Peer Review File

Manuscript Title: Dietary restriction impacts health and lifespan of genetically diverse mice

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

Di Francesco and colleagues investigated the effects of CR (20% and 40%) and IF (1 day and 2 days of fasting per week) on the health and survival of 960 genetically diverse female mice. All interventions extended lifespan, although only CR reduced the mortality doubling time. They show that, IF did not extend lifespan in mice with high pre-intervention bodyweight and show evidence for the ability of mice to maintain bodyweight through periods of handling, as the top predictor of lifespan with additional predictors being changes in immune cells, red blood cell distribution width (RDW), and retention of adiposity in late life. They found that lifespan is heritable, and that genetic background has a larger influence on lifespan than dietary interventions. A significant association for lifespan on chromosome 18 that explained 4.3% of the diet-adjusted variation in lifespan. They propose that improving health does not strongly correlate with extending lifespan and that metabolic parameters may be inappropriate endpoints for evaluating aging interventions in preclinical models and clinical trials.
This is an important study since it involves a large number of mice and some of the major dietary interventions known to extend longevity and improve health. The use of genetically diverse mice is a major strength of the study. I have a number of questions and comments:

How do the DO mice compare to the HET3 mice? Why females?

How did the per month restriction compare? Did mice on the 1 day/week or 2 day/week fasting overfeed during the refeeding days? By how much? This is very important and should be measured and included in the results so it is clear what the overall monthly CR was for all groups.

It looks like the 1D and 2D fasting groups have a decrease in fat mass equivalent to that of the 40% CR yet they display an absolute lean body mass even higher than that of controls. This is very important and should be emphasized, especially when considering these mice as models for human studies. I also think the authors should show relative lean body mass for all groups and possibly lean over fat mass.

The authors should also point out the 15-20% lean body mass reduction caused by 4 months of 20-40% CR (between 6 and 10 months). This is a major decrease in lean mass considering that it is occurring in young mice and may increase even further in the 10-15 month period according to BW loss.
The authors write that “weight loss at any age was associated with reduced lifespan (Fig. 2c, second row). Can the authors separate the weight loss associated with cancer and other diseases and conditions

and the weight loss caused by aging? Is the same correlation observed?
What is the difference in heavy and light mice pre-DR? Do they have similar body composition and different weights or different fat levels?

Referee #2 (Remarks to the Author):

In this study Andrea Di Francesco, Gary Churchill and colleagues explored regulators of health and lifespan in a genetically diverse mouse population subjected to different regimens of dietary restriction, graded levels of caloric restriction (20% and 40%) and intermittent fasting (1 or 2 days of fasting/wk). Lifespan was heritable (h2 = 0.24) and genetic predictors had a larger impact than dietary interventions. Stress resilience was the top within diet predictor of lifespan, while other contributors were changes in immune cells and red blood cell distribution width and maintenance of adipose tissue later in life.
Below a few issues that can be improved in the current study.
Major comments:
1. The authors collected an impressive set of phenotypic traits, which are reported aggregated by dietary groups. A fundamental aspect of genetic studies and one of the goals of genetic populations is to generate high genetic and phenotypic heterogeneity. We suggest quantifying and showing the distribution of traits across the population in a more extended way (a clustered heatmap summarizing the main traits could be an example). This could be especially helpful for the QTL mapping section, and indicate the potential of the collected traits in identifying several genetic associations not reported here. Additionally, some analyses (such as QTL mapping) are reported for certain phenotypes but not others. It should be clarified which analyses were on which traits and how multiple testing was accounted for.
2. 40% caloric restriction seems quite severe for mice. Mice have a much higher metabolism than to human. How has this amount of CR been chosen? The mice continue to lose weight throughout their life, never reaching metabolic balance, as the authors state. This amount of caloric restriction not only impacts fat mass but the mice also lose lean mass. The authors conclude that body weight loss at any age is deleterious for lifespan, but I wonder how much this conclusion can be translated to human. Similarly, 48h of fasting is a very long time for mice and not everywhere allowed by ethical committees.
3. CR fed at 3PM, shifts the food intake timing compared to ad libitum fed mice, this could have an impact on many metabolic and health parameters.
4. What might be the impact of social dynamics of mice housed together competing for food in the DR? Can this be estimated (e.g., are heavier mice less likely to be found housed together than we might expect by chance)?
5. Fig. 1f and Ext. Fig 5: There is no clear indication here of the timepoints used for phenotypes, since they were measured at different times. As the authors point out later, having an increase in body mass in early life is related to litter size in mice and has been linked with reduced longevity before, but the relationship is inverted at old age. However, it is not clear on the figure which timepoints are being considered. The same applies to all longitudinal traits, it should be clarified in the figure which timepoint(s) were used and whether these correlations changed with age. In addition, it would be interesting to see which traits were associated with lifespan irrespective of dietary intervention. A

quantitative color legend for Fig 1f is also missing.
6. Along the manuscript there are mentions of circadian oscillations affecting lifespan and beautiful oscillating patterns are present for example in Extended Data Fig. 2. It might be very interesting to quantify those patterns to assess and verify their impact on lifespan.
a. Related to this, Day/night-based Delta RQ is used as a proxy for fasted vs. fed, however CR mice clearly have shifted behavioral patterns with a spike in food intake at feeding time rather than at night. The change in Delta RQ observed in CR mice is more likely due to this shift in behavior, rather than a metabolic change (i.e., Delta RQ is biased).
7. The trait with the strongest association with lifespan was resilience to BW loss after phenotyping. This observation is quite interesting as it highlights the importance of resilience to stress as an aging trait. It would have been interesting to go a bit more in depth on this finding, such as with QTL mapping of this trait, and finding closely-associated traits.
8. The findings about RDW as a predictor of reduced lifespan is potentially quite interesting, however there are a few concerns:
a. RDW depends on multiple factors, including most notably the mean corpuscular volume (MCV), and is generally not analyzed in isolation from other blood parameters. The results of RDW would be more informative if they were put into the context of other blood metrics.
b. The mediation analysis has severe limitations that must be acknowledged. Although the text refrains from claiming causality outright, it is implied throughout (e.g., l. 336-337: “2D and 40% interventions contributed to reduced lifespan acting through the RDW pathway (…)”). The limitations and assumptions should be clearly discussed to make clear what can and cannot be inferred from the analysis.
c. The identification of a significant lifespan QTL on chromosome 18 which colocalizes with one for RDW is potentially very interesting but very little is done to build on this result by, e.g., narrowing down the list of potential candidate genes for either trait. Possible approaches could include
i. Searching for cis-eQTLs at the same site in RNAseq datasets available in several organs in the DO and CC populations, and other large mouse populations.
ii. Searching human databases which include blood counts, such as the UK biobank, for associations between RDW and variants in any of the candidate genes.
iii. Other database searches, such as K.O. repositories, tissue expression maps and automated literature searches
iv. …
9. Methods and statistics are not well described or missing. For example, the method used to estimate heritability is never described. FDR correction is mentioned, was this correcting for all 194 phenotypes, are the multiple timepoints also taken into account? “Correlation analysis” is mentioned (l. 122) but this seems to refer to linear regression? l. 264 “The percent of total circulating lymphocytes in mice at all timepoints was positively associated with lifespan”, no statistics are given and the method used is only mentioned in the Figure caption but never described. Description of the results should include effect sizes or coefficients (where applicable) and p-values.
10. Why were linear models and ANOVA used to determine associations with lifespan? Are their respective assumptions not likely to be violated?
11. PLL does not eliminate survivorship bias. It attenuates it in the later stages of life but also increases it at other stages (for example, CR 40% mice appear to grow faster prior to DR in Fig. 2b).

12. l. 216-218 The fact that FI does not differ between diets (PLL scale) does not make it a reliable marker for biological aging. First, failure to reject the null hypothesis of a statistical test (that the trajectories are the same) is not the same as accepting it. Second, if it were a reliable marker, then there should be differences between the different DR to reflect differing rates of aging, which is not what is observed.
13. Several metabolic changes are labeled as “health benefits” based (as far as is justified in the text) on the fact they have previously “been suggested as potential mechanisms that mediate the extension of lifespan associated with DR” (l. 246-247). This is, at best, a weak reason to label them health benefits. More importantly, the results then presented suggest that these “health benefits” “are not mediators of longevity effects” (l. 251-252). If these results are accurate, then these metabolic changes are not health benefits.
Minor comments:
1. Could the associations observed between lean mass and lifespan be due to differences in body size?
2. The last paragraph of the introduction could be structured better (goes from explaining this study to background information back to this study, without making clear whether DRiDO is the current study or a previous one).
3. Line 82: a loss in percentage is mentioned but from the Extended Data Fig. 1 it is not clear that the vertical axis is in percentage. An interesting representation could be the evolution of BW (and the other parameters of Figure S1d) as a function of the initial BW, with time on x-axis and % of initial BW on the y axis.
4. Lines 85-89: the description is difficult to follow. A visual representation (a diagram, for example a timeline with daily interventions) could help. Fig. 1a appears intended to illustrate this but is not very clear and misleading (since CR mice were not fed on weekends).
5. Extended Data Fig. 2: It is unclear why only year 2 is used to generate panel a. Can the use of military time instead of Zeitgeber time be justified here? There are several mentions of circadian rhythmicity that could justify the use of Zeitgeber time. The temporal x-axis of the figures is not entirely clear. We also suggest specifying which model of metabolic cages are used. Do you observe similar patterns across different cages for the same strain and across strains? We suggest showing the heterogeneity of the groups differences across strains since this is a key aspect in a genetic population. We suggest showing only the 4-hours or 1-hour aggregated data and not both as this doesn’t seem to provide additional information except if used for some specific analyses absent here. Similar comments apply for Extended Data Fig. 3. For well distance, a useful metric to show could be the cumulative run distance over a given period of time.
6. Fig. 1c: add legend for circles and triangles.
7. l. 94-95 The lack of apparent difference in lifespan between interventions does not need any particular explanation, there are few deaths occurring prior to 18 months of age (10-20% mortality) and it is unlikely that the statistical power of the study would be sufficient to pick up any differences before.
8. Line 151: we suggest using *, **, *** for different levels of significance in Fig. 2c. This comment applies to all other figures too.
9. Line 180: We suggest showing an additional panel with all the mice and report the overall p-value mentioned in the text.
10. Ext. Figure 4b-c: Barplots show proportions, not numbers, adjust legend.

11. l. 232-234: Wheel running also declines with age for CR 40% mice (Ext. Fig. 8h).
12. Line 251-252: We suggest a more careful rephrasing of the statement, as all those analyses are based on correlation or Kaplan-Meier modelling. This suggests the absence of a primary mediating role but doesn't exclude the joint mediation role of some of the tested biomarkers when taken together for example, or their action through intermediate effectors.
13. Lines 289-290: it would be extremely interesting to also compute the percentage of explained lifespan variation for the (DR)x(genetic background) interaction term.
14. Lines 307-309 / Figure 5a: we suggest reporting the LOD significance score based on a permutation strategy as a horizontal line in Fig 5a and Extended Data Fig. 10d. It might also be very interesting to include a QTL model allowing the presence of a (DR)x(genetic background) interaction term.
15. Line 321: Although, as stated, the genetic mapping of physiological traits was not the focus, we suggest reporting a summary figure of empirical QTL peaks p-values for all traits (a Manhattan plot with only significant points) to have an overview of the traits with significant associations and justify the very specific selection of RDW in Fig 5d.
16. Line 348: We suggest rephrasing the statement. In the manuscript the authors showed the absence of association of several parameters with lifespan between DR groups, without doing proper predictive analyses.
17. Extended Data Fig. 6a: We suggest adding a trend line for each dietary group, to better compare the average BW evolution across groups.
18. A statistical result of "$p < 2.2e-16$" is reported multiple times throughout the paper. This is likely due to reporting standards in R summaries, however the full p-value should be reported where possible (unless the threshold is of greater importance, e.g., $FDR < 0.01$).

