nature portfolio

Peer Review File

A metabolomic profile of biological aging in 250,341 individuals from the UK Biobank

Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

This study identified aging-related NMR biomarkers and investigated their associations with various aging-related health outcomes utilizing data from an extensive cohort of 250,341 individuals enrolled in the UK Biobank. Subsequently, the authors constructed a metabolomic aging score and compared it with previously established metabolomics-based scores and other aging measures. Next, they investigated the personalized metabolomic changing patterns and identified potential anti-aging and pro-aging NMR biomarkers based on longitudinal data from 13,263 individuals. This study offers a comprehensive examination of the potential metabolomic profile of biological aging within a large-scale population. However, several concerns warrant attention:

1. Currently, there have been numerous studies on biological age and aging within the realm of metabolomics [1-3]. The authors should make proper comparisons (quantitative/qualitative), and state the novelty clearly of this study over the previous works. In particularly, it is necessary to further elucidate the advancements in terms of methodology and research findings.
Reference:
[1] Buergel, Thore, et al. "Metabolomic profiles predict individual multidisease outcomes." Nature Medicine 28.11 (2022): 2309-2320.
[2] Talmor-Barkan, Yeela, et al. "Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease." Nature medicine 28.2 (2022): 295-302.
[3] Deelen, Joris, et al. "A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals." Nature communications 10.1 (2019): 3346.
2. The importance of this study lies in establishing the ground truth of biological aging. However, the term "an overall health indicator" mentioned in Line 78 lacks clarity. Further clarification is needed to delineate how this indicator is determined.
3. Furthermore, the ground truth encompassed the date of death, a variable strongly correlated with chronological age. It is imperative to determine whether the authors adjusted for chronological age as a covariate in subsequent analyses and validated the independence of these identified metabolomic factors in relation to biological aging. Failure to correct for age in subsequent analyses could undermine the ability to ascertain whether the identified metabolomic factors exert an independent effect on biological age.
4. The results section elucidates extensive comparisons with some previous studies, showing a significant overlap in many metabolomic indicators. However, the aspects where there is no overlap require explanation for their absence. Currently, the authors merely list these discrepancies, such as the contents within lines 110-122. It is necessary for the authors to further analyze and discuss the reasons for inconsistencies with previous work, such as whether they may be due to errors between different measurement platforms rather than their impact on biological aging mechanisms.
5. The clinical significance of both the Metabolomic Aging Score and the change in Metabolomic Aging Score (ΔMet) remains ambiguous, necessitating further clarification. It is imperative to elucidate the distinctions between them and how they individually reflect aging status. Additionally, clarification is needed on whether they should be combined to assess one's aging or if focus should be placed solely on one of them. Furthermore, it is crucial to investigate whether there are variations in their significance among populations with different aging statuses.
6. In the prediction of mortality, the Metabolomic Aging Score does not exhibit an advantage when

compared to chronological age. Instead, it only demonstrates an advantage in diagnosing specific diseases. Does this limitation imply that the score inadequately reflects overall systemic aging or different stages of aging?
7. Furthermore, with the extension of follow-up time, the advantage of the Metabolomic Aging Score diminishes while chronological age outperforms. Does this suggest a limitation in the applicable aging period for this score? Further clarification is required regarding its clinical significance.
8. Is there a difference in the performance of the metabolomic score across different chronological age groups? Does it only hold significance in populations with older chronological ages?
9. Participants were categorized into three biological age groups based on various aging measures and were further classified into early-onset patients, late-onset patients, and disease-free individuals for analysis. The aging and disease information, along with distribution characteristics among these subgroups, need to be provided to assess potential biases in the analysis.
10. The impact of the research findings and the clinical significance of the identified metabolic markers, as well as the subsequent constructed score and rate, need to be clearly discussed, respectively.
11. The discussion section requires further specificity regarding the distinct findings or novel significance uncovered in this study compared to other metabolomic studies. Particularly, it should delve into the reasons behind the differences observed in this study, rather than merely reiterating the results sections.
12. The age range and racial diversity of the population are limited in this study. It remains unclear to what extent the metabolomic profile and the constructed aging score are applicable outside the studied population.

Reviewer #2 (Remarks to the Author):

In this work, the authors investigated the metabolomic profile of biological aging in 250,341 individuals from the UK Biobank, uncovering significant associations and causal links between aging-related metabolomic biomarkers and multiple aging-related adverse health outcomes. Furthermore, they developed an innovative Metabolomic Aging Score, which demonstrated exceptional performance in predicting mortality risk and effectively identifying populations with accelerated aging compared to other aging measures.
1. In lines 67-69, "we aimed to identify aging-related metabolomic biomarkers out of 325 biomarkers (with 76 additional biomarker ratios not available in the original data but of potential biological implications)" is mentioned. Clarifying the selection criteria and the rationale behind these specific ratios would help understand their significance. What is the definition of biomarker in the manuscript? What does "325 biomarker" refer to?
2. What criteria were used to categorize the cohort into "young, middle, and old" groups? Please provide a detailed explanation.
3. The authors employed LASSO Cox regression modeling for biomarker selection, with all-cause mortality as the endpoint of interest. However, the correlation between the selected metabolites and age has not been addressed.
4. In lines 324-325, what method was employed to calculate the correlation between ΔMet and ΔFI? Does a correlation coefficient of 0.06 indicate a significant correlation?
5. For the anti/pro-aging biomarkers, a deeper discussion on the biological relevance, their

interconnections, and the physiological pathways they implicate in aging and age-related diseases could enrich the study's contribution to understanding aging mechanisms.
6. More related advances (e.g. VIEW 2023, 4, 20220038) should be included and discussed.
7. Throughout the manuscript, abbreviations used in figures and tables should be spelled out in their legends (e.g., Fig. 6).
8. For Fig. 3, please specify the significance of the data represented in the legend.

Reviewer #3 (Remarks to the Author):

The paper develops a new metabolite based clock to measure survivor in first instance and then is tested against many different outcomes.

I think overall the paper is ok in the part where the model for the clock is trained as it does seem to improve on other existing metrics in an out of sample dataset.

Where the paper falls short is in the second part where the focus is all on the metabolites chosen to be included in the model.
There are in this respect a several questions/issues which would need addressing:
1) The level of overlap with previous studies in a bit worrying as it shows lack of replication. How much do the previous scores correlate with the newly developed one?
2) Another issue is the treatment of the high correlation between the metabolites which make choice between one or the other subject to be influenced by the background noise. In this context it is possible that the chosen metabolites may be the most predictive but not necessarily the causal ones. This will make the resulting model formally correct but makes interpreting results difficult. Given the sometimes close to one correlation between multiple metabolites needs proper treatment. There are many choices: the first is to select the metabolite most representative of a group of tightly correlated traits as marker of the whole group; the second is to use a variable reduction technique to find a latent variable that represents the whole group.
I don’t think that in this case the conclusions that the biomarkers could be causal are justified and the MR analysis should have not been limited just to the selected representative metabolites. This may also explain the discrepancy between the selected metabolites and those used previously.
3) It is unclear to me why the authors did not just rerun the GWAS analysis on the metabolites in UK biobank instead of relying only on existing summary statistics for less metabolites (42). The summary statistics are there why not use them?
4) It is unclear to me why the authors have decided that 0.05 was a reasonable p-value threshold for the MR analysis as they have performed many tests. They have applied multiple test correction in other parts of the paper why not here?
5) The colocalization analysis is puzzling as they report PPH4 for the each result, however it is unclear to which SNP it refers to given each analysis has used multiple SNPs. Did they test for colocalization each pair of locus-trait? How was the multivariable framework handled in this case?
6) MR-Egger tests only a specific type of pleiotropy (directional) and can be used but just with the aim of testing this type of assumption violations. It would have been better to use other types of approaches (i.e. MR-median).

In conclusion I think that overall the paper is relatively sound in the score building and testing part but much weaker on the interpretation part and needs to be revied especially in its interpretation.

