1955).
51. Buang, Y., Wang, Y.-M., Cha, J.-Y., Nagao, K. & Yanagita, T. Dietary phosphatidylcholine alleviates fatty liver induced by orotic acid. *Nutrition* **21**, 867–873 (2005).
52. Ferreira, A. V. M. et al. Fenofibrate prevents orotic acid—Induced hepatic steatosis in rats. *Life Sci.* **82**, 876–883 (2008).
53. Durschlag, R. P. & Robinson, J. L. Species specificity in the metabolic consequences of orotic acid consumption. *J. Nutr.* **110**, 822–828 (1980).
54. Jung, E.-J., Kwon, S.-W., Jung, B.-H., Oh, S.-H. & Lee, B.-H. Role of the AMPK/SREBP-1 pathway in the development of orotic acid-induced fatty liver. *J. Lipid Res.* **52**, 1617–1625 (2011).
55. Chen, Y. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.* **55**, 44–53 (2023).
56. Bucaciuc Mracica, T. et al. MetaboAge DB: a repository of known ageing-related changes in the human metabolome. *Biogerontology* **21**, 763–771 (2020).
57. Akagi, K. et al. Dietary restriction improves intestinal cellular fitness to enhance gut barrier function and lifespan in D. melanogaster. *PLoS Genet.* **14**, e1007777 (2018).
58. Sharma, A. et al. Musashi expression in intestinal stem cells attenuates radiation-induced decline in intestinal permeability and survival in Drosophila. *Sci. Rep.* **10**, 1–16 (2020).
59. Katewa, S. D. et al. Peripheral circadian clocks mediate dietary restriction-dependent changes in lifespan and fat metabolism in drosophila. *Cell Metab.* **23**, 143–154 (2016).
60. Seabold, S., Perktold, J. statsmodels: Econometric and statistical modeling with python. *In*: 9th Python in Science Conference (2010).
61. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
62. Hagberg, A., Schult, D. & Swart, P. Exploring network structure, dynamics, and function using NetworkX. In: (eds. Varoquaux, G., Vaught, T. & Millman, J.) 11–15 (Pasadena, CA USA, (2008).
63. Long, T. et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
64. Elsworth, B. et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv* https://doi.org/10.1101/2020.08.10.244293. (2020).
65. Shin, S. Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
66. Pilling, L. C. et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging* **9**, 2504–2520 (2017).
67. Atkins, J. L. et al. A genome-wide association study of the frailty index highlights brain pathways in ageing. *Aging Cell* **20**, e13459 (2021).
68. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
69. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes article. *Nat. Genet.* **50**, 559–571 (2018).
70. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
71. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
72. Pierce, B. L., Ahsan, H. & Vanderweele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**, 740–752 (2011).
73. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
74. Greco M, F. D., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* **34**, 2926–2940 (2015).
75. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
76. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *Int. J. Epidemiol.* **47**, 1264–1278 (2018).
77. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. *JAMA* **326**, 1614–1621 (2021).

Acknowledgements

T.A.U.H. was supported by NIH and NIA award F31AG062112, NIH/NIA training grant T32AG000266-24, and RF1 AG 068908–02. V.P.N was supported by the Glenn Foundation. K.A.W. was supported by NIH and NIA award F31AG052299, the NIH/NIA training grant T32AG000266-23, and the Buck CatalystX gift from Alex and Bob Griswold. This work was funded by grants from the American Federation of Aging Research (R.B.B. and P.K.); NIH grants R56AG038688, R21AG054121, RO1AG045835 (P.K.), R01GM120430 (R.B.B.), AG049494 (D.P.), and S10OD021562 (D.R.); the Hevolution Foundation (P.K.), and the Larry L. Hillblom Foundation. Additionally, D.P. received support from USDA cooperative agreement USDA/ARS 58-8050-9–004. We thank the Bloomington *Drosophila* Stock Center, and the Vienna *Drosophila* Stock Center for providing the flies used.

Author contributions

T.A.U.H., V.P.N., K.A.W., R.B.B., and P.K. conceptualized and designed the research. T.A.U.H., V.P.N., E.M.C., and D.R. performed the research, while K.A.W., D.R., and D.P. obtained the *Drosophila melanogaster* metabolic profiles. Data analysis was conducted by T.A.U.H. and V.P.N., who also provided the R and Python code used in the study. The original draft of the paper was written by T.A.U.H., V.P.N., and K.A.W., with D.P., R.B.B., E.M.C., and P.K. contributing to the review and editing process. Finally, D.P., R.B.B., J.C., and P.K. provided experimental guidance and funding for the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52909-y.

Correspondence and requests for materials should be addressed to Pankaj Kapahi.

Peer review information *Nature Communications* thanks Luke Pilling and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024