Referee #3 (Remarks to the Author):

Review of "Regulators of health and lifespan extension in genetically diverse mice on dietary restriction"

The authors present a very detailed analysis of the effects of dietary restriction (DR) and intermittent fasting (IF) on longevity, measures of health, and underlying molecular and physiological processes in female mice. The attention to detail is impressive and the analyses are appropriate in many instances, although the number of analyses may raise issues of false positives in certain settings or the interpretation of the results of the analyses can be problematic. Overall, I found the paper very informative and likely to raise more questions than it answers, which is not a bad thing given interest in lifespan extension and how to go about vetting potential lifespan extending interventions. For example, the fact that genetic background was a more important factor than the DR effect in mediating lifespan is hugely important for the community to acknowledge and probe further. I have a few important concerns the authors should address, all outlined below.

1. My biggest question is not detail-oriented, but simply: 'why only females?' There is no discussion about this in the paper and yet it is such an obvious question. The overarching goal of the paper is to identify and characterize factors that influence lifespan, including genetic factors, and sexual

dimorphism seems to be an obvious one to consider. If there is a reason that is well-known about why this choice was made, it should be emphasized early and also brought up in the discussion section.
2. There is a discussion of batch effects in the Methods section, but it does not provide detail. What was the magnitude of the batch effects, relative to the DR/IF effect, the genetic effects, etc.? This is very important since it speaks to important unmeasured sources of variation (e.g., more aggressive mice in one cage) or simply noise. I would publish these numbers and mention then in the main paper in a few sentences given how much emphasis there is on exploring sources of variation affecting DR/IF/longevity/etc.
3. The authors are careful in discussing the various analysis results in the different analysis settings (e.g., survival vs. batch effects vs. QTL mapping, etc.), which is crucial for a paper with as many different data types as this paper. This raises questions about false positive results, however. Some test settings make explicit use of strategies for controlling for multiple comparisons (e.g., longitudinal analyses make use of an FDR strategy; QTL analysis assume genome-wide significance levels; etc.) whereas others do not appear to control for multiple comparisons (e.g., analyses reflected in Figures 2, 3 and 4). I am not an advocate for correcting for the totality of statistical tests performed, but some attention to this must be paid. For example, in Figure 2f the 40% difference group effect may reflect noise, as with Figure 2i and the 2D and 20% groups. Even more pronounced are the results in Extended Figures 6-8
4. Related to item 3, Many of the analyses, especially around *differences* in lifespan among intervention groups, should be tested formally by aggregating the data, and not tested independently. Figure 2f and 2i are good examples. It is not clear if this was done, but single analysis models with all mice included using, e.g., indicator variables reflecting intervention groups, are appropriate; e.g., one large ANOVA or regression model. Interaction terms could be exploited in this model.
5. The partial correlation matrix analysis is fascinating, and I felt it was one of most exciting analyses the in light of the results presented in Extended Data Figure 4d and 4e. However, the interpretation here needs real care. For example, one interpretation of Extended Data Figure 5b would be that one could change CD4+ T Cell abundance or eosinophil abundance and affect lifespan, given that those might reflect more downstream processes modulated by DR/IF that ultimately impact lifespan. This may be the case, but are the authors committed to that kind of interpretation? If not, why not?
6. What are the genes in the linked region on chromosome 18 that make the most sense biologically? I am not expecting functional analyses with this paper, but, e.g., do any of the genes in the region have orthologs in humans that exhibit association with anything, let alone lifespan? Are they implicated in anything of note from the literature?
7. The authors should consider joint survival and longitudinal analysis in settings when the data are available to enable this.
8. In my opinion the interpretation of DR and IF interventions are absolutely confounded by what makes up the calories when food is eaten. I can't imagine that, for example, holding the timing of restricted feeding constant, diets with nutrition poor calories (i.e., don't contain essential nutrients) would work as well as diets that are nutrition rich (e.g., contain essential nutrients) in DR/IF settings. I don't know mouse chow, but why this issue is not explored seriously is a mystery to me and deserves comment in any paper exploring DR and IF. Also, as the authors point out, mice may live a long time, but if they are sickly and never put themselves in harm's way by isolating or accepting protection from others, then this may speak more to behavior than a direct MOA of a dietary intervention.

Author Rebuttals to Initial Comments:

To the Reviewers,

We would like to express our sincere gratitude for your thorough and constructive feedback on our manuscript. Your insightful comments and suggestions have been invaluable in guiding our revisions and enhancing the overall quality of our work.

We are pleased to inform you that we have carefully addressed each of your concerns and incorporated your recommendations into the revised manuscript. We trust that you will find that the manuscript is now significantly improved, both in terms of clarity and scientific rigor.

We appreciate your time and effort in reviewing our work and believe that the changes made have strengthened the manuscript considerably. We hope that the revised version meets your approval, and we look forward to your further comments.

Andrea and Gary on behalf of all authors.

Referees' comments:

Referee #1 (Remarks to the Author):

Di Francesco and colleagues investigated the effects of CR (20% and 40%) and IF (1 day and 2 days of fasting per week) on the health and survival of 960 genetically diverse female mice. All interventions extended lifespan, although only CR reduced the mortality doubling time. They show that, IF did not extend lifespan in mice with high pre-intervention bodyweight and show evidence for the ability of mice to maintain body weight through periods of handling, as the top predictor of lifespan with additional predictors being changes in immune cells, red blood cell distribution width (RDW), and retention of adiposity in late life. They found that lifespan is heritable, and that genetic background has a larger influence on lifespan than dietary interventions. A significant association for lifespan on chromosome 18 that explained 4.3% of the diet-adjusted variation in lifespan. They propose that improving health does not strongly correlate with extending lifespan and that metabolic parameters may be inappropriate endpoints for evaluating aging interventions in preclinical models and clinical trials.
This is an important study since it involves a large number of mice and some of the major dietary interventions known to extend longevity and improve health. The use of genetically diverse mice is a major strength of the study. I have a number of questions and comments:

1. How do the DO mice compare to the HET3 mice? Why females?

We thank the reviewer for raising this point. Aging studies in genetically diverse mice can reduce the risk of generalizing strain-specific idiosyncratic responses (Nadon et al., 2017 PMCID: PMC5514387[59]). They also provide an opportunity to identify genetic factors that affect aging and response to dietary interventions. DO mice and the UM-HET3 mice are models with genetically driven heterogeneity in age-related traits. We chose to work with the DO mice because they have more genetic variants and higher mapping resolution than UM-HET3 mice. DO and UM-HET3 mice have distinct ancestral origins. UM-HET3 mice descend from a 4-way cross between a female BALB and a male C57BL/6 (CByB6F1/J) as the mother and a cross between a female C3H and a male DBA (C3D2F1/J) as the father. This lineage ensures that the mitochondrial DNA, inherited maternally, and the Y chromosome, inherited paternally, remain consistent across all UM-HET3 individuals. Conversely, the breeding scheme for DO mice involves random mating, leading to a broader range of genetic diversity, including variability in

mitochondrial genotype. We have included a brief discussion of these differences in the revised manuscript (**Page 9 L7-10**)

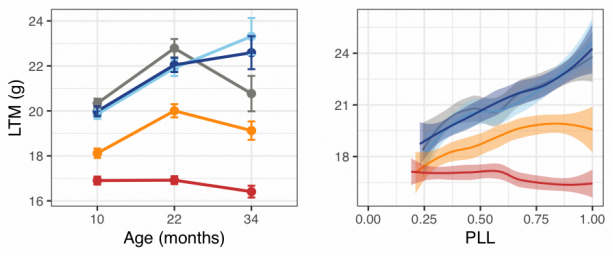
Regarding the use of female mice, we had concerns about male mice fighting. When male DO mice are cohoused, aggressive behavior can arise in ~10% of pens, and we were concerned that the incidence of aggression might be increased in resource-limited experimental groups, such as dietary restriction. Considering the cost of long-term housing and phenotyping, the possibility that DR could exacerbate aggression, and the potential for bias due to removing aggressive animals from the study, or loss of weaker animals due to fighting wounds, we chose to do a females-only study. While we acknowledge that this is not ideal because we will miss the opportunity to see sex differences, there are many studies that use only male mice. Our study tips the balance the other way. We emphasize that the study is female-only throughout the revised manuscript (**Page 9 L12-16**).

2. How did the per month restriction compare? Did mice on the 1 day/week or 2 day/week fasting overfeed during the refeeding days? By how much? This is very important and should be measured and included in the results, so it is clear what the overall monthly CR was for all groups.

We agree that food consumption and feeding behavior are important measures and have clarified this in the revised text. Our quantification of food intake is described in the Results section (**Page 2 L42 - Page 3 L8**). Food intake in the main study cohort was assessed in metabolic cages. In this setting mice are isolated and placed in an unfamiliar environment. There were additional concerns about the accuracy of the food hopper scale in the metabolic cages, especially when smaller amounts of food were added daily to feed CR mice. For these reasons, we established an independent cohort of 160 mice and maintained them under identical conditions to the main cohort of 960 mice. At 6.9, 8.3, and 9.9 months of age, we weighed food daily throughout the full seven-day feeding cycle, which is highly accurate and allows technicians to account for food hoarding and 'crumbling'. The daily data (**ED Fig 1a**) demonstrate compensatory feeding following the fasting periods for IF mice, consistent with the metabolic cage data from the main study cohort (**ED Fig 2a**). The total food consumption per week averaged across three weeks (**ED Fig 1b,c**) indicates that the 1D mice consumed a comparable amount of food to *ad libitum* (AL) mice, whereas 2D mice exhibited an approximately 12% reduction in overall food consumption. This assessment of cumulative food intake conducted over three non-consecutive weeks accurately reflects the home-cage food consumption in each of the DR groups, based on three non-consecutive weeks of assessments rather than monthly.

3. It looks like the 1D and 2D fasting groups have a decrease in fat mass equivalent to that of the 40% CR yet they display an absolute lean body mass even higher than that of controls. This is very important and should be emphasized, especially when considering these mice as models for human studies. I also think the authors should show relative lean body mass for all groups and possibly lean over fat mass.

We agree that the body composition changes in response to DR are important and have placed greater emphasis on these findings in the revised manuscript. The loss of lean mass under CR would be an undesirable outcome for humans. Unlike CR mice, IF mice have increasing lean mass throughout life that is indistinguishable from the AL mice. However, this should not be solely interpreted as increased muscle mass (**Page 10 L1-6**). The apparent difference between IF and AL at 34 months of age reflects differential survival which is accounted for in the PLL-scale plot (**Fig. 2h** and illustrated below in **Reviewer Fig. 1**), i.e., their end-of-life trajectories are similar.



Reviewer Fig. 1. Lean tissue mass across all groups (grey: AL, light blue: 1D, dark blue: 2D, yellow: 20%, red: 40%) by age (left) and percent of life lived (right).