Response to reviewers' comments

Reviewer #1:

This study identified aging-related NMR biomarkers and investigated their associations with various aging-related health outcomes utilizing data from an extensive cohort of 250,341 individuals enrolled in the UK Biobank. Subsequently, the authors constructed a metabolomic aging score and compared it with previously established metabolomics-based scores and other aging measures. Next, they investigated the personalized metabolomic changing patterns and identified potential anti-aging and pro-aging NMR biomarkers based on longitudinal data from 13,263 individuals. This study offers a comprehensive examination of the potential metabolomic profile of biological aging within a large-scale population. However, several concerns warrant attention:

Response: We thank the reviewer for providing a comprehensive summary of our research.

Comment 1: Currently, there have been numerous studies on biological age and aging within the realm of metabolomics [1-3]. The authors should make proper comparisons (quantitative/qualitative), and state the novelty clearly of this study over the previous works. In particularly, it is necessary to further elucidate the advancements in terms of methodology and research findings. Reference: [1] Buergel, Thore, et al. "Metabolomic profiles predict individual multidisease outcomes." Nature Medicine 28.11 (2022): 2309-2320. [2] Talmor-Barkan, Yeela, et al. "Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease." Nature medicine 28.2 (2022): 295-302. [3] Deelen, Joris, et al. "A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals." Nature communications 10.1 (2019): 3346.

Response: We thank the reviewer for providing these references. In our revised manuscript, we have implemented the following changes:

(1) Included qualitative comparisons between aging-related biomarkers in our study and those of previous publications; we also discuss potential reasons for differences between studies. Please refer to our response to **Comment 4**.

(2) Included quantitative comparisons between *MetaboHealth* (the metabolomic score predicting all-cause mortality in the third reference listed above) and our *Metabolomic Aging Score* to investigate their differences in all-cause mortality prediction across different follow-up intervals.

> ***Results, page 11, line 237-240:*** "The *Metabolomic Aging Score* in our study was highly correlated with *MetaboHealth* (Pearson's r=0.68) and was moderately correlated with chronological age (r=0.29) and the frailty index (r=0.32), while had the weakest correlation with LTL (r=0.12) (p-values<5E-7, Extended Figure 6)."
>
> ***Results, page 11, line 245-246***: "Compared to *MetaboHealth*, the frailty index and LTL, the *Metabolomic Aging Score* had the highest accuracy in mortality risk prediction across all follow-up intervals (Figure 3a-b)."
>
> ***Results, page 12, line 256-260***: "The residuals of the four biological aging metrics regressed against chronological age displayed slightly reduced predictive performance across seven follow-up intervals (1y, 2y, 3y, 4y, 5y, 10y and 15y). Still, the residuals of the *Metabolomic Aging Score* outperformed others and had similar prediction accuracy as chronological age across one-year to five-year intervals (*p*-values of 0.32, 0.32, 0.67, 0.53 and 0.36, respectively) (Extended Figure 7)."

(3) Highlighted the significance and potential clinical application of the *Metabolomic Aging Score* in short-term mortality risk prediction, identification of individuals experiencing accelerated aging and prediction of aging-related diseases.

> ***Discussion, page 22-23, line 486-498***: "The characteristics of the plasma metabolome provide this score with prospective clinical value, particularly in monitoring high-risk populations such as frail seniors for whom complex physiological examinations may be challenging[67]. Additionally, it serves as a tool for detecting subtle metabolomic changes indicative of pathological patterns, facilitating early interventions[68]. Our subsequent analyses revealed the *Metabolomic Aging Score* as a significant factor in discriminating future early-onset patients of multiple aging-related diseases who exhibited an accelerated pace of aging compared with their peers[69], even after adjusting for chronological age and other potential confounders. Importantly, this score offered additional and complementary predictive signals beyond traditional risk factors indicative of aging-related disease risks.
>
> Hence, we proposed the potential application of the *Metabolomic Aging Score* in several clinical scenarios, including monitoring short-term mortality risk, discriminating aging-accelerated populations, and enhancing disease risk prediction when combined with traditional risk factors."

(4) Highlighted the novelty and methodological strengths compared to previous research (References 1 and 2 listed above), for example examining potential causal relationships linking the NMR biomarkers to disease onset, instead of focusing only on observational associations.

> ***Discussion, page 20-21, line 444-453***: "In contrast to previous metabolomics-based studies that focused narrowly on selected biomarkers' associations with specific diseases or their contribution to predictive performance[16,17], our study harnessed WGS data from 95,372 individuals alongside comprehensive NMR metabolomic profiles. This enabled us to characterize the genetic architecture of the plasma metabolome[60], yielding 325 NMR GWAS summary statistics for downstream analysis. Through MVMR analysis, we discovered 197 candidate biomarker-disease causal pairs, connecting 136 NMR biomarkers with 20 aging-related diseases. Among them, 85 pairs were further supported by colocalization analysis, providing further insights into how these risk or protective biomarkers may affect disease onset. Moreover, the GWAS summary statistics for 325 NMR biomarkers provide a repository for future exploration[61]."

Comment 2: The importance of this study lies in establishing the ground truth of biological aging. However, the term "an overall health indicator" mentioned in Line 78 lacks clarity. Further clarification is needed to delineate how this indicator is determined.

Response: Thank you for bringing this to our attention. In our original manuscript, "an overall health indicator" referred to all-cause mortality. Aging is the progressive deterioration in physiological functions, leading to an age-dependent decrease in rate of survival. At the individual level, death is the ultimate outcome of aging, and at the populational level, mortality rates often indicate overall health trends and provide value insights into broader health patterns and risks. Thus, we described it as an overall health indicator. We have modified the wording and added further clarifications in the revised manuscript.

> ***Results, page 4-5, line 85-92***: "To identify metabolomic biomarkers representative of biological aging among 325 highly-correlated NMR biomarkers, we developed a Least Absolute Shrinkage and Selection Operator (LASSO) Cox proportional hazards model[22] with all-cause mortality as the predicted outcome. At the individual level, aging is a ubiquitous biological process accompanied by loss of physiological functions which ultimately leads to death[23]; at the populational level, mortality rates often indicate overall health trends and provide value insights into broader health patterns and risks[24]. Thus, all-cause mortality risk was used as a measure of global aging status here and further as a ruler to compare different aging metrics[5]."

Comment 3: Furthermore, the ground truth encompassed the date of death, a variable strongly correlated with chronological age. It is imperative to determine whether the authors adjusted for chronological age as a covariate in subsequent analyses and validated the independence of these identified metabolomic factors in relation to biological aging. Failure to correct for age in subsequent analyses could undermine the ability to ascertain whether the identified metabolomic factors exert an independent effect on biological age.

Response: Thank you for your insightful comment. In our revised manuscript, we have clearly described any adjustments for chronological age in the analyses:

(1) In the analysis exploring the associations between 54 aging-related biomarkers with multiple aging-related frailty phenotypes, we included chronological age as a covariate.

> ***Results, page 6, line 132-134***: "The 54 aging-related representative biomarkers predictive of all-cause mortality were further investigated by testing their associations with multiple aging-related frailty deficits[30]. Chronological age was included as a covariate in multivariable logistic regression models."

(2) In the analysis evaluating the performance of different aging metrics at predicting all-cause mortality, we regressed the four biological aging measures against chronological age and took the residuals as orthogonal components independent of chronological age for downstream analyses.

> ***Results, page 12, line 252-260***: "Considering the correlations of the four biological aging metrics with chronological age, we regressed them each against chronological age and extracted the residuals to investigate whether they retained predictive information independent of chronological age (Supplementary Table 14).
>
> The residuals of the four biological aging metrics regressed against chronological age displayed slightly reduced predictive performance across seven follow-up intervals (1y, 2y, 3y, 4y, 5y, 10y and 15y). Still, the residuals of the *Metabolomic Aging Score* outperformed others and had similar prediction accuracy as chronological age across one-year to five-year intervals (*p*-values of 0.32, 0.32, 0.67, 0.53 and 0.36, respectively) (Extended Figure 7)."

In the age-stratified analysis, we also focused on the residuals against chronological age to examine their associations independent of chronological age (***Results, page 12, line 261-271***).

(3) In the analysis where we used Cox PH models to estimate hazard ratios for all-cause mortality.

Results, page 13, line 272-274: "The residuals of the *Metabolomic Aging Score* were a significant risk factor of short-term all-cause mortality. Age-stratified analysis further confirmed this observation (Extended Figure 8)."