DEXA analysis provides two independent measures of body composition: LTM and PercFat - an absolute measure of body size and a relative measure of fat content. However, the data can be transformed to other quantities including total tissue mass (TTM), fat tissue mass (FTM), percent lean mass (PercLean = 1-PercFat), fat-to-lean ratio (FLRatio = FTM/LTM), and lean-to-fat ratio (LFRatio = LTM/FTM). Following the reviewers' suggestion, we added the non-redundant traits to our analysis, which proved to be helpful. PercFat has the strongest association with lifespan, but the absolute measures LTM and TTM revealed diet x trait interactions on lifespan and highlighted some key differences between CR and IF interventions (**Fig 2h,i** and **ED Fig 5d-e**). (**Page 4 L43 – Page 5 L8**)

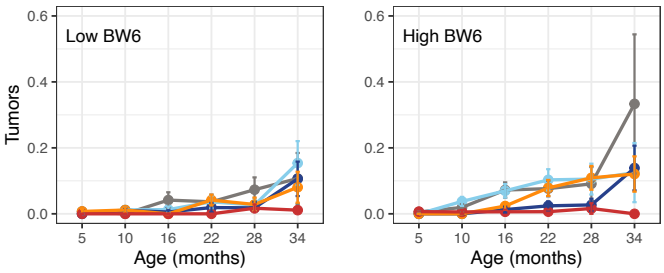
4. The authors should also point out the 15-20% lean body mass reduction caused by 4 months of 20-40% CR (between 6 and 10 months). This is a major decrease in lean mass considering that it is occurring in young mice and may increase even further in the 10-15 month period according to BW loss.

We agree the loss of lean mass is extreme and would be an undesirable outcome for both mice and humans. We highlight this in the revised manuscript text (**Page 10 L1-6**). Indeed, body weight loss in the post-intervention period is associated with shorter lifespan (**Fig. 2f**). Despite this, the CR mice appear to be healthy, and they live longer. Data in **Fig 2h** suggests that the lean mass in CR mice stabilizes between 10 and 22 months of age. The lack of pre-intervention DEXA scans was a shortcoming of this study – we can compare diet-specific average loss of lean mass in **Fig. 2h**, but we cannot track changes in individual mice across the pre- to post-intervention transition.

5. The authors write that "weight loss at any age was associated with reduced lifespan (Fig. 2c, second row). Can the authors separate the weight loss associated with cancer and other diseases and conditions and the weight loss caused by aging? Is the same correlation observed?
What is the difference in heavy and light mice pre-DR? Do they have similar body composition and different weights or different fat levels?

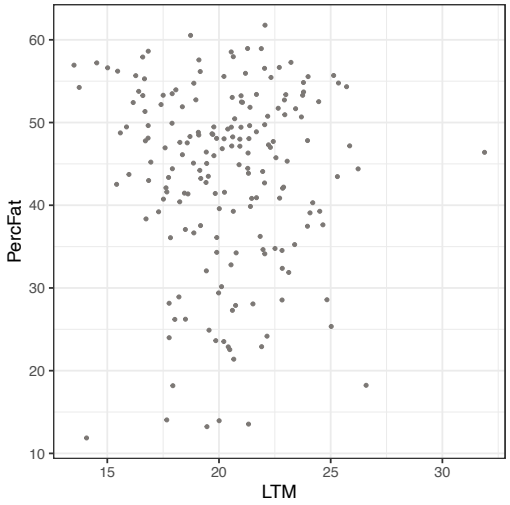
We were not able to determine whether mice have specific diseases. We recorded the presence of palpable masses as part of the frailty exam and in our experience most of these are neoplastic. We expect that many cancers were not detected. Mice with cancers typically lose weight[88], but we are not able to determine for any given mouse that weight loss was due to cancer or other disease. Age-associated weight loss typically occurs near the end of life (**Fig 2b**) and it could reflect a terminal disease state (**See Page 11 L37 - Page 12 L5**).

We observed that the incidence of palpable masses (i.e., tumors) and another indicator of serious illness (distended abdomen, DA) are reduced in mice with lower BW at 6 months. This could contribute to the increased lifespan of smaller mice (**Fig 2e**). Also, mice on 40% CR have greatly reduced rates of tumors and DA that could contribute to lifespan extension. Tumor incidence in other DR groups is moderately reduced. We felt that these observations were sufficiently important to warrant new figure panels (**Page 11 L37 – Page 12 L5**, and new **ED Fig 6c,d,** and illustrated below in **Reviewer Fig. 2**).



Reviewer Fig. 2. Tumors in mice examined at 6-month interval. Mice were stratified by body weight at 6 months.

As noted above, we did not measure body composition prior to the start of DR. However, when examining AL mice at 10 months of age, we observed that lean mass and adiposity are uncorrelated (r = -0.044, p = 0.55, **Reviewer Fig 3**). From this observation we infer that DO mice display a wide range of body shapes and sizes, including large lean mice and small fat mice.



Review Figure 3: Scatterplot of lean mass x adiposity for 10-month-old AL mice.

Referee #2 (Remarks to the Author):

In this study Andrea Di Francesco, Gary Churchill and colleagues explored regulators of health and lifespan in a genetically diverse mouse population subjected to different regimens of dietary restriction, graded levels of caloric restriction (20% and 40%) and intermittent fasting (1 or 2 days of fasting/wk). Lifespan was heritable (h2 = 0.24) and genetic predictors had a larger impact than dietary interventions. Stress resilience was the top within diet predictor of lifespan, while other contributors were changes in immune cells and red blood cell distribution width and maintenance of adipose tissue later in life. Below a few issues that can be improved in the current study.

Major comments:

1. The authors collected an impressive set of phenotypic traits, which are reported aggregated by dietary groups. A fundamental aspect of genetic studies and one of the goals of genetic populations is to generate high genetic and phenotypic heterogeneity. We suggest quantifying and showing the distribution of traits across the population in a more extended way (a clustered heatmap summarizing the main traits could be an example). This could be especially helpful for the QTL mapping section, and indicate the potential of the collected traits in identifying several genetic associations not reported here. Additionally, some analyses (such as QTL mapping) are reported for certain phenotypes but not others. It should be clarified which analyses were on which traits and how multiple testing was accounted for.

We appreciate the reviewer's acknowledgment of the scale of these data and although we explored several options to visualize them. The clustered heat maps were too dense and were unreadable. The correlation network in **Extended Data Figure 8**, provides a sense of the correlation structure for a cross-section of the trait data.

As suggested, we carried out QTL mapping analysis of all traits – discussed further below – and constructed a summary Manhattan plot of QTL peaks (**Extended Data Figure 9**, details provided in **Suppl. Data Table 12).** We also released the study data in an online QTL mapping tool[51] (https://churchilllab.jax.org/qtlviewer/DRiDO). These help to convey the scope of the data.

Regarding multiple testing, for all tests we computed an unadjusted and a false discovery rate (FDR) adjusted p-value. We used the Benjamini-Hochberg one-step method of adjustment. For the longitudinal analysis, we computed 6 test statistics for each of 204 traits and applied the FDR adjustment across traits and within each test (**Suppl. Table 6**). For the association with lifespan, we computed 2 primary tests (the first is a Diet and BW adjusted overall association of the trait with lifespan and the second is a trait x diet interaction test) and 5 post-hoc tests (within diet tests of association). We computed these tests on 720 trait-timepoint combinations and computed the FDR adjustment across trait-timepoint and within test (**Suppl. Table 7**). We applied a stringent selection criterion (FDR < 0.01) to identify significant traits that are reported in the manuscript. However, unless otherwise indicated as p^{adj}, we report unadjusted p-values in the text and figure captions. We chose to report the unadjusted p-values because they have a well characterized property – uniformly distributed when the null hypothesis is true – that make them easy to interpret. Whereas FDR adjusted p-values have variable and poorly characterized distributions that are dependent on other traits. All test statistics are reported in the supplementary tables – which are very useful for exploring the data beyond what we can capture in this short report.

2. 40% caloric restriction seems quite severe for mice. Mice have a much higher metabolism than to human. How has this amount of CR been chosen? The mice continue to lose weight throughout their life, never reaching metabolic balance, as the authors state. This amount of caloric restriction not only impacts fat mass but the mice also lose lean mass. The authors conclude that body weight loss at any

age is deleterious for lifespan, but I wonder how much this conclusion can be translated to human. Similarly, 48h of fasting is a very long time for mice and not everywhere allowed by ethical committees.

We agree with the reviewer regarding the severity of 40% CR in mice studies. However, early studies of CR in rodents by Weindruch et al. (1986; PMID: 3958810[92]) demonstrate that levels of restriction up to 65% significantly extend lifespan. The common level of restriction in the literature is 30% and we intended to bracket that level to determine if the more extreme (40% CR) intervention might reduce lifespan in some mice. This was motivated by the study of Liao et al. (2010; PMCID: PMC3476836) who reported genotype-specific lifespan response (increase and decrease) in mice on 40% CR. We hypothesized that the population average effect of 40% CR in DO mice could be neutral or possibly reduced lifespan. We also aimed to identify phenotypic predictors of shorter lifespan on 40% CR – and while we confirmed the findings of Liao et al.[37] (2011; PMCID: PMC3685291), that mice that maintain higher adiposity in late life have longer lifespan, we did not see evidence of reduced lifespan in any identifiable subpopulation of 40% CR mice.

Similarly, the 2D IF intervention is extreme. Previous studies with mice show that cycles of fasting lasting up to 60h are feasible and safe (Raffaghello et al., 2008, PMCID: PMC2448817; Lee et al., 2012, PMCID: PMC3608686; Cheng et al., 2014, PMCID: PMC4102383; Shi et al., 2012, PMCID: PMC3527202; Di Biase et al., 2016, PMCID: PMC5388544). We carried out pilot experiments to confirm that DO mice would tolerate the 2D IF intervention. There were no deaths, but we observed fluctuations in body weight (that we later confirmed, see **ED Fig 1d**). The 2D IF and 40% CR interventions were reviewed and approved by the Jackson Lab ACUC.

The reviewer raises an important point regarding the challenges in extrapolating these findings to humans. The most rigorous investigation of CR in humans was carried out by the Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy (CALERIE) Consortium (É. Le Bourg, L. Redman, 2018; PMID: 29897135). In CALERIE 1, prescribing 25% CR resulted in participants achieving 18% CR over 6 months. In CALERIE 2, which was a randomized controlled trial involving over 220 healthy, nonobese individuals undergoing 25% CR for 2 years, initial compliance was 19.5 ± 0.8% over the first 6 months, but significantly decreased to 9.1 ± 0.7% over the subsequent 18 months. These investigations underscore the challenge of sustaining reduced calorie intake over an extended period. Participants experienced 10% weight loss at 6, 12, and/or 24 months, primarily attributed to fat mass loss, although loss of fat-free mass was also noted. Participants demonstrated improvements in the quality of life and alterations in various health parameters often indicative of age-related or metabolic diseases. Notably, the seminal study on semistarvation by Keys et al. (1950; "The Biology of Human Starvation. Minneapolis: Univ. Minn. Press. 2 vol") revealed that an estimated 40% CR sustained over 24 weeks, induced by dietary restriction and increased exercise, resulted in a body weight loss of 25%, with approximately two-thirds as fat mass (FM) and one-third as fat-free mass. This level of CR led to severe adverse effects, both physical (e.g., chronic weakness, reduced aerobic capacity, and painful lower limb edema) and psychological (e.g., emotional distress, confusion, apathy, depression, hysteria, hypochondriasis, suicidal thoughts, and loss of sex drive), which manifested after 6 weeks. While there may be no simple correspondence with the degree of CR or duration of IF in mice, based on these studies, we expect that the 'human equivalent' of the 40% CR or 2D IF interventions would not be well tolerated. We have included these considerations in the Supplemental Discussion section (**Page 12 L36 – Page 13 L2**).

3. CR fed at 3PM, shifts the food intake timing compared to ad libitum fed mice, this could have an impact on many metabolic and health parameters.

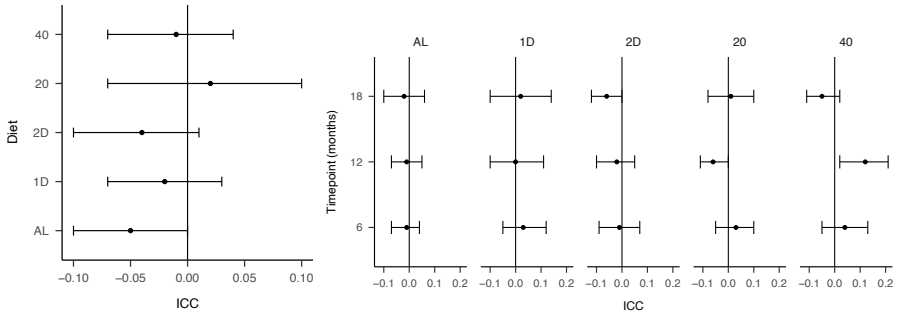
Thank you for raising this point. Indeed, the 3pm feeding and the triple-feeding on Friday for CR mice are integral aspects of our interventions. We made these choices in part to avoid imposing unusual

work hours on our technicians. The 3 p.m. feeding time closely approximates the normal (AL) onset of feeding, which for AL mice peaks at the beginning of the dark cycle. Since CR mice eat all available food in a short time, the peak of feeding time occurs earlier. Feeding just prior to the onset of the dark cycle has been recently shown to maximize lifespan extension in mice subjected to 30% CR (Acosta-Rodriguez, Science, 2022[29]). The feast-famine cycle induced by the Friday triple feeding has been employed in other studies of CR (Weindruch et al. 1986; PMID: 3958810[92]), but to our knowledge the health impacts have never been directly assessed. We now address these issues in the Methods section (**Page 29 L19-37**).

4. What might be the impact of social dynamics of mice housed together competing for food in the DR? Can this be estimated (e.g., are heavier mice less likely to be found housed together than we might expect by chance)?

The impact of social dynamics among mice housed together under DR is an intriguing aspect to consider. Mice are social animals. They interact extensively and rely on one another to stay warm. The choice to house mice in groups of eight was made to provide an optimal environment and to minimize the proportion of 'last' mice who outlive all of their cage mates. There is no practical way to ensure that all mice receive exactly the allocated amount of food, but previous studies have demonstrated the efficacy of CR in group-housed animals (Ikeno et al., 2005, PMID: 16424282**)**. We fed the CR cohorts by placing food directly in the floor of the pen and observed that mice 'claim' pellets and disperse to consume their share. We have highlighted this observation in the revised manuscript (**Page 29 L19-37**).

To address the concern about the effects of competition, we estimated the within-cage intraclass correlation coefficients (ICC) for lifespan and for body weight at 6, 12 and 18 months of age. When mice in the same pen are more similar (ICC > 0) or more dissimilar (ICC < 0) to one another than to a randomly selected mouse from the study, the ICC will deviate from zero. If competition is affecting these traits, we would expect to see negative ICC values, i.e., overdispersion of body weight and/or lifespans within housing groups. We observed that most of the 95% confidence intervals for ICC span zero, which is consistent with the absence of competition effects on these traits. This analysis is not described in the manuscript, but we provide a summary figure below.



Reviewer Figure 4: Intra-class (cage) correlation coefficients with 95% CI shown by diet groups for (left) lifespan and (right) body weight at 6, 12, 18, months of age. The CIs largely overlap with zero and all fall within +/- 10%, consistent with null or small within-cage correlations.

5. Fig. 1f and Ext. Fig 5: There is no clear indication here of the timepoints used for phenotypes, since they were measured at different times. As the authors point out later, having an increase in body mass in early life is related to litter size in mice and has been linked with reduced longevity before, but the

relationship is inverted at old age. However, it is not clear on the figure which timepoints are being considered. The same applies to all longitudinal traits, it should be clarified in the figure which timepoint(s) were used and whether these correlations changed with age. In addition, it would be interesting to see which traits were associated with lifespan irrespective of dietary intervention. A quantitative color legend for Fig 1f is also missing.

Thank you for pointing this out. We expanded the figure that shows the phenotyping schedule (now **ED Fig 3a**) to provide a more fine-grained view of the timing of measurements. Details of the phenotyping schedule are provided in **Suppl. Table 4**. Additionally, we note that AgeInDays is recorded for every test-domain in the data file (AllData.csv), to indicate the precise age at which data were collected for individual mice.

The partial correlation network (now **ED Fig. 8a**) is based on traits from the first post-intervention assessment for each phenotyping domain. Age at assessment has been added to the figure legend. The traits in the waterfall plot (now **ED Fig 8b** - with color legend added) were selected to represent each of the clusters in **ED Fig 8a** with a preference for traits with the strongest average within-cluster correlations and least missing data. The trait-lifespan associations can change with age. We focused on the earliest post intervention measurements in this analysis to capture the early response to diet and to minimize survivorship bias. Additional details about the network analysis are provided in **Suppl. Table 10**.

Significance of trait associations with lifespan (tabulated in **Suppl Table 7**) were computed using a regression model with diet and bodyweight as covariates, i.e., Lifespan ~ Diet + BW6 + BWTest + Trait. They are adjusted for diet and bodyweight and thus are diet independent. (For BW traits, we did not include the BW covariates (see **Page 34 L4-22**). To visualize the strength of association we present plots, e.g., **Fig 2f**, stratified by diet and display the within-diet BW adjusted correlations ("r = xx"). We replaced the KM plots from the original submission with scatterplots to provide a more informative view of the data and to be consistent with the linear regression methods used in their analysis.

An important point to emphasize is that the reported associations with lifespan first account for the effects of diet, i.e., we test the Trait term in the linear regression Lifespan ~ Diet + BW + Trait. We also tested the Diet:Trait interaction which would indicate the presence of a diet-specific effect on lifespan. Surprisingly few traits showed a significant interaction effect (10-month TTM, LTM and FTM being the notable exceptions). **Suppl Table 7** can be sorted by the Diet x Trait interaction p-value to see this. For most traits that are associated with lifespan the direction of the effects is consistent across diets, but with any test of interaction low power is a concern and failure to reject the null hypothesis of constant effect sizes across diets should not be interpreted as proof of a constant effect. This interaction test has 4 degrees of freedom, and a multiple testing adjustment was applied, exacerbating concerns about power. This is why we chose to display the stratified data.

6. Along the manuscript there are mentions of circadian oscillations affecting lifespan and beautiful oscillating patterns are present for example in Extended Data Fig. 2. It might be very interesting to quantify those patterns to assess and verify their impact on lifespan.

Thank you for your suggestion regarding the potential impact of circadian oscillations on lifespan. In response, we added new traits, including DeltaEE, DeltaRQ, DeltaVCO2, DeltaVO2, DeltaVH2O, DeltaBodyMass, to capture the magnitude of circadian oscillations (re: DeltaRQ, see below). It's worth noting that the circadian timing of fluctuations in these traits was tightly coupled to the feeding schedule. We present **ED Fig2** early in the results because these changes are the direct response to the timing of feeding that constitutes the DR interventions. While the patterns of oscillations are indeed visually striking, our quantitative analysis did not reveal any significant associations with lifespan.

When implementing an intervention like caloric restriction, it is not sufficient to say “we reduced food intake by 20%”. The details matter, including the timing of feeding daily and across the course of a week. This is why we present the data in **ED Figure 2** before addressing the other effects of the interventions. The circadian and weekly patterns of feeding and its immediate consequences provide the details necessary to precisely define the intervention.

a. Related to this, Day/night-based Delta RQ is used as a proxy for fasted vs. fed, however CR mice clearly have shifted behavioral patterns with a spike in food intake at feeding time rather than at night. The change in Delta RQ observed in CR mice is more likely due to this shift in behavior, rather than a metabolic change (i.e., Delta RQ is biased).

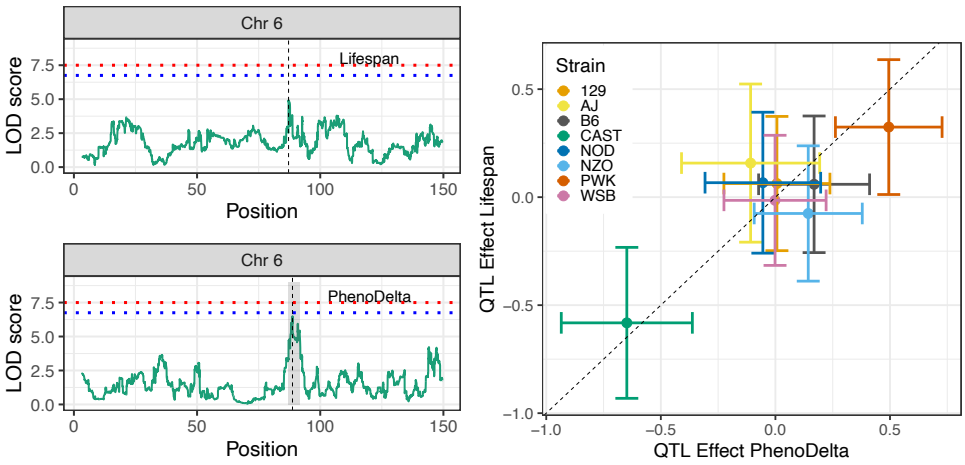
Thank you for highlighting this concern. The original trait for change in RQ (now ChangeRQ – see **Suppl Table 5**) was computed by taking the difference between the 5th and 95th percentile of RQ values across each 24-hour period, 2pm - 2pm and then taking the median of these daily ranges. Importantly, this calculation is not based on the dark-light cycle. We also computed RQ.dark and RQ.light which are tied to the light cycle and may be affected by the shift in peak feeding time.

In the revised data, we retained the original trait “ChangeRQ” and have added a new trait “DeltaRQ” that is the difference between the 5th and 95th percentiles of RQ across the entire week (**Page 31 L28-37**). This captures the full range of metabolic states that a mouse experiences (see new **ED Fig 2e**). The original ChangeRQ did not account for the extended periods of low RQ that occurred during the IF fasting periods. As expected, the new DeltaRQ increases in relation to the duration of fasting periods. CR leads to an increase in DeltaRQ because mice self-impose daily cycles of feeding and fasting, whereas in IF mice the changes occur over the course of each week (**Page 3 L14-16**). These adjustments should provide a more accurate reflection of metabolic changes in response to dietary interventions.

7. The trait with the strongest association with lifespan was resilience to BW loss after phenotyping. This observation is quite interesting as it highlights the importance of resilience to stress as an aging trait. It would have been interesting to go a bit more in depth on this finding, such as with QTL mapping of this trait, and finding closely-associated traits.

Thank you for bringing our attention to this observation. The PhenoDelta traits stand apart from most of the other lifespan predictors in correlation clustering (**ED Fig 3e**), with its nearest neighbors being RDW and HDW.