(4) In the analysis identifying future early-onset patients of aging-related diseases using the *Metabolomic Aging Score* we adjusted for chronological age and other confounders. Please see our response to **Comment 9**.

(5) In the longitudinal analysis exploring the association of the *Metabolomic Aging Rate* with mortality risk.

Results, page 17, line 364-369: "Among the 13,263 individuals with revisit data, the residuals of the *Metabolomic Aging Rate* regressed against the baseline score were a significant risk factor (HR=1.46 per S.D., 95%CI: 1.37-1.55, *p*-value < 2.2E-16) for all-cause mortality, after adjusting for chronological age. Next, we conducted age-stratified analysis to explore potentially varied associations between the *Metabolomic Aging Rate* and all-cause mortality risk across three age groups (40-50, 51-60 and 61-70 years) while adjusting for chronological age."

Comment 4: The results section elucidates extensive comparisons with some previous studies, showing a significant overlap in many metabolomic indicators. However, the aspects where there is no overlap require explanation for their absence. Currently, the authors merely list these discrepancies, such as the contents within lines 110-122. It is necessary for the authors to further analyze and discuss the reasons for inconsistencies with previous work, such as whether they may be due to errors between different measurement platforms rather than their impact on biological aging mechanisms.

Response: Thank you for your insightful comments. In our revised manuscript, we have discussed three reasons for non-replicated aging-related NMR biomarkers, including (1) different sources of sampling (plasma or serum), (2) different cohort characteristics (including socioeconomic and health features), and (3) different profiling background and noise in model training.

Discussion, page 21, line 457-473: "Despite the replication of certain findings, there were several differences from previous studies, particularly related to lipoproteins, which might be attributed to: (1) different sampling sources, e.g., NMR metabolomic biomarkers in the UK Biobank were measured from EDTA plasma samples, whereas *MetaboAge* utilized serum metabolomics[19] and *MetaboHealth* employed a

mixture of EDTA plasma and serum samples from various subcohorts[18]. Previous studies have indicated that the source of sampling, whether plasma or serum, can impact metabolomic profiling[62]; (2) different cohort characteristics: individuals included in our study were between 40 to 70 years old, whereas the studies used for *MetaboHealth* and *MetaboAge* had a broader baseline age range spanning from 18 to 109 years. Additionally, differences in other health or socioeconomic features might also impact results; (3) different profiling backgrounds and noise in model training: although NMR biomarkers in *MetaboHealth* and *MetaboAge* were measured using the same platform as in the UK Biobank, there were disparities in the number and variety of measured biomarkers. Moreover, due to the high collinearity among NMR biomarkers, the selection of aging or mortality-related biomarkers during model training may be influenced by the metabolomic background unique in each study[63]. Non-replicated aging-related biomarkers across different studies could, in fact, provide concordant and shared predictive information because of their inherent correlation[63]."

Comment 5: The clinical significance of both the Metabolomic Aging Score and the change in Metabolomic Aging Score (ΔMet) remains ambiguous, necessitating further clarification. It is imperative to elucidate the distinctions between them and how they individually reflect aging status. Additionally, clarification is needed on whether they should be combined to assess one's aging or if focus should be placed solely on one of them. Furthermore, it is crucial to investigate whether there are variations in their significance among populations with different aging statuses.

Response: Thank you for your insightful feedback pointing out the lack of (1) discussion on their clinical significance, (2) clarification of their differences and (3) stratified analyses across different aging-status groups. We have included several revisions and updated analyses in the manuscript:

(1) Clarification of the clinical significance and potential applications. We discussed the potential clinical value and application of the *Metabolomic Aging Score* given its superior performance over chronological age and other biological aging metrics in short-term mortality risk prediction, its discrimination of future early-onset patients of multiple aging-related diseases, and its improvement over traditional risk factors in disease risk prediction. (Please refer to the response to ***Comment#1, the third point.***)

(2) Discussion of the distinctions between the *Metabolomic Aging Score* and the *Metabolomic Aging Rate*.

In our updated analysis, we primarily focused on the clinical significance of *Metabolomic Aging Rate* (defined as ΔMet/ Follow-up intervals) which we believe provide more refined metabolomic

information on the rate of change towards the aging pathology. We removed the analysis on the ΔMet and made a clear distinction between the *Metabolomic Aging Score* and the *Metabolomic Aging Rate*.

> ***Discussion, page 23, line 505-513***: "While the *Metabolomic Aging Score* aided in discerning biologically older individuals within a peer group, the *Metabolomic Aging Rate* detected subtler differences in the rate of change in the aging-related metabolomic profile among individuals with similar *Metabolomic Aging Scores*. Drawing an analogy to the concepts of distance and speed in physics, the combination of both can estimate how far a person will reach. We believed that combining the *Metabolomic Aging Score* (reflecting current distance along the route of biological aging) with *Metabolomic Aging Rate* (reflecting current speed of biological aging) would yield more predictive power and insights into personal aging status and future disease or mortality risk[71]."

(3) Stratified analysis across different chronological age groups. We performed age-stratified analysis for three age groups as 40-50, 51-60, 61-70 to investigate whether the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* had differential performance across these groups. Please see our response to **Comment 8** for further details.

Comment 6: In the prediction of mortality, the Metabolomic Aging Score does not exhibit an advantage when compared to chronological age. Instead, it only demonstrates an advantage in diagnosing specific diseases. Does this limitation imply that the score inadequately reflects overall systemic aging or different stages of aging?

Response: Thank you for raising this important point. It is worth noting that in short-term mortality risk prediction with follow-up intervals extending from one to five years, the *Metabolomic Aging Score* outperformed chronological age. We have discussed potential reasons for the improved prediction of short-term vs long-term mortality risk.

> ***Discussion, page 22, line 478-485***: "It is noteworthy that the *Metabolomic Aging Score* demonstrated optimal predictive capability for short-term (one to five years) mortality risk, surpassing chronological age and yielding the highest accuracy. This superiority might stem from the dynamic nature of metabolomic profiles, which reflects immediate influences from intrinsic and extrinsic factors on current health status[64]. Cellular metabolic reactions undergo rapid changes, while the resulting products or waste built up across the body because of delayed or impaired clearance over the course of aging[65]. However, the predictive signals carried by these biomarkers may decay due to the plethora of changes that occur over longer periods of follow-up[66]."

In terms of disease risk prediction and early-onset patient discrimination, the *Metabolomic Aging Score* exhibited generally good performance, albeit with some degree of disease-specificity, as the reviewer noted. This finding is consistent with previous studies exploring the potential advantages and limitations of metabolomics for disease prediction (PMID:36138150). It is plausible that the *Metabolomic Aging Score* primarily reflects aging-related changes more intimately linked to the plasma metabolome. Aging is a complicated and multidimensional progress. Instead of using a single biological age measure, combining metrics from multiple domains, for example metabolomic aging scores, epigenetic aging scores, and proteomic aging scores, will provide more comprehensive and in-depth aging profiles. We have discussed this limitation in the revised manuscript and called for future combinations of *Metabolomic Aging Score* with other aging metrics.

> ***Discussion, page 24, line 531-536***: "Lastly, aligning with earlier research into the strengths and limitations of metabolomics for disease risk prediction, our findings underscored that the predictive power of the plasma metabolome exhibited some degree of disease specificity[16]. It is plausible that the metabolic fingerprint varied in its contribution to different disease mechanisms[44], which underscored the challenge of applying a one-size-fits-all approach in metabolomics-based risk prediction[5]."

> ***Discussion, page 24-25, line 539-544***: "However, it is important to note that our intention in devising this score was not to recommend a singular authoritative metric of biological aging. Instead, this score captures aging-related signal at the metabolome level. Given the multifaceted nature of aging[1], future research should integrate diverse aging-related metrics from multiple dimensions, for example combining proteomic aging scores[77] and epigenetic aging scores[78] with the *Metabolomic Aging Score* to unveil a more comprehensive profile of aging[5]."

Comment 7: Furthermore, with the extension of follow-up time, the advantage of the Metabolomic Aging Score diminishes while chronological age outperforms. Does this suggest a limitation in the applicable aging period for this score? Further clarification is required regarding its clinical significance.