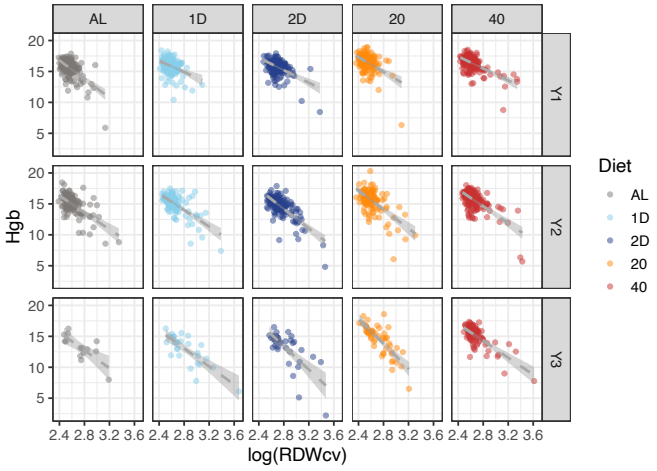
QTL mapping peak for PhenoDelta identified a suggestive association on chr 6 at 88.7Mb with LOD = 6.73 (suggestive threshold is LOD > 6.0). We observed a concordant, albeit non-significant, peak for lifespan on chr 6 at 87.4Mb, LOD = 5.24, and the allele effects of the two QTL align well (**Reviewer Fig. 5**), which is consistent with a shared genetic driver of PhenoDelta and Lifespan. The region around the peak is gene dense and complex with no obvious functional candidate. While this is of interest to us, the lack of significance of the QTL makes us hesitant to report the finding here. These and other genetic associations are discoverable using our online tool (https://churchilllab.jax.org/qtlviewer/JAC/DOAgingIntervention).



Reviewer Figure 5: (a) LOD curves of QTL on chromosome 6 for Lifespan and PhenoDelta at 10 months of age. (b) Allele effects at the QTL peak showing that both traits are lower in the presence of CAST alleles.

8. The findings about RDW as a predictor of reduced lifespan is potentially quite interesting, however there are a few concerns:
a. RDW depends on multiple factors, including most notably the mean corpuscular volume (MCV), and is generally not analyzed in isolation from other blood parameters. The results of RDW would be more informative if they were put into the context of other blood metrics.

RDW is the coefficient of variation of the size of red blood cells (RDWcv = standard deviation of corpuscular volume / MCV). We confirmed that RDW and MCV are only weakly correlated (r = -0.076, p = 0.023), as expected. MCV is not among the top lifespan associated traits. The trait most strongly correlated with RDW is HDW. HDW has the same diet response, but attenuated (Diet:p^{adj} = 5.6e-09, **ED Fig 7f**). RDW and HDW are also correlated with the hemoglobin traits (Hgb, Hct and NumRBC) and mice with elevated RDW have lower Hgb (**Reviewer Figure 6**). At 10 months of age, Hgb values for all groups are within normal mouse ranges (15.5 and above). After that, however, Hgb decreases quite substantially in all groups. The falling Hgb levels are likely contributing to the subsequent rise in both RDW and HDW, i.e., a consequence of age-associated anemia. The hemoglobin traits respond to DR (Hgb, Diet:p^{adj} < 2.2e-16, **Fig 4g**), but the pattern of response is distinct from RDW (**Fig 4f**). We conclude that CR is protective against age-related anemia while IF is not. Note that the hemoglobin traits are tightly correlated and we chose Hgb to represent the group. (**Page 6 L14-26, Page 11 L1-8**)



Reviewer Figure 6: Relationship of RDW to Hgb.

b. The mediation analysis has severe limitations that must be acknowledged. Although the text refrains from claiming causality outright, it is implied throughout (e.g., l. 336-337: “2D and 40% interventions

contributed to reduced lifespan acting through the RDW pathway (…)"). The limitations and assumptions should be clearly discussed to make clear what can and cannot be inferred from the analysis.

We appreciate the reviewer's insightful comments regarding the limitations of our mediation analysis. We recognize that the regression-based assessment of the RDW mediation has weaknesses (e.g., Gastonguay et al., PMCID: PMC10158839). Nonetheless, the 'causal steps' analysis (Baron and Kenny, PMID: 3806354) of RDW as a mediator of the CAST genotype on chr 18 looks convincing:

CAST -> Lifespan | RDW p = 0.00225
RDW -> Lifespan | CAST p < 2.2e-16
CAST -> RDW | Lifespan p = 1.93e-07
Lifespan -> RDW | CAST p < 2.2e-16

Considering the general concerns about mediation analysis, we have revised the text to avoid making implicit claims of causality.

c. The identification of a significant lifespan QTL on chromosome 18 which colocalizes with one for RDW is potentially very interesting but very little is done to build on this result by, e.g., narrowing down the list of potential candidate genes for either trait. Possible approaches could include
i. Searching for cis-eQTLs at the same site in RNAseq datasets available in several organs in the DO and CC populations, and other large mouse populations.
ii. Searching human databases which include blood counts, such as the UK biobank, for associations between RDW and variants in any of the candidate genes.
iii. Other database searches, such as K.O. repositories, tissue expression maps and automated literature searches
iv. …

i) We used online tools (https://qtlviewer.jax.org/) to look at cis-eQTL for genes in the QTL region. We considered all cis-eQTL with LOD > 6 in liver, kidney, heart, and bone. We identified two cis-eQTL with effects driven by CAST: *Dsc2* has low CAST expression in liver, and the gene model *Gm6378* has high CAST expression in kidney. We cannot rule out other genes that may have cis-eQTL in other tissues or at different life stages. As noted in the text, *Dsc2* is one of several genes in the QTL regions involved in cell junction formation and it has been implicated to have a role in heart disease in human studies. We also note that several genes, including *Garem1* have private CAST missense variants that could potentially affect their function independent of effects on gene expression.

ii) We analyzed GWAS summary data from UK Biobank. Consistent with previous reports[53], we observed a significant association between RDW and lifespan. We obtained coordinates of (>400) GWAS associations with RDW, but none were concordant with the mouse chr18 locus or colocalized with GWAS associations for human lifespan. While mildly disappointing, it is likely that distinct polymorphic loci are present between mice and humans (**Page 8 L5-9**, **Page 11 L10-13**).

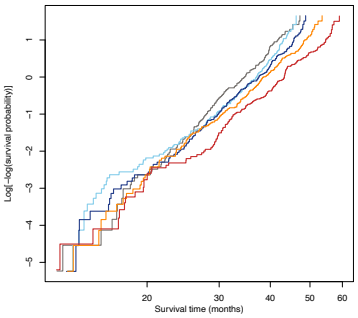
iii) We identified that *Garem1* knockout has a significant effect on spleen size and morphology in mice, suggesting a plausible link to RDW via recycling of red blood cells. None of the other genes has a reported blood phenotype in the mouse knockout consortium database (https://www.mousephenotype.org/).
Also see: https://www.informatics.jax.org/marker/phenotypes/MGI:2685790

9. Methods and statistics are not well described or missing. For example, the method used to estimate heritability is never described. FDR correction is mentioned, was this correcting for all 194 phenotypes, are the multiple timepoints also taken into account? “Correlation analysis” is mentioned (l. 122) but this seems to refer to linear regression? l. 264 “The percent of total circulating lymphocytes in mice at all timepoints was positively associated with lifespan”, no statistics are given and the method used is only mentioned in the Figure caption but never described. Description of the results should include effect sizes or coefficients (where applicable) and p-values.

Thank you for drawing our attention to these points. We have added the requested information for heritability analysis (**Page 35 L8-14**) and FDR correction (**Page 33 L38-40**, Page 34 L14-16; also see **Reviewer 3, comment 3**). We reorganized the **Results** to clarify the statistical analysis of BW-Diet-PLL effects on traits and the Lifespan-trait associations. These are the analyses from which most interpretation of the data follows. A complete tabulation of these results including unadjusted and multiple-test adjusted p-values can be found in **Suppl Tables 6 & 7**. We assessed trait-lifespan associations using a linear regression model Lifespan ~ Diet + BW6 + BWTest + Trait (without the BW covariates for direct and delta BW traits). Data were transformed to rank normal scores so that regression coefficients are comparable across traits and can be interpreted as adjusted correlations (**Page 34 L18-22**). We have added statistics throughout the text to support claims of significance. Statistics that are also reported in the supplement tables are shown in figure captions, and statistics not listed in the supplemental tables are cited in the main text. We have reported effect sizes on lifespan using Diet (and BW) adjusted regression coefficients as in (new) **Fig. 2d** to illustrate the magnitude of the BW effect on lifespan and how it changes with age.

10. Why were linear models and ANOVA used to determine associations with lifespan? Are their respective assumptions not likely to be violated?

We initially carried out all analysis with both CoxPH and linear regression/ANOVA models, and both approaches supported our main findings. We confirmed that the linear regression residuals are bell-shaped, and given the minimal censoring in our lifespan data,the assumptions of linear regression appeared to be met. On the other hand, the proportional hazards assumption for the Cox model is questionable based on diagnostic plots (**Reviewer Figure 7**) and significance of the slope term in fitted Gompertz models. Therefore, we chose to report the results from linear regression analyses. The panel plots (e.g., **Fig 2f, g**) have been redrawn as scatterplots to avoid the impression that we used survival analysis methods to evaluate the association. Where we have applied survival analysis (log-rank) tests, results are displayed as Kaplan-Meier curves (**Figs. 1b**, **2d**, **5g**, **ED Figs. 4d**, **10c**) and the reported p-values are indicated as Log-rank:p = 0.xx. (**Page 33, L10-17**). We believe this approach ensures clarity and transparency in our methodology, accurately reflecting the statistical techniques employed and addressing concerns regarding assumptions and interpretation.



Reviewer Figure 7: Estimated log-hazard by diet group. If the proportional hazards assumption holds, the lines should be parallel.

11. PLL does not eliminate survivorship bias. It attenuates it in the later stages of life but also increases it at other stages (for example, CR 40% mice appear to grow faster prior to DR in Fig. 2b).

A good point. Still, we find that PLL is useful for making comparisons relative to the end of life and we have retained the PLL plots throughout. We have modified our interpretation of PLL in text (**Page 4, L10-11**). Thank you for bringing this important aspect to our attention.

12. l. 216-218 The fact that FI does not differ between diets (PLL scale) does not make it a reliable marker for biological aging. First, failure to reject the null hypothesis of a statistical test (that the trajectories are the same) is not the same as accepting it. Second, if it were a reliable marker, then there should be differences between the different DR to reflect differing rates of aging, which is not what is observed.

We have removed the claim that FI is a marker for biological aging. This is a complex question that we feel would take us too far from the focus of the paper. However, we do wish to clarify here for the reveiwer. Specifically for the Frailty Index score, the interaction term in the regression model with age is significant (FI ~ BW + Diet + Age + Diet:Age, p_diet.age = 2.19e-05) indicating varying rates of increase in FI across diets on the chronological time scale, with the highest rate observed in the AL group. In contrast, the interaction term in the regression with PLL is not significant (FI ~ BW + Diet + PLL + Diet:PLL, p_diet.pll = 0.633), which is consistent with constant rate of change of FI across diets on PLL scale. This disparity in interaction effects between chronological age and PLL underscores an important point: if diets indeed influence the rate of aging, and if lifespan serves as a reliable proxy for biological aging, we expect to observe constant slopes on the PLL scale, but not on the chronological age scale. This finding supports the notion that PLL may provide valuable insights into the dynamics of aging that are not apparent when examining chronological age alone. We believe this clarification strengthens the interpretation of our findings regarding FI and its relationship to biological aging. Thank you for prompting us to further elucidate this aspect of our analysis.

13. Several metabolic changes are labeled as “health benefits” based (as far as is justified in the text) on the fact they have previously “been suggested as potential mechanisms that mediate the extension of lifespan associated with DR” (l. 246-247). This is, at best, a weak reason to label them health benefits. More importantly, the results then presented suggest that these “health benefits” “are not mediators of longevity effects” (l. 251-252). If these results are accurate, then these metabolic changes are not health benefits.

We appreciate the reviewer's clarification and apologize for the confusion – this was not how we justified the reference to health benefits. We have modified the text to provide clarity and highlight that that there are two distinct points:

First point: it is common in human studies and clinical practice to interpret reduced adiposity and lower fasting glucose as beneficial responses. This interpretation is reasonable in light of the high prevalence of obesity and metabolic disease in (modern/western) human populations. Whether weight loss and reduced fasting glucose should be viewed as beneficial for mice is less clear[56] (**Page 10 L18-20**).

Second point: others have proposed that changes in metabolic traits, specifically reduced adiposity and lower fasting glucose, mediate the lifespan extension effects of CR <<Martin et al., 2009. PMCID: PMC2852022; Redman and Ravussin, 2011. PMCID: PMC3014770>>. Our findings contradict this hypothesis. In fact, we see significant *suppression* of lifespan extension by several measures of metabolic health. This indicates that lifespan extension and improved health are not synonymous (**Page 10 L8-15**).

Minor comments:

1. Could the associations observed between lean mass and lifespan be due to differences in body size?

Yes. Lean mass is a proxy for body size. We note that 'lean mass' as assessed by DEXA includes muscle mass, soft organs, and fluids.

2. The last paragraph of the introduction could be structured better (goes from explaining this study to background information back to this study, without making clear whether DRiDO is the current study or a previous one).

The text has been updated see **Page 2 L24-33**.

3. Line 82: a loss in percentage is mentioned but from the Extended Data Fig. 1 it is not clear that the vertical axis is in percentage. An interesting representation could be the evolution of BW (and the other parameters of Figure S1d) as a function of the initial BW, with time on x-axis and % of initial BW on the y axis.

We only measured body composition before and after fasting and we plotted the change (After-Before) vs the baseline (Before). We considered plotting percent change, but the very small values for Fat (g) and Adiposity (%) create infinite or very large percentage changes. Description of the results is modified in the text to be consistent with the plot (**Page 30 L4-6, Page 16 L11-14**)

4. Lines 85-89: the description is difficult to follow. A visual representation (a diagram, for example a timeline with daily interventions) could help. Fig. 1a appears intended to illustrate this but is not very clear and misleading (since CR mice were not fed on weekends).

The text has been revised for clarity (**Page 2 L42 – Page 3 L8**). We have redrawn **Fig 1a** to more accurately summarize the feeding schedule. The timing of food consumed is shown in **ED Fig 1b**, and **ED Fig2a**.

5. Extended Data Fig. 2: It is unclear why only year 2 is used to generate panel a. Can the use of military time instead of Zeitgeber time be justified here? There are several mentions of circadian rhythmicity that could justify the use of Zeitgeber time. The temporal x-axis of the figures is not entirely clear. We also suggest specifying which model of metabolic cages are used. Do you observe similar patterns across different cages for the same strain and across strains? We suggest showing the heterogeneity of the groups differences across strains since this is a key aspect in a genetic population. We suggest showing only the 4-hours or 1-hour aggregated data and not both as this doesn't seem to provide additional information except if used for some specific analyses absent here. Similar comments apply for Extended Data Fig. 3. For well distance, a useful metric to show could be the cumulative run distance over a given period of time.

Thanks for the helpful suggestions. We have made several revisions to **ED Fig 2** based on your feedback. We now show only the 4-hour summaries of the data, as we agree these provide sufficient information. Additionally, we have included data from all three years. The labeling of the temporal (x) axis is clarified in the figure caption and the dark-light cycle (6am-6pm) is shown as shading for better visualization. We used the Sable Systems Int., Promethion cages, as detailed in the Methods section (**Page 31 L28-37**), and a protocol is provided in the FigShare files for reference.
The individual-level data for many of the channels are noisy and it is difficult to distinguish measurement noise from individual (genetic) variation – as such, overlay plots of individual traces are often not interpretable. We added panels to show individual summary values (**ED Fig 2e-g**), including cumulative wheel running to better show the substantial individual variation. The time series plots with

incremental wheel running are better for showing the mid-day activity burst in 40% CR mice and we retained these.

6. Fig. 1c: add legend for circles and triangles.

A legend had been added to **Fig 1c**.

7. l. 94-95 The lack of apparent difference in lifespan between interventions does not need any particular explanation, there are few deaths occurring prior to 18 months of age (10-20% mortality) and it is unlikely that the statistical power of the study would be sufficient to pick up any differences before.

Agreed. The text was removed.

8. Line 151: we suggest using *, **, *** for different levels of significance in Fig. 2c. This comment applies to all other figures too.

Fig 2c as well as heatmaps in **Fig 3a,b** and **Fig 4a,b** have been updated as suggested.

9. Line 180: We suggest showing an additional panel with all the mice and report the overall p-value mentioned in the text.

We added new traits for DeltaBW from 6 months to 15, 18, 21 and 24 months of age and dropped change from 6 to 20 that was shown in original figure. Summarizing weight change across 3-month increments is consistent with other summary body weight data. We computed overall diet-adjusted association with lifespan for the DeltaBW traits as reported in **Suppl Table S7**. Change in body weight from 6 to 18 months (Y2C_BW_Delta6) had a significant association with lifespan (Trait:p^{adj} = 5.8e-11), followed by 6 to 21 months and 6 to 15 months changes (ranked 8, 9, and 10 among top associations with lifespan). The diet-trait interaction is not significant (Diet x Trait:p = 0.36) which is consistent with a common effect across diet groups (i.e., a diet-independent effect on lifespan). The figure (now **Fig 2f**) has been revised to show effect size as within diet correlations (r = xx). The diet-stratified test statistics can be found in **Suppl Table 7**. We appreciate this comment as it motivated changes to make reporting of significances tests consistent throughout the text.

We want to note here that this result (shown in **Fig 2f**) is a key observation because it directly contradicts the common notion that lifespan extension on CR is due to reduced obesity and related metabolic changes. Here we see that the opposite is true – mice that lose more weight on CR have shorter lifespans compared to mice in the same intervention groups that retain weight. See **comment 13** above.

10. Ext. Figure 4b-c: Barplots show proportions, not numbers, adjust legend.

Revised as suggested.

11. l. 232-234: Wheel running also declines with age for CR 40% mice (Ext. Fig. 8h).

Revised text (**Page 3 L17-20**) and replaced **ED Fig 8h** with **ED Fig 2g**.

12. Line 251-252: We suggest a more careful rephrasing of the statement, as all those analyses are based on correlation or Kaplan-Meier modelling. This suggests the absence of a primary mediating role but doesn't exclude the joint mediation role of some of the tested biomarkers when taken together for example, or their action through intermediate effectors.

Demonstrating mediation is a challenging problem with many potential pitfalls. On the other hand, demonstrating that a variable is not a mediator is relatively easy. As noted above, we presented several examples of measured traits, including RDW and PhenoDelta at 10 months, where the direct effect of diet on lifespan is in the opposite direction of the indirect effect. We further addressed the potential for a joint mediating effect in the partial correlation network analysis (**ED Fig. 8a, b**), where the pattern of opposing effects persists even when other variables are taken into account. We interpret the measured traits as surrogates/proxies for unobserved biological processes that are suppressing or mediating effects of diet on lifespan (**Page 7 L3-9**). Also see **Reviewer 2 comment 13**.

13. Lines 289-290: it would be extremely interesting to also compute the percentage of explained lifespan variation for the (DR)x(genetic background) interaction term.

This is an interesting suggestion, and we spent some effort attempting to implement it using the R/lme4qtl software. However, the computations which require inversion of 5 large matrices proved to be infeasible. We feel that further research on the computational methods may be needed to estimate interaction variance components with DO mouse data.

14. Lines 307-309 / Figure 5a: we suggest reporting the LOD significance score based on a permutation strategy as a horizontal line in Fig 5a and Extended Data Fig. 10d. It might also be very interesting to include a QTL model allowing the presence of a (DR)x(genetic background) interaction term.

We added the permutation-based significance thresholds as requested. The genome-wide alpha = 0.05 'significant' LOD threshold is ~7.5, and the genome-wide 'suggestive' threshold is ~6.0 for both traits (Lander and Kruglyak, 1995; PMID: **7581446**).

We ran QTL analysis for all traits to search for Diet x Genotype interaction. The genome-wide threshold is high (LOD > 19) which is not surprising for a test with (5-1)*(8-1) = 28 degrees of freedom (df). Despite this, using the R/qtl2 software we identified a significant Diet-interaction QTL for lifespan on chromosome 5. However, the within-diet allele effect patterns were not consistent with the strength of the intervention. The high df associated with the test raises concerns about over-fitting and low power. In response, we reanalyzed the data using a different method (Wright et al., 2022, PMID: 35838135) and did not see any significant QTL. While identifying diet-specific genetic effects would be of great interest, we are circumspect about the chr 5 finding. Note that the interaction genome scan can be executed in our online tool for anyone who wishes to explore.

15. Line 321: Although, as stated, the genetic mapping of physiological traits was not the focus, we suggest reporting a summary figure of empirical QTL peaks p-values for all traits (a Manhattan plot with only significant points) to have an overview of the traits with significant associations and justify the very specific selection of RDW in Fig 5d.

We carried out QTL mapping on all of the traits and summarized the results in new **ED Fig. 9** and **Suppl. Table 12**. In addition, we implemented an online QTL mapping tool (https://churchilllab.jax.org/qtlviewer/DRiDO) to support exploration of the data (**Page 7 L20-26**).

16. Line 348: We suggest rephrasing the statement. In the manuscript the authors showed the absence of association of several parameters with lifespan between DR groups, without doing proper predictive analyses.

We appreciate the reviewer's suggestion and clarification. We used the term ‘predictor’ to indicate a significant association based on regression analysis of lifespan. This is a common usage and given the time-order of the variables is sensible. We interpret the reviewer’s comment ‘proper predictive analysis’ to mean an analysis with independent training and test sets to quantify predictive accuracy. While we acknowledge the value of such analyses, we believe that incorporating them would expand the scope of the paper beyond its intended focus. We are not aiming to predict lifespan *per se*. Therefore, we have modified the language to avoid ‘predictor’ and use ‘association’ to indicate a significant regression relationship between trait and lifespan.

17. Extended Data Fig. 6a: We suggest adding a trend line for each dietary group, to better compare the average BW evolution across groups.

Trend lines have been added to **ED Fig 4a**.

18. A statistical result of “p < 2.2e-16” is reported multiple times throughout the paper. This is likely due to reporting standards in R summaries, however the full p-value should be reported where possible (unless the threshold is of greater importance, e.g., FDR < 0.01).

We reported p < 2.2e-16 intentionally, following the best practices recommendation (Huber, 2019, PMID: 30921532) to avoid reporting precise p-values that are so extreme that distinctions among them are meaningless. The non-truncated versions are provided in **Suppl Tables 6 & 7** (**Page 33 L5-6**). The exact cutoff for reporting extremely small p-values is somewhat arbitrary but consider that the ratio of the Planck length to the diameter of the visible universe is ~75 orders of magnitude.

Referee #3 (Remarks to the Author):

Review of “Regulators of health and lifespan extension in genetically diverse mice on dietary restriction”

The authors present a very detailed analysis of the effects of dietary restriction (DR) and intermittent fasting (IF) on longevity, measures of health, and underlying molecular and physiological processes in female mice. The attention to detail is impressive and the analyses are appropriate in many instances, although the number of analyses may raise issues of false positives in certain settings or the interpretation of the results of the analyses can be problematic. Overall, I found the paper very informative and likely to raise more questions than it answers, which is not a bad thing given interest in lifespan extension and how to go about vetting potential lifespan extending interventions. For example, the fact that genetic background was a more important factor than the DR effect in mediating lifespan is hugely important for the community to acknowledge and probe further. I have a few important concerns the authors should address, all outlined below.