Response: We agree that the superior prediction of short-term mortality risk (one year to five years) narrows the scope of application of the *Metabolomic Aging Score*. We have discussed potential reasons for better predictive performance of short-term than long-term risk and stated clearly that its clinical significance primarily relates to short-term mortality risk prediction. Please see our response for Comment 6.

Comment 8: Is there a difference in the performance of the metabolomic score across different chronological age groups? Does it only hold significance in populations with older chronological ages?

Response: Thank you for this question. We have conducted additional age-stratified analyses to investigate the predictive performance of the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* across different chronological age groups.

(1) Age-stratified analyses of the *Metabolomic Aging Score*.

> ***Results, page 12, line 261-271***: "We next conducted age-stratified analysis to explore the predictive performance of the *Metabolomic Aging Score* across different age groups (40-50, 51-60 and 61-70 years). A 10-year age span was considered sufficient to allow for differences in physiological and aging status between groups. Within the 40-50 age group, the residuals of the *Metabolomic Aging Score* had similar predictive performance as chronological age with no significant difference in AUCs across all the follow-up intervals (smallest *p*-value=0.43). Within the 51-60 and 61-70 age groups, the residuals of the *Metabolomic Aging Score* outperformed chronological age in mortality risk prediction across all follow-up intervals (*p*-values <0.05 except for *p*=0.069 in the 61-70 group with one-year interval). It had the best predictive performance in the 51-60 group with an AUC of 86.8% for one-year mortality risk, whereas chronological age had an AUC of 56.0% (Figure 3c-d, Supplementary Table 15)."

> ***Results, page 13, line 272-279***: "The residuals of the *Metabolomic Aging Score* were a significant risk factor of short-term all-cause mortality. Age-stratified analysis further confirmed this observation (**Extended Figure 8**). The residual of the *Metabolomic Aging Score* exhibited the greatest impact in the 51-60 age group, with a one-year mortality HR per S.D. of 3.5 (95%CI: 2.6-4.8) and a 15-year mortality HR per S.D. of 1.9 (95%CI: 1.8-2.1), followed by the 61-70 age group with a one-year mortality HR per S.D. of 2.3 (95%CI: 1.8-2.9) and a 15-year mortality risk HR per S.D. of 1.6 (95%CI: 1.5-1.7), and the 40-50 age group with a one-year mortality HR per S.D. of 1.8 (95%CI: 0.6-5.2) and 15-year mortality HR per S.D. of 1.6 (95%CI: 1.3-1.8)."

(2) Age-stratified analyses of the *Metabolomic Aging Rate*.

> ***Results, page 17, line 367-372***: "Next, we conducted age-stratified analysis to explore potentially varied associations between the *Metabolomic Aging Rate* and all-cause mortality risk across three age groups (40-50, 51-60 and 61-70 years) while adjusting for chronological age. Compared with those in the 40-50 age group, the residuals of the *Metabolomic Aging Rate* exhibited stronger associations with mortality risk

among those in the 51-60 and 61-70 age groups (p-values of 1.07E-02, 2.79E-15 and 8.06E-19, with HR per S.D. of 1.42, 1.53 and 1.43, respectively) (Extended Figure 12)."

Through the age-stratified analyses, we found that the *Metabolomic Aging Score* was more predictive of mortality risk among those in the 51-60 and 61-70 age groups, compared to those in the 40-50 age group. The *Metabolomic Aging Rate* exhibited similar associations. One possible reason for this might be the relative underrepresentation of those in their 40-50s in the UK Biobank, resulting in an unbalanced distribution of the three age groups during model training (the whole sample size for each group as 40-50, 51-60 and 61-70, respectively: 65,530, 88,001 and 96,803). We added a relevant section to the discussion in the revised manuscript.

Discussion, page 23-24, line 517-521: "However, there are certain limitations to our current study. First, participants included in our study aged between 40 to 70 years at the first recruitment, with a predominant proportion falling within the 50-70 age group. The limited age range and imbalanced distribution among age groups might introduce potential bias in model training[73] and diminish the generalizability of the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* to broader populations."

Comment 9: Participants were categorized into three biological age groups based on various aging measures and were further classified into early-onset patients, late-onset patients, and disease-free individuals for analysis. The aging and disease information, along with distribution characteristics among these subgroups, need to be provided to assess potential biases in the analysis.

Response: We have examined the distribution of several health and socioeconomic characteristics in the early-onset, late-onset and disease-free groups for each disease group, and provide relevant summary statistics in ***Supplementary Note 2***. We have subsequently adjusted potential confounders (including baseline age, sex, BMI, systolic blood pressure, Townsend Deprivation Index, alcohol taking frequency, and smoking status) in multinominal logistical regression analyses. After controlling for these confounders, the *Metabolomic Aging Score* remained a significant prognostic factor to discriminate future early-onset, late-onset and disease-free groups of multiple aging-related diseases.

Results, page 14, line 305-313: "Between-group differences in the distribution of several basic physiological and socioeconomic characteristics (including chronological age, sex, BMI, systolic blood pressure, Townsend Deprivation Index, alcohol taking frequency, and smoking status) were noticed for each disease (Supplementary Note 2). To address potential confounding effects derived from these baseline differences, we conducted a multinomial logistical regression with each disease status ("Disease-free",

"Early-onset" and "Late-onset") as the dependent variable, the above-mentioned confounders as covariates, and the *Metabolomic Aging Score* as the independent variable. The baseline *Metabolomic Aging Score* remained a significant factor distinguishing future early-onset, late-onset, and disease-free groups (Supplementary Table 16)."

Comment 10: The impact of the research findings and the clinical significance of the identified metabolic markers, as well as the subsequent constructed score and rate, need to be clearly discussed, respectively.

Response: Thank you for your valuable feedback. We have put more emphasis on the value and biological functions of the 54 aging-related biomarkers, and subsequently highlighted the importance of the 325 GWAS summary statistics revealing the genetic architecture of the plasma metabolome which could facilitate future studies. Potential clinical applications of the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* are discussed more clearly in the revised manuscript as well.

Discussion, page 20, line 435-443: "Here, we have presented the largest aging-related metabolomic profile to date based on 325 NMR biomarkers from 250,341 individuals recruited in the UK Biobank. 54 aging-related representative metabolomic biomarkers were identified based on their prediction of all-cause mortality. These aging-related biomarkers are involved in diverse biological functions and multiple metabolic pathways[57], which might serve as potential anti-aging intervention targets and facilitate further exploration of the mechanism of aging-related metabolic disorders. High-resolution analysis of the refined composition and structure of multiple lipoprotein-related biomarkers, thanks to NMR profiling[58], contributes greatly to unraveling the intricate roles of lipid metabolism in the process of aging[59]."

Discussion, page 20-21, line 444-453: "In contrast to previous metabolomics-based studies that focused narrowly on selected biomarkers' associations with specific diseases or their contribution to predictive performance[16,17], our study harnessed WGS data from 95,372 individuals alongside comprehensive NMR metabolomic profiles. This enabled us to characterize the genetic architecture of the plasma metabolome[60], yielding 325 NMR GWAS summary statistics for downstream analysis. Through MVMR analysis, we discovered 197 candidate biomarker-disease causal pairs, connecting 136 NMR biomarkers with 20 aging-related diseases. Among them, 85 pairs were further supported by colocalization analysis, providing further insights into how these risk or protective biomarkers may affect disease onset. Moreover, the GWAS summary statistics for 325 NMR biomarkers provide a repository for future exploration[61]."

For a discussion of the clinical significance of the *Metabolomic Aging Score* and the *Metabolomic Aging Rate*, please see our response to **Comment 5**.

Comment 11: The discussion section requires further specificity regarding the distinct findings or novel significance uncovered in this study compared to other metabolomic studies. Particularly, it should delve into the reasons behind the differences observed in this study, rather than merely reiterating the results sections.

Response: Thank you for your constructive suggestions. Following your previous suggestions described above, especially **Comments 1, 4, 5** and **10**, we have revised the ***Discussion*** section thoroughly to emphasize the novelty and significance of our findings. We have also reviewed possible reasons for the discrepancy between our study and previous ones (please see our responses to **Comments 1** and **4**).