1. My biggest question is not detail-oriented, but simply: ‘why only females?’ There is no discussion about this in the paper and yet it is such an obvious question. The overarching goal of the paper is to identify and characterize factors that influence lifespan, including genetic factors, and sexual dimorphism seems to be an obvious one to consider. If there is a reason that is well-known about why this choice was made, it should be emphasized early and also brought up in the discussion section.

We appreciate the opportunity to address your question regarding our decision to exclusively study female mice, as well as in response to feedback from **Reviewer 1, comment 1**. This choice stemmed from several considerations, primarily concerning potential issues related to male aggression. In our experience, male DO mice when cohoused can exhibit aggressive behavior, with approximately 10% of pens displaying such behaviors. Given the substantial investment in long-term housing and phenotyping, we were concerned about the impact of DR on exacerbating aggression in male mice.

Additionally, there was a risk of introducing bias by removing aggressive animals from the study, potentially skewing the results. While we recognize that our decision may not be ideal, it was made to ensure the integrity and reliability of our findings. It's worth noting that many studies in the field have traditionally focused on male mice, often based on the misconception that females exhibit greater variability. By opting for a female-only approach, we aimed to provide a balanced perspective and contribute to the existing body of literature. Throughout the manuscript, we explicitly highlight the female-only nature of our study to ensure transparency and clarity (**Page 9 L12-16**).

2. There is a discussion of batch effects in the Methods section, but it does not provide detail. What was the magnitude of the batch effects, relative to the DR/IF effect, the genetic effects, etc.? This is very important since it speaks to important unmeasured sources of variation (e.g., more aggressive mice in one cage) or simply noise. I would publish these numbers and mention then in the main paper in a few sentences given how much emphasis there is on exploring sources of variation affecting DR/IF/longevity/etc.

Thank you for bringing attention to this aspect of our study. We appreciate the opportunity to provide further details on our methodology and its implications for our findings. Firstly, we want to emphasize that the most critical outcome in this study, and the only one for which we evaluate components of variance is lifespan (variance attributable to heritability and diet), and no batch correction was applied to lifespan or body weights. We confirmed that there were no significant lifespan associations for birth cohort ($p = 0.54$), parity ($p = 0.27$), coat color ($p = 0.14$), or mitochondrial genotype ($p = 0.29$). We have

now also confirmed that there were no significant co-housing effects on lifespan and bodyweight (see **Reviewer 2, comment 4).**

However, some of the measured traits were potentially susceptible to batch effects due to factors including collection date or technician. To facilitate a uniform analysis of traits and avoid having to use complicated linear mixed models, we preadjusted the raw data of all continuous outcome measures (except lifespan and body weight) for possible batch effects. All traits were corrected for collection date. We note that scheduling was coordinated such that on any given collection date, there were approximately equal numbers of mice from each diet group (although by year 3 differential mortality affected this balance). As such, there should be little potential for confounding. For some traits, we expected that other factors such as coat color or technician might be important. We selected the adjustment variables *a priori* and retained them even if not significant. For most traits, we fit

```
trait ~ (1|age) + (1|diet) + (1|mouse) + (1|date) + bwtest
```

using R/lme4 and we adjusted each observation by subtracting the estimated random effect (BLUP) of *date* (collection date). For the majority of traits the adjustments were negligible. Here are some examples showing proportion of explained variance attributable to various factors:

Frailty Score:

MouseID	Date	Age	Tech	Coat
1.604	3.189	80.829	14.363	0.015

Fasted Glucose:

MouseID	Date	Age
58.05	17.74	24.22

Percfat:

MouseID	Date	Tech	Age
65.07	0.52	0.0092	34.40

Frailty score had a substantial technician effect, and a negligible coat color effect. Fasted glucose had a substantial collection date effect - likely due to calibration of the glucometer. Percent fat, measured by DEXA had non-significant date and technician effects.

We feel that a complete enumeration of batch effects would be lengthy and complex to present. However, we have expanded our discussion of batch correction in Methods (**Page 33 L1-4**), and in **Supplemental Table 5** we list the covariates used for batch correction. We note that all source data are provided in the FigShare files along with descriptions of how data from each phenotyping domain were pre-processed. We hope this explanation adequately addresses your concerns.

3. The authors are careful in discussing the various analysis results in the different analysis settings (e.g., survival vs. batch effects vs. QTL mapping, etc.), which is crucial for a paper with as many different data types as this paper. This raises questions about false positive results, however. Some test settings make explicit use of strategies for controlling for multiple comparisons (e.g., longitudinal analyses make use of an FDR strategy; QTL analysis assume genome-wide significance levels; etc.) whereas others do not appear to control for multiple comparisons (e.g., analyses reflected in Figures 2, 3 and 4). I am not an advocate for correcting for the totality of statistical tests performed, but some attention to this must be paid. For example, in Figure 2f the 40% difference group effect may reflect

noise, as with Figure 2i and the 2D and 20% groups. Even more pronounced are the results in Extended Figures 6-8

We acknowledge the importance of addressing the issue of false positive results, - please also see **Reviewer 2, comment 1**. We have revised the Methods for clarity (**Page 33 L44 – Page 34 L2**). Briefly, the main conclusions of the paper derive from two sets of univariate analyses (summarized in **Suppl Tables 6 and 7)**. These tables are valuable resources for further exploration of the data. In addition, there are survival analyses (**Figs. 1b, 2e, and 5g**), a multivariate network analysis (**ED Fig. 8**), and the genetic analyses (**Fig. 5** and **ED Figs. 9-10**). Some reorganization of the presentation was done to better delineate these, specifically relocating the network analysis to follow the univariate trait analysis.

The longitudinal analysis (**Suppl Table 6**) is based on fitting a generalized additive mixed model to each of ~200 traits that were repeatedly measured on individual mice over the course of the study (**Methods Page 33 L19-).** We repeated the analysis using both PLL and Age, which is largely redundant (but see **Reviewer 2, minor comment 12**). The significance tests are based on likelihood ratio tests and reported as p-values (denoted as Diet:p = xx, PLL:p = xx, Diet x PLL:p = xx, Age:pp = xx, Diet x Age:p = xx to correspond with **Suppl Table 6**). For every test, we computed a false discovery rate (FDR) adjustment across the traits using the one-step Benjamini-Hochberg method (denoted using a superscript p^{adj}).

The association of traits with lifespan (**Suppl Table 7**) was assessed at each measured timepoint using a linear regression model (Lifespan ~ BW6 + BWTest + Diet + Trait + Diet x Trait) as described in Methods (**Page 34 L4-**). We report testing of the Trait and Diet x Trait terms (Trait:p = xx and Diet x Trait:p = xx) and FDR adjusted p-values (p^{adj}) across 724 traits.

We used multiple-test adjusted p-values to select the traits highlighted in the manuscript but, unless indicated (as p^{adj}), we have reported unadjusted p-values.

4. Related to item 3, Many of the analyses, especially around *differences* in lifespan among intervention groups, should be tested formally by aggregating the data, and not tested independently. Figure 2f and 2i are good examples. It is not clear if this was done, but single analysis models with all mice included using, e.g., indicator variables reflecting intervention groups, are appropriate; e.g., one large ANOVA or regression model. Interaction terms could be exploited in this model.

We agree that aggregating the data and testing for interaction effects using single analysis models with all mice included is appropriate. As noted above (**Reviewer 2 comment 5**), there are some concerns about the power of the interaction test, so we don't not assert that effects are uniform across diets when adjusted p-value for the interaction is not significant. To address this, we now report the unadjusted Diet x Trait interaction p-value e.g., in **Figs 2f and 2g** captions. We report effect sizes as diet-stratified correlations (shown in figure panels) – the diet-specific test statistics can be found in **Suppl Table 7** for completeness, but they are not reported or used to interpret the data in the manuscript.

5. The partial correlation matrix analysis is fascinating, and I felt it was one of most exciting analyses the in light of the results presented in Extended Data Figure 4d and 4e. However, the interpretation here needs real care. For example, one interpretation of Extended Data Figure 5b would be that one could change CD4+ T Cell abundance or eosinophil abundance and affect lifespan, given that those might reflect more downstream processes modulated by DR/IF that ultimately impact lifespan. This may be the case, but are the authors committed to that kind of interpretation? If not, why not?

This is a good point. We acknowledge the importance of exercising caution in interpreting the results of the partial correlation analysis, particularly in terms of inferring causality. We want to clarify that our intention was not to suggest a direct causal relationship between the measured variables and lifespan. This is now clarified in the text (**Page 7 L3-9**). The measured variables are best thought of as surrogates for unmeasured latent factors that are the true causal mediators. This analysis allows us to establish that there are multiple pathways mediating the effects of DR on lifespan, as there is no single variable that captures a large proportion of the mediated effect, and there are both mediating (consistent direction of effects) and suppressing (inconsistent direction of effect) paths from diet to lifespan. Despite the wide scope of measured traits, they fall short of fully explaining the DR effects.

6. What are the genes in the linked region on chromosome 18 that make the most sense biologically? I am not expecting functional analyses with this paper, but, e.g., do any of the genes in the region have orthologs in humans that exhibit association with anything, let alone lifespan? Are they implicated in anything of note from the literature?

The annotated protein coding genes are shown in **ED Fig 10d-e.** The most relevant genes are discussed in the text (**Page 7 L41 – Page 8 L5** also **Reviewer 2 comment 8**). While some of the genes have known function that would make them plausible candidates, but there is no single gene candidate that stands out.

7. The authors should consider joint survival and longitudinal analysis in settings when the data are available to enable this.

This is an interesting suggestion. Our analysis is already complex and the suggested analysis would present some challenges such as accounting for asynchronous measurement times and, because there is a sharp drop-off of the numbers of surviving animals in year 3, many of our measured traits have effectively two time points. While methods to address these challenges exist, the additional analysis would be substantial and seems best reserved for future work.

8. In my opinion the interpretation of DR and IF interventions are absolutely confounded by what makes up the calories when food is eaten. I can't imagine that, for example, holding the timing of restricted feeding constant, diets with nutrition poor calories (i.e., don't contain essential nutrients) would work as well as diets that are nutrition rich (e.g., contain essential nutrients) in DR/IF settings. I don't know mouse chow, but why this issue is not explored seriously is a mystery to me and deserves comment in any paper exploring DR and IF. Also, as the authors point out, mice may live a long time, but if they are sickly and never put themselves in harm's way by isolating or accepting protection from others, then this may speak more to behavior than a direct MOA of a dietary intervention.

We agree that the choice of diets for aging studies is an important consideration. We used the same diet for all mice in this study – LabDiets 5K0G is optimized to provide the essential nutrients required by mice. The diet spec sheet is included in the supplemental files (Figshare/Protocols). How to formulate optimal diets for aging mice remains a long-standing open question (McDonald 1997 PMID: 9164251), but recent studies do support the use of 5K0G for aging mice (Lee et al, 2023, PMID: 37079216).