Comment 12: The age range and racial diversity of the population are limited in this study. It remains unclear to what extent the metabolomic profile and the constructed aging score are applicable outside the studied population.

Response: Thank you for your concern about the generalizability of our findings to other populations, which have been well-documented in the UK Biobank (PMID: 37106081). We have highlighted these considerations in the discussion.

> ***Discussion, page 23-24, line 517-526***: "However, there are certain limitations to our current study. First, participants included in our study aged between 40 to 70 years at the first recruitment, with a predominant proportion falling within the 50-70 age group. The limited age range and imbalanced distribution among age groups might introduce potential bias in model training[73] and diminish the generalizability of the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* to broader populations. Second, the underrepresentation of more deprived and less healthy individuals in the UK Biobank is well documented[74]. Thus, application of our findings outside the studied population might warrant caution. Although we used a sub-cohort recruited from Scotland as an out-of-sample dataset for validation given distinct demographic and health-related characteristics across different regions[25], still, the external validation in another cohort is warrant."

Reviewer #2:

In this work, the authors investigated the metabolomic profile of biological aging in 250,341 individuals from the UK Biobank, uncovering significant associations and causal links between aging-related metabolomic biomarkers and multiple aging-related adverse health outcomes. Furthermore, they developed an innovative Metabolomic Aging Score, which demonstrated exceptional performance in

predicting mortality risk and effectively identifying populations with accelerated aging compared to other aging measures.

Response: We thank the reviewer for the thoughtful summary of our work and for highlighting the significance and novelty of our findings.

Comment 1: In lines 67-69, "we aimed to identify aging-related metabolomic biomarkers out of 325 biomarkers (with 76 additional biomarker ratios not available in the original data but of potential biological implications)" is mentioned. Clarifying the selection criteria and the rationale behind these specific ratios would help understand their significance. What is the definition of biomarker in the manuscript? What does "325 biomarker" refer to?

Response: The biomarkers described in our manuscript refer to metabolites measured from EDTA plasma samples using the Nightingale Health NMR platform. The 325 NMR biomarkers in our study included 249 NMR biomarkers available to download from UK Biobank (168 NMR biomarkers in absolute levels and 81 in derived ratios), and an additional 76 biomarker ratios not available in the original data. We derived these 76 biomarker ratios following the guidance of a previous study on quality control and removal of technical variation in the NMR metabolomics data (PMID: 36720882). The selection criteria were: (1) an additional 20 lipid fractions were derived for three more lipoprotein classes (very low density lipoprotein, low density lipoprotein and high density lipoprotein) and for total lipids because the original data only covered fractions for 14 lipoprotein subclasses; (2) the level of total cholesterol was decomposed into free cholesterol and esterified cholesterol and their ratio because these measures might indicate lipoprotein atherogenicity (PMID: 29736097); (3) polyunsaturated fatty acids were decomposed into omega-3 and omega-6 fatty acids and relevant ratios were derived because these two polyunsaturated fatty acids serve distinct biological functions (PMID: 30347877). We have clarified the selection criteria and rationale for including these additional 76 NMR biomarkers in the revised manuscript and in ***Supplementary Table 1*** describing each biomarker in more detail.

> ***Methods, page 25, line 551-565***: "Nuclear Magnetic Resonance (NMR) metabolomics data available in the UK Biobank (updated in July 2023) included 249 metabolomic biomarkers (168 in absolute concentrations and 81 in derived ratios) in EDTA plasma samples from approximately 280,000 participants. Technical variation present in the UK Biobank NMR data was removed using the "ukbnmr" R package[21]. 76 additional biomarker ratios with potential biological implications but not available in the original data were also computed and included in our analysis, which, together, resulted in 325 NMR biomarkers[21]. The inclusion criteria and rationale behind these additional 76 biomarker ratios could be summarized as follows:

(1) Supplementing 20 additional lipoprotein fractions for three lipoprotein classes (low density lipoprotein, very low density lipoprotein, and high density lipoprotein) and for total serum lipids on the basis of original 14 lipoprotein sub-classes provided by UK Biobank; (2) Decomposing total cholesterol into free cholesterol and esterified cholesterol and subsequently deriving more refined ratios for each lipoprotein class and subclass; (3) Decomposing polyunsaturated fatty acids into omega-3 fatty acids and omega-6 fatty acids and subsequently deriving more refined ratios comprised of omega-3 and omega-6 fatty acids. "

Comment 2: What criteria were used to categorize the cohort into "young, middle, and old" groups? Please provide a detailed explanation.

Response: In our revised manuscript, we have added age-stratified analyses of the predictive performance and association with all-cause mortality for both the *Metabolomic Aging Score* and the *Metabolomic Aging Rate* (see **Comment 8, Reviewer 1**). Instead of dividing the sample into "young", "middle" and "old" groups based on the interquartile range, we defined three groups based on their baseline age (40-50, 51-60 and 61-70), in line with previous research (PMID: 34461040). The 10-year age span likely captures differences in physiological and aging status between groups. We have clarified this in the revised manuscript.

Results, page 12, line 261-264: "We next conducted age-stratified analysis to explore the predictive performance of the *Metabolomic Aging Score* across different age groups (40-50, 51-60 and 61-70 years). A 10-year age span was considered sufficient to allow for differences in physiological and aging status between groups."

Comment 3: The authors employed LASSO Cox regression modelling for biomarker selection, with all-cause mortality as the endpoint of interest. However, the correlation between the selected metabolites and age has not been addressed.

Response: Thank you for your comment. To address the correlation between selected NMR biomarkers and chronological age, we conducted the following analyses:

(1) We explored the correlation between each selected biomarker and chronological age.

Results, page 6, line 115-120: "Next, we investigated their correlations with different aging metrics including chronological age, the frailty index (FI) which is a clinical indicator of accumulated health deficits[27], and leukocyte telomere length (LTL) which is a measure of cell division[28]. Among the 54 aging-

related representative biomarkers, 49 were significantly correlated with chronological age, 51 with the FI, and 50 with LTL (BH-adjusted *p*-values < 0.05). There were 35 aging-related biomarkers with consistent correlations with all aging metrics (Extended Figure 3)."

(2) We included chronological age as a covariate in subsequent analyses to validate the independence of selected metabolomic biomarkers in relation to biological aging. Please see our response to **Comment 3 from Reviewer 1**.

Comment 4: In lines 324-325, what method was employed to calculate the correlation between ΔMet and ΔFI? Does a correlation coefficient of 0.06 indicate a significant correlation?

Response: Thank you for your question. In our original manuscript, the Pearson's correlation between ΔMet and ΔFI was calculated. While this coefficient suggests a modest correlation, its statistical significance depends on the sample size. With larger sample sizes, the impact of random error is reduced (PMID: 27747028). The revisit sample size as 13,263 was adequate to indicate strong evidence against the null hypothesis (*p-value*=6.6E-12). In our revised manuscript, we focused primarily on the potential clinical value of the *Metabolomic Aging Rate* (calculated as ΔMet/Follow-up intervals) and removed relevant analysis comparing ΔMet and ΔFI.

Comment 5: For the anti/pro-aging biomarkers, a deeper discussion on the biological relevance, their interconnections, and the physiological pathways they implicate in aging and age-related diseases could enrich the study's contribution to understanding aging mechanisms.

Response: Thank you for your insightful comment. The NMR biomarkers measured in the UK Biobank were lipid-focused, which are not well covered in current metabolite databases for pathway and enrichment analysis. To overcome this limitation and to provide an indication of their biological functions, we have manually retrieved relevant biological knowledge for the non-lipid anti-/pro-aging biomarkers, and we checked the overlap between the remaining lipid-related biomarkers and candidate causal biomarkers for 20 aging-related diseases identified in the MVMR analyses to annotate their disease associations and potential functions. Details for each anti-aging or pro-aging biomarker are provided in **Supplementary Table 18**.

Results, page 18-19, line 404-425: "Out of the 40 pro-aging and anti-aging biomarkers identified here, we manually retrieved relevant biological functions and disease associations for 13 non-lipid biomarkers. For the remaining 27 lipid-related biomarkers characterized by intricate correlations and uncertain biological

roles, we examined their overlap with the candidate causal biomarkers of 20 aging-related diseases identified in the MVMR analysis.

Among the 15 pro-aging biomarkers, GlycA marks the level of inflammatory cytokines in circulation and predicts cardiovascular and severe infection risk[48]. Impaired and dysregulated glycolysis was identified as a cardiovascular disease mechanism[49] and a relevant biological aging process[50], which might account for the pro-aging effects of glucose, lactate and citrate, three metabolites involved in glycolysis. Acetone and 3-Hydroxybutyrate, two ketone bodies whose elevated concentrations in circulation have been linked with ketoacidosis, a complication of uncontrolled diabetes and a significant risk factor of mortality, might exert their pro-aging effects via inducing oxidative stress[51]. Among the 25 anti-aging biomarkers, albumin serves multiple biological functions including inhibiting endothelial apoptosis and protecting against inflammation and oxidative stress[52]. Several lipoprotein-related and polyunsaturated fatty acid-related biomarkers were identified as potential causal metabolomic biomarkers protective against aging-related diseases in the MVMR analyses, including XS_VLDL_FC_pct (free cholesterol to total lipids in very small VLDL percentage) with type 2 diabetes and COPD, PUFA_by_MUFA (polyunsaturated fatty acids to monounsaturated fatty acids ratio) with Parkinson's disease, and LDL size (average diameter for LDL particles) with Alzheimer's disease. The biological role and disease associations of other pro-/anti-aging biomarkers are listed in Supplementary Table 18."

Comment 6: More related advances (e.g. VIEW 2023, 4, 20220038) should be included and discussed.

Response: We have incorporated relevant advances mentioned in the study by Chen et al. and other recently published metabolomics-based research. Please also refer to our response to **Comment 1**, **Reviewer 1**.

Introduction, page 3, line 60-64: "Metabolomics, which integrates intrinsic biological changes with extrinsic exposures[15], carries systemic information across the body[16,17]. The advancement in multiple spectroscopy technologies, for example high-throughput and cost-effective nuclear magnetic resonance (NMR) analysis[18,19] and the application of various machine learning algorithms, have promoted population-scale metabolomics research, with great potential for disease prediction[20]."

Comment 7: Throughout the manuscript, abbreviations used in figures and tables should be spelled out in their legends (e.g., Fig. 6).

Response: Thank you for bringing this negligence to our attention. We have revised the figures and tables accordingly.

Comment 8: For Fig. 3, please specify the significance of the data represented in the legend.

Response: We have revised the legend of Figure 3 to specify the statistical significance of the results. Additionally, we have carefully reviewed all other figures and tables to ensure consistency.

Reviewer #3:

The paper develops a new metabolite-based clock to measure survivor in first instance and then is tested against many different outcomes. I think overall the paper is ok in the part where the model for the clock is trained as it does seem to improve on other existing metrics in an out of sample dataset. Where the paper falls short is in the second part where the focus is all on the metabolites chosen to be included in the model.

Response: We thank the reviewer for the thoughtful assessment of our manuscript. We appreciate the recognition of the strengths of our findings and have revised our manuscript according to the reviewer's insightful suggestions.

There are in this respect a several questions/issues which would need addressing:

Comment 1: The level of overlap with previous studies in a bit worrying as it shows lack of replication. How much do the previous scores correlate with the newly developed one?

Response: The Pearson's correlation between the *Metabolomic Aging Score* developed in our study and *MetaboHealth* (another metabolomic score predicting all-cause mortality) was 0.68, indicating a moderate to strong correlation. We have also provided an in-depth discussion of differences between our study and previous research. Please see our response to **Comment 4, Reviewer 1**.

Comment 2: Another issue is the treatment of the high correlation between the metabolites which make choice between one or the other subject to be influenced by the background noise. In this context it is possible that the chosen metabolites may be the most predictive but not necessarily the causal ones. This will make the resulting model formally correct but makes interpreting results difficult. Given the sometimes close to one correlation between multiple metabolites needs proper treatment. There are many choices: the first is to select the metabolite most representative of a group of tightly correlated traits as marker of the whole group; the second is to use a variable reduction technique to find a latent variable

that represents the whole group. I don't think that in this case the conclusions that the biomarkers could be causal are justified and the MR analysis should have not been limited just to the selected representative metabolites. This may also explain the discrepancy between the selected metabolites and those used previously.

Response: Thank you for your insightful comments and suggestions. Indeed, in our original analyses, we restricted the Mendelian Randomization analyses to the selected 54 aging-related biomarkers, potentially omitting causal metabolomic biomarkers. In our revised analyses exploring potential causal relationships between the metabolomic biomarkers and aging-related diseases, we first explored the genetic correlations between the 325 NMR biomarkers and subsequently included all of them as exposures in multivariable Mendelian Randomization. We calculated GWAS summary statistics for each of the 325 NMR biomarkers using WGS data of a subset of 95,372 individuals in our study. The quality control of the sequencing data and the methodological details are provided in ***Methods, page 27-28, line 594-621***. Please also refer to our response to **Comment 3** below.

> ***Results, page 7-8, line 155-178***: "Moving beyond cross-sectional associations, we investigated potential causal relationships linking metabolomic biomarkers with aging-related disease onset. It is possible that the NMR biomarkers that are causally related to aging-related diseases might not be among those selected by the LASSO model given the high collinearity as an inherent feature of metabolomics data[32]. Thus, we extended the search for causal biomarkers of aging-related diseases to all 325 NMR biomarkers.
>
> A genome-wide association study (GWAS) for each NMR biomarker was conducted using WGS data from a subset of 95,372 individuals. Variants with a minor allele frequency (MAF) > 0.1% were included after quality control (Methods). Based on GWAS summary statistics of 325 NMR biomarkers, we calculated pairwise genetic correlations via linkage disequilibrium score regression (LDSC)[33]. An extensive genetic correlation profile was identified underlying these metabolomic biomarkers, especially for the lipid and lipoprotein-related biomarkers (Extended Figure 5 and Supplementary Tables 8-9).
>
> Given the considerable pleiotropy of significant loci and correlations between the NMR biomarkers, we performed multivariable Mendelian randomization (MVMR) analysis to allow for multiple correlated exposures and pleiotropic instrumental variables[34]. All 325 NMR biomarkers were collectively included as exposures and twenty chronic non-communicable diseases (including sixteen leading causes of global disability-adjusted life-years (DALYs) amongst the elderly[35] and four additional common geriatric diseases) were included as the outcomes (Supplementary Table 10).
>
> Among a total of 6500 pairs between 325 NMR biomarkers and 20 aging-related diseases, 197 pairs, involving 136 NMR biomarkers and 20 diseases, were identified as candidate causal biomarker-disease pairs using both the MVMR-IVW and MVMR-Egger methods (*p*-value threshold with Bonferroni-correction of 5E-04) (Figure 2)."

Comment 3: It is unclear to me why the authors did not just rerun the GWAS analysis on the metabolites in UK biobank instead of relying only on existing summary statistics for less metabolites (42). The summary statistics are there why not use them?

Response: Thank you for your constructive suggestion. In response, we conducted a re-analysis of the genetic architecture for each NMR biomarker, resulting in the generation of GWAS summary statistics for all 325 biomarkers included in our study. We plan to provide these summary statistics available for public use, which we believe will serve as a valuable resource for future research endeavors. Considering the large number of variants (~17,000,000 sites) and the big file size (~1.7GB in zipped VCF format) for each GWAS summary statistics, we will submit them to the *GWAS Catalog* upon the publication of this manuscript. The methodological details of our GWAS are described in the revised manuscript.