Regarding the appearance of the aged CR mice, we understand how the reviewer got the impression that they look 'sickly'. In fact, the mice look remarkably healthy, with the qualifier that many reached advanced age. This is confirmed by the frailty score (see **Fig. 3c**). The CR mice are objectively colder (**Fig. 3e**), they lose body weight and lean mass (**Fig. 2h, i**), and they display food seeking behavior that indicates hunger (**ED Fig. 2d,g**). There may be additional effects of CR, e.g., on reproductive function, that were not observed here. Considering the heightened interest (and some self-administration) of purported life-extending diets by people, we wished to be clear that the interventions we have applied

to mice are not directly translatable to humans and may have undesirable effects. It is our opinon that intolerable levels of restriction would be needed to achieve similar levels of lifespan extension in humans (**Page 12 L40 – Page 13 L2**; also see **Reviewer 2, Comment 2**).

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have addressed many of my concerns and suggestions. However, there are still a number of issues that I think should be addressed:

I believe that it is important to increase the reporting on the negative effects of CR and on the positive effects of 1D and 2D IF. I believe this needs to be done in the abstract and more clearly in the results and discussion sections

1) 40% CR causes a strong loss in lean mass decrease (15%) but 20% CR also causes loss in lean mass already at 10 months maintained until 34 months whereas 1D and 2D show a remarkable increase in lean mass at both 22 and 34 months. By 34 months the lean mass of 1D and 2D is about 40% higher compared to 40% CR. Furthermore, the increase or stabilization in lean mass between 22 and 34 months in the IF groups is impressive but barely mentioned.

2) While 1D and 2D IF increase lean mass they still show a reduction in fat mass. It would be good to add a lean/fat loss figure

3) The fact, that 1D and 2D are achieving lifespan extension with no or minimal calorie restriction should be emphasized since now we know that we is being compared is not equivalent, meaning 2 groups of mice are truly calorie restricted and in fact even the 20% CR is restricted by 27% whereas the 1D consumes more calories than controls. These are fundamental issues and should be front and center, especially because a 27-43% CR below normal calorie intake in humans would not even be considered, also for the reasons shown below.

4) The main figure show a decrease in effector CD4 cells in 40% CR but the extended figure 7 shows a major reduction in B cells, monocytes and NK cells in 40% CR. Whereas these changes may not be consequential in the sterile/protected vivarium environment they could have major consequences in the wild. This is another key part of the data that needs to be included in the abstract and much more clear in the manuscript.

Referee #2 (Remarks to the Author):

I applaud the thorough job done on revising this study. Two minor things to be addressed:
(1) the website listed under the QTL analyses section did not work (https://churchilllab.jax.org/qtlviewer/JAC/DOAgingIntervention), it should be changed.

(2) please explain how the threshold of QTL mapping was calculated.

Johan Auwerx

Referee #3 (Remarks to the Author):

Second Review of. “Regulators of health and lifespan extension in genetically diverse mice on dietary restriction”

The authors have done an excellent job responding to my and, from what I can gather, the other reviewers’ concerns. The emphasis on Dr and IF findings that are statistically significant and biologically meaningful improve the value of the paper. The big take away messages to my mind are the complex nature of dietary restriction, that genetics is more important than an DR/IF intervention for lifespan, the fact that metabolic phenotypes are not key players in lifespan, 'resilience to stress' is a key predictor of lifespan, and the obvious distinctions between humans and mice wrt to longevity enhancement. I believe the authors have responded adequately to my individual concerns, but there is still an overall potential for false positives across all the assays and analyses that would be hard to quantify. I found the revised manuscript much more readable and better organized, laying out the issues and analyses so that readers would know that many are ‘hypothesis generating’ or upset current thinking in some way but are not pivotal.

One minor issue is with the mediation analysis. I had brought this up in my previous review and I think the authors responded appropriately, but their response raises an additional question which concerns what the processes might be that the immune cells are acting as surrogates for? Obviously, there is no way of knowing this, but, for example, what other phenotypes did the immune cells correlate with? If DR/IF do affect immune cells, and this affect leads to downstream changes resulting in lifespan, then of course one could look for other ways of modulating those immune cells – beyond DR and IF – with the belief that this will affect lifespan. Therefore 1-2 sentences on insight into what those immune cell parameters may have correlated with that may get at this would be of value. It was hard to go through the various results to pick this out easily (e.g., do the immune cell parameters correlate with lack of infections, a better balance in cell populations, diversity in cell populations, effect on other processes like cellular communication, etc.?).

Referee #3 (Remarks on code availability):

I did NOT look at all the code, but did use the website that displays the results. Like all websites providing graphical summaries of analyses, it is a bit clumsy but it works. I think the real value is in letting others have access to the data.

Author Rebuttals to First Revision:

Response to Reviewer Comments (second review)

Referee #1 (Remarks to the Author)

The authors have addressed many of my concerns and suggestions. However, there are still a number of issues that I think should be addressed:

I believe that it is important to increase the reporting on the negative effects of CR and on the positive effects of 1D and 2D IF. I believe this needs to be done in the abstract and more clearly in the results and discussion sections

Keeping in mind that our findings indicate distinct negative (and beneficial) effects of both CR and IF, we added text to the abstract (see page 1: *Health effects differed between interventions and exhibited inconsistent relationships with lifespan extension. 40% CR had the strongest lifespan extension effect but led to loss of lean mass and changes in the immune repertoire that could confer susceptibility to infections. IF did not extend lifespan in mice with high pre-intervention body weight, and 2-day IF was associated with disruption of erythroid cell populations.*) and we have expanded on this point in the Discussion (see page 8, third paragraph of Discussion). In our revised manuscript, we have ensured that the results and discussion sections thoroughly address the negative effects associated with CR and highlight the positive outcomes observed with 1D and 2D IF (see below). However, due to the word count constraints, it is not feasible to include these details in the abstract.

1) 40% CR causes a strong loss in lean mass decrease (15%) but 20% CR also causes loss in lean mass already at 10 months maintained until 34 months whereas 1D and 2D show a remarkable increase in lean mass at both 22 and 34 months. By 34 months the lean mass of 1D and 2D is about 40% higher compared to 40% CR. Furthermore, the increase or stabilization in lean mass between 22 and 34 months in the IF groups is impressive but barely mentioned.

We agree that the body composition effects distinguish the CR and IF interventions. We highlight the complex and distinct associations of LTM with lifespan in (see page 5: *higher LTM was associated with shorter lifespan in IF mice but with longer lifespan in 40% CR mice*). A possible interpretation of this finding is that a progressive increase in organ size other than muscle mass may negatively impact lifespan.

2) While 1D and 2D IF increase lean mass they still show a reduction in fat mass. It would be good to add a lean/fat loss figure

In the previous revision, we expanded the presentation of various measures of adiposity to include fat tissue mass (FTM), lean tissue mass (LTM), total tissue mass (TTM = FTM + LTM), and adiposity (FTM/TTM). DEXA analysis provides two independent measures of body composition (TTM and adiposity) and the rest are calculated. We examined all reasonable combinations and selected the four representations above because they each revealed unique aspects of intervention response. We found that the ratios LTM/FTM and FTM/LTM were redundant with adiposity and chose to leave them out of the displayed data. Both traits were analyzed and the results are summarized in Extended Data Tables 6 and 7.

3) The fact, that 1D and 2D are achieving lifespan extension with no or minimal calorie restriction should be emphasized since now we know that we is being compared is not equivalent, meaning 2 groups of mice are truly calorie restricted and in fact even the 20% CR is restricted by 27% whereas the 1D consumes more calories than controls. These are fundamental issues and should be front and center, especially because a 27-43% CR below normal calorie intake in humans would not even be considered, also for the reasons shown below.

It is now clear from our own work and that of others (References 2 - 6) that both the timing of feeding and reduced caloric intake play a role in the lifespan extension effects of DR. At the time we designed our study this was an unresolved question. There are several practical constraints to implementing DR in mice. It remains to be seen if someone can construct a DR intervention strategy that cleanly separates caloric intake from timing. In our study the two effects are partially confounded, i.e., 2D IF have a reduced calorie intake, and both CR groups experience daily and weekly periods of fasting. (See Page 3, line 29: *Notably, 1D IF mice experienced an extended median lifespan without a reduction in net caloric intake*.)

4) The main figure show a decrease in effector CD4 cells in 40% CR but the extended figure 7 shows a major reduction in B cells, monocytes and NK cells in 40% CR. Whereas these changes may not be consequential in the sterile/protected vivarium environment they could have major consequences in the wild. This is another key part of the data that needs to be included in the abstract and much more clear in the manuscript.

Agreed, the B cells, monocytes, and NK cells are the most dramatic cases across the FLOW traits where 40% CR appears to exacerbate age related changes. See new text in the Abstract and Discussion (third paragraph) that emphasizes the potential for increased susceptibility to infection.

This is a hypothesis that remains to be tested.

Referee #2 (Remarks to the Author)

I applaud the thorough job done on revising this study. Two minor things to be addressed:
(1) the website listed under the QTL analyses section did not work (https://churchilllab.jax.org/qtlviewer/JAC/DOAgingIntervention), it should be changed.
(2) please explain how the threshold of QTL mapping was calculated.

Johan Auwerx

Thank you.

The website address has been corrected to https://churchilllab.jax.org/qtlviewer/DRiDO.

Explanation of QTL thresholds added to **Methods**: *QTL mapping*.

Referee #3 (Remarks to the Author)

Second Review of. “Regulators of health and lifespan extension in genetically diverse mice on dietary restriction”

The authors have done an excellent job responding to my and, from what I can gather, the other reviewers’ concerns. The emphasis on Dr and IF findings that are statistically significant and biologically meaningful improve the value of the paper. The big take away messages to my mind are the complex nature of dietary restriction, that genetics is more important than an DR/IF intervention for lifespan, the fact that metabolic phenotypes are not key players in lifespan, 'resilience to stress' is a key predictor of lifespan, and the obvious distinctions between humans and mice wrt to longevity enhancement. I believe the authors have responded adequately to my individual concerns, but there is still an overall potential for false positives across all the assays and analyses that would be hard to quantify. I found the revised manuscript much more readable and better organized, laying out the issues and analyses so that readers would know that many are ‘hypothesis generating’ or upset current thinking in some way but are not pivotal.

Thank you for providing a succinct summary of the results. We think you hit all the main points and have reiterated these points in the revised **Discussion**. The initial review comments were helpful in pushing us to reorganize the presentation of the analysis – the comment here is appreciated.

One minor issue is with the mediation analysis. I had brought this up in my previous review and I think the authors responded appropriately, but their response raises an additional question which concerns what the processes might be that the immune cells are acting as surrogates for? Obviously, there is no way of knowing this, but, for example, what other phenotypes did the immune cells correlate with? If DR/IF do affect immune cells, and this affect leads to downstream changes resulting in lifespan, then of course one could look for other ways of modulating those immune cells – beyond DR and IF – with the belief that this will affect lifespan. Therefore 1-2 sentences on insight into what those immune cell parameters may have correlated with that may get at this would be of value. It was hard to go through the various results to pick this out easily (e.g., do the immune cell parameters correlate with lack of infections, a better balance in cell populations, diversity in cell populations, effect on other processes like cellular communication, etc.?).

In addition to lifespan, the FLOW variables are inter-correlated and correlated with CBC variables that also measure immune cell populations. We agree that it is highly unlikely that direct intervention on immune cell populations would extend lifespan. (see revised text, page 7: *Although the selected traits are unlikely to be direct causal mediators of the DR effects on lifespan, they serve as surrogates for underlying mechanisms that mediate the effects of DR. Identification of these mechanisms could provide new targets for interventions*). The age-related changes in lymphocyte proportion are likely due to changes in underlying hematopoietic stem cell population (PMID: **21664115). This is an active area of research, and the mechanisms remain to be elucidated.**