> ***Methods, page 27-28, line 594-621***: "A subset of individuals with available whole-genome sequencing data in UK Biobank 200k release was included in the first place. Then, samples failed to pass quality requirements (UKB Field ID: 23093), samples with sex chromosome aneuploidy (UKB Field ID: 22019), samples with discordant genetic sex (UKB Field ID: 22001) and self-reported sex (UKB Field ID: 31) were excluded from further analysis. Finally, samples whose genetic ethnic group belonged to Caucasian (UKB Field ID: 22006) were included, resulting in a final sample size of 95,372 individuals for downstream GWAS analysis.
>
> We used the whole-genome sequencing data from the UKB 200k release in GraphTyper joint call pVCF format on UKB RAP[82]. Multiallelic variants were decomposed into biallelic variants using bcftools (v1.15.1). Quality control for SNPs and indels were performed based on following inclusion criteria[83]: (1) alternative alleles with AAscore>0.5; (2) variant sites with the tag "FILTER=PASS"; (3) Hardy-Weinberg P-value>10E-15; (4) genotype missing rate<10%. Further, only common variant sites (MAF>0.1%) were included in GWAS analysis. NMR data processed after quality control was inverse rank normalized with age, age^2, sex, age*sex, age^2*sex, BMI, medication status, smoking status, alcohol intake frequency, fasting time, assessment center, genetic PC 1-10 adjusted as covariates. GWAS analyses were conducted using STAAR framework (individual variant analysis provided within) which was suitable for biobank-scale WGS studies with abundant functional annotations to promote the power of association analysis[84,85]. Significance threshold was defined as 5E-09 for GWAS analysis including low-frequency variants (MAF>0.1%)[86] divided by 5 principal components which together accounted for >80% variation in 325 biomarker levels instead of dividing the total number of metabolomic biomarkers included in the study which was too conservative and stringent given the high collinearity among these biomarkers[61]. Thus, we chose 1E-09 as an appropriate *p-value* threshold to claim statistical significance.

GWAS summary statistics for each biomarker was further clumped to identify independent loci accounting for linkage disequilibrium between variants with a clumping window size of 500kb around the index variant (*p-value <1E-09*) and a linkage disequilibrium r2 threshold as 0.1."

Comment 4: It is unclear to me why the authors have decided that 0.05 was a reasonable p-value threshold for the MR analysis as they have performed many tests. They have applied multiple test correction in other parts of the paper why not here?

Response: Thank you for pointing out this inconsistency in our approach across different analyses. In our revised manuscript, we chose a Bonferroni-corrected p-value threshold as 5E-04 (i.e., 0.05/5/20) for the MVMR analysis (0.05 refers to the alpha level; 5 refers to the number of principal components which together accounted for more than 80% variation in the 325 metabolomic biomarkers; 20 refers to the number of aging-related diseases included in our study). The reporting of statistically significant causal relationships was revised accordingly throughout the manuscript.

Comment 5: The colocalization analysis is puzzling as they report PPH4 for the each result, however it is unclear to which SNP it refers to given each analysis has used multiple SNPs. Did they test for colocalization each pair of locus-trait? How was the multivariable framework handled in this case?

Response: Thank you for raising this point. In our revised manuscript, we tested for colocalization for each biomarker-disease causal relationship pair with statistically significant *p*-values for both the IVW and MR-Egger methods (***Methods, page 31, line 687-701***). Several colocalized causal variants were discussed in detail, and the full list of the colocalized causal variant (or sets) with their posterior probabilities and relevant annotations from VEP are provided in **Supplementary Table 12**.

Results, page 9-11, line 201-231: "Because results from Mendelian randomization analysis might be biased when the exposure and the outcome have distinct but correlated causal variants (for example in linkage disequilibrium)[36], we subsequently conducted colocalization analysis for each pair of 197 candidate causal relationships to explore whether the biomarker and the disease shared the same causal variant, which also provided further insights into the underlying mechanism. Of the 197 candidate causal pairs, 85 had a posterior probability of hypothesis four (PPH4, which implies that two traits share the same causal variant[37]) greater than 80%. For each colocalized NMR biomarker-disease pair, we annotated the colocalized causal SNP with its target gene and associated phenotypes with the Ensembl Variant Effect Predictor[38] (Supplementary Table 12). Generally, 85 colocalized pairs pointed to 22 causal variants in 19 target genes, indicating the existence of pleiotropic causal variants linking multiple NMR biomarkers with aging-related

diseases. For example, rs11591147, a missense variant in the protein-coding region of PCSK9 exhibited the most extensive pleiotropy as a shared causal variant for 19 NMR biomarker-disease pairs. Aligning with the well-known roles of PCSK9 in LDL-receptor recycling and coronary artery disease risk[39], rs11591147 was identified as a causal variant linking 19 lipid and lipoprotein-related biomarkers with hyperlipidemia, ischemic heart diseases, and arterial fibrillation and flutter. rs429358, a missense variant in the APOE locus, was a shared causal variant for five lipoprotein-related biomarkers with four diseases (Alzheimer's disease, senile cataract, hypertensive heart disease, chronic kidney disease). rs7412, another missense variant in the APOE locus linked seven lipoprotein-related biomarkers with ischemic heart disease and hyperlipidemia.

Investigation into relevant biological functions and associated phenotypes of colocalized variants also provided insight into potential pathways or mechanisms through which the metabolomic biomarker may impact disease onset. For example, rs1260326, a missense variant in GCKR, which encodes glucokinase regulators[40], was identified as a shared causal variant linking lactate and five lipoprotein-related biomarkers with type 2 diabetes. However, rs1260326 was also found to be associated with circulating leptin levels, C-reactive protein levels, fructose-bisphosphate aldolase B levels, gamma-glutamyl transpeptidase (GGT) levels, and serum IGF-1 and IGF binding protein (IGFBP)-3 levels[41]. These enzymes and biological molecules associated with rs1260326 might provide a hint towards the mechanism through which lactate and lipoprotein-related biomarkers are involved in disease onset."

Comment 6: MR-Egger tests only a specific type of pleiotropy (directional) and can be used but just with the aim of testing this type of assumption violations. It would have been better to use other types of approaches (i.e. MR-median).

Response: Thank you for suggesting additional approaches. The median-based approaches in MR allows the IV assumption to be violated in a more general way with a maximum tolerance of 50% invalid IVs. We carefully checked for pleiotropy of instrumental variables (IVs) included in the MVMR and discovered that a large proportion of IVs were pleiotropic and significantly associated with multiple NMR biomarkers in our study (p-value < 1E-09 in raw GWAS summary statistics). This phenomenon was especially pronounced among the lipid-related biomarkers. Under this scenario, each exposure was assigned a large proportion of "invalid" IVs extending the tolerance of median-based methods as a threshold of 50%. The MR-Egger method in a multivariable context allows all the IVs to be pleiotropic (a tolerance of invalid IVs of 100%), which is the optimal approach for dealing with highly correlated exposures and a large number of pleiotropic IVs (PMID: 28960498). Causal pairs identified here using MVMR-IVW and MVMR-Egger were proposed as candidates, and those with supporting evidence from colocalization analyses are likely more robust. We reminded our readers to be cautious when interpreting these candidate causal relationships.

Discussion, page 24, line 526-531: "Third, it is crucial to interpret candidate causal relationships identified in MVMR analysis with caution, as only rigorous randomized controlled trials are the gold standard for testing cause-effect relationships[75]. Moreover, the MVMR analysis detected the direct effect of an exposure on the outcome independent of other correlated exposures, rather than the overall effect on that outcome because the exposure might influence the outcome indirectly via other related exposures in a multivariable scenario[76]."

Comment 7: In conclusion I think that overall the paper is relatively sound in the score building and testing part but much weaker on the interpretation part and needs to be revied especially in its interpretation.

Response: Thank you for your constructive feedback on our manuscript. We appreciate your positive comments regarding the score building and testing sections. Regarding the concerns you have raised about the interpretation, we have thoroughly reviewed your comments along with those of the other reviewers. We have revised the manuscript to improve its clarity and depth, and we believe these enhancements have greatly strengthened our paper.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I thank the authors for the changes made. The manuscript has been improved. I enclose few comments below.

1. Figure legends are missing both in manuscript and the figure files. It is difficult to follow the results.
2. That statement mentioned that “Still, the residuals of the Metabolomic Aging Score outperformed others and had similar prediction accuracy as chronological age across one-year to five-year intervals”. The conclusion “Hence, we proposed the potential application of the Metabolomic Aging Score in several clinical scenarios, including monitoring short-term mortality risk” may be inappropriate because Metabolomic Aging Score has no significant advantage compared to chronological age.
3. I agree that Metabolomic Aging Rate (defined as ΔMet/ Follow-up intervals) may provide more refined metabolomic information on the rate of change towards the aging pathology. Therefore, the application of aging assessment could be focus more on ΔMet.

Reviewer #2 (Remarks to the Author):

The paper can be accepted.

Reviewer #3 (Remarks to the Author):

I think the manuscript is greatly improved and much clearer.

I still don't agree with using MR-egger as sensitivity analysis as it detects only very specific problems. It indeed allows for 100% of potentially invalid instruments, but this applies only if they are all affected by the same bias in the same direction otherwise it has the same exact limitations of other methods. Adding further sensitivity analyses should not burden the authors and would strengthen the results.

Finally in the discussion about the MR results I would add a sentence about the fact that with metabolites even with MVMR collinearity will affect the results (MVMR is in essence a mutivariable linear regression) and thus which specific metabolite affects which trait should be considered cautiously.

Otheriwse I have no other comments.

Response to reviewers' comments

Reviewer #1

I thank the authors for the changes made. The manuscript has been improved. I enclose few comments below.

Response: Thank you for your positive feedback and your previous comments.

Comment 1: Figure legends are missing both in manuscript and the figure files. It is difficult to follow the results.

Response: Thank you for pointing out this oversight. We have added the legends to each figure file and attached a separate figure legend file. We apologize for any inconvenience this may have caused.

Comment 2: That statement mentioned that "Still, the residuals of the Metabolomic Aging Score outperformed others and had similar prediction accuracy as chronological age across one-year to five-year intervals". The conclusion "Hence, we proposed the potential application of the Metabolomic Aging Score in several clinical scenarios, including monitoring short-term mortality risk" may be inappropriate because Metabolomic Aging Score has no significant advantage compared to chronological age.

Response: Thank you for your insightful comment. We acknowledge that although the residual of the *Metabolomic Aging Score* demonstrated similar prediction accuracy as chronological age, it did not surpass it. We agree that this distinction should be more clearly communicated. Our intention was to highlight the potential of the *Metabolomic Aging Score* as a complementary tool rather than a superior alternative.

> "We propose the application of the *Metabolomic Aging Score* as a complementary tool in various clinical scenarios, where it can be used alongside chronological age and other clinical parameters to provide a more comprehensive assessment. These include monitoring short-term mortality risk, identifying aging-accelerated populations, and enhancing disease risk prediction when used in combination with traditional risk factors." **(Discussion, pages 22-23, lines 497-501)**

Comment 3: I agree that Metabolomic Aging Rate (defined as ΔMet/ Follow-up intervals) may provide more refined metabolomic information on the rate of change towards the aging pathology. Therefore, the application of aging assessment could be focus more on ΔMet.

Response: We appreciate your perspective on focusing the aging assessment more on the *Metabolomic Aging Rate* and the change in *Metabolomic Aging Score*. We agree that the *Metabolomic Aging Rate* provides a more refined resolution of personal aging trajectories and warrants further exploration. Regrettably, the longitudinal analysis in our study was constrained by a limited sample size with revisit data (~5% of the entire cohort) and a limited re-assessment frequency (only once). Consequently, we have highlighted the need for future studies, including our own forthcoming research, to focus on refining and improving this rate. We believe that as larger and more longitudinal cohorts are established and designed, biological aging rate metrics will be improved and applied to facilitate more effective and more personalized aging assessments.

> "Aging is a gradual and dynamic biological process. Longitudinal studies are crucial for capturing indicative changes in bio-signals over time and providing insights into the evolving nature of aging pathology[80,81]. The *Metabolomic Aging Rate* reflects the rate of change towards aging pathology. With more samples from follow-up visits, increased reassessment frequency, and the development of large-scale longitudinal cohorts[82], this rate can be further refined and improved. Such advancements would contribute to more precise and personalized aging assessments. Future research should focus more on the analysis of biological aging rates, as exemplified in our study, which offers a higher resolution of the pace of aging." **(Discussion, pages 23-24, lines 518-525)**

Reviewer #2

The paper can be accepted.

Response: Thank you. We are grateful for your previous comments which have contributed to the improvement of our work.

Reviewer #3

I think the manuscript is greatly improved and much clearer.

Response: Thank you for your positive feedback and your earlier comments.

Comment 1: I still don't agree with using MR-egger as sensitivity analysis as it detects only very specific problems. It indeed allows for 100% of potentially invalid instruments, but this applies only if they are all affected by the same bias in the same direction otherwise it has the same exact limitations of other methods. Adding further sensitivity analyses should not burden the authors and would strengthen the results.

Response: Thank you for your valuable feedback. To address these concerns, we have included additional sensitivity analyses including MVMR-Median and MVMR-Lasso and clarified the advantage of each method under different scenarios.

> "Four MVMR analysis methods were utilized: MVMR-IVW, MVMR-Egger, MVMR-Median, and MVMR-Lasso. Each method provides valid estimates of causal effects under varying sets of relaxed assumptions[105]: MVMR-IVW provides unbiased estimates when all genetic variants are valid IVs or, in the presence of invalid IVs, if the pleiotropy is balanced and the InSIDE (Instrument Strength Independent of Direct Effect) assumption is met; MVMR-Egger is robust to directional pleiotropy and provides unbiased estimates even when all IVs are invalid, provided that the InSIDE assumption is met; MVMR-Median provides unbiased estimates when at least 50% of the weights come from valid IVs, allowing for the IV assumptions to be violated in a more general manner than MVMR-Egger; MVMR-Lasso identifies valid IVs and accounts for pleiotropy caused by invalid IVs without loss of power and without the requirement for the InSIDE assumption. Each method offers unique advantages tailored to specific scenarios. When combined, they provide complementary and supportive evidence for a candidate causal estimate, thereby strengthening the overall findings." **(Methods, page 32, lines 706-718)**

We observed that these four methods provided overall consistent results. The MVMR-IVW and MVMR-Egger methods yielded wider confidence intervals than the MVMR-Lasso and MVMR-Median methods (**Extended Figure 6a-d**). Consequently, we primarily based our interpretations of the MVMR analysis on the estimates from the MVMR-IVW method, as its wider confidence interval suggest a more conservative estimation.

Comment 2: Finally in the discussion about the MR results I would add a sentence about the fact that with metabolites even with MVMR collinearity will affect the results (MVMR is in essence a

mutivariable linear regression) and thus which specific metabolite affects which trait should be considered cautiously.

Response: Thank you again for your insightful comment. In our revised analysis, we have tried to solve this problem by extracting the full rank matrix (with a tolerance for determining rank set at 1E-07) from the original "Exposure-Instrumental variable" marginal association matrix. Redundant vectors were eliminated which yielded 287 NMR biomarkers to include in the following analysis. This step would alleviate the issue of multicollinearity and lead to more robust estimations (candidate causal pairs dropped from 197 to 14 under the Bonferroni-corrected *p*-value threshold compared with the previous analysis).

> "Different from univariable Mendelian randomization, MVMR requires the marginal association matrix between the IVs and exposures to be of full column rank. Multicollinearity within the exposure-IV association matrix can lead to unstable estimates and inflated type-I errors, thereby reducing statistical power[104]. To meet this requirement, we extracted a full rank matrix (with a tolerance for determining rank set of 1E-07) and eliminated redundant vectors from the exposure-IV marginal association matrix (β_{xi} matrix). We retained a total of 3,167 candidate IVs and 287 exposures in the full-rank association matrix for further analysis. After harmonizing allele effects between the exposures and outcomes, a total of 2,164 IVs were included in the MVMR analysis **(Supplementary Table 11)**." **(Methods, pages 31-32, lines 697-705)**

Despite this, the issue of multicollinearity could not be eliminated, and as you suggested, we have drawn the readers' attention to this matter.

> "Importantly, similar to multivariable regression, multicollinearity due to the inclusion of correlated exposures could lead to unstable estimates and reduced statistical power in MVMR. Therefore, careful consideration is needed when determining which specific biomarkers influence which diseases[87]." **(Discussion, pages 24, lines 541-544)**

Otheriwse I have no other comments.

Response: Thank you once again for your thorough review.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I have no further comments.

Reviewer #3 (Remarks to the Author):

Nothing to add