

# Heterogeneous metabolomic aging across the same age and prediction of health outcome

## Authors:

Xueqing Jia<sup>1</sup>, Jiayao Fan<sup>1</sup>, Xucheng Wu<sup>1</sup>, Xingqi Cao<sup>1</sup>, Lina Ma<sup>2</sup>, Zeinab Abdelrahman<sup>3</sup>, Daniele Bizzarri<sup>4</sup>, Erik B van den Akker<sup>4</sup>, P. Eline Slagboom<sup>4</sup>, Joris Deelen<sup>4</sup>, Dan Zhou<sup>1</sup>, Zuyun Liu<sup>1\*</sup>

## Authors' affiliations:

<sup>1</sup>Center for Clinical Big Data and Analytics Second Affiliated Hospital, and Department of Big Data in Health Science School of Public Health, The Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province, Zhejiang University School of Medicine, Hangzhou 310058, Zhejiang, China

<sup>2</sup>Department of Geriatrics, National Research Center for Geriatric Medicine, Xuanwu Hospital, Capital Medical University, Beijing, China

<sup>3</sup>Centre for Public Health, Queen's University of Belfast, Belfast BT12 6BA, UK

<sup>4</sup>Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands

\*Corresponding author

Zuyun Liu, Professor, Center for Clinical Big Data and Analytics Second Affiliated Hospital, and Department of Big Data in Health Science School of Public Health, The Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province, Zhejiang University School of Medicine, Hangzhou 310058, Zhejiang, China. Telephone: +86-0571-87077127. Email: Zuyun.liu@outlook.com or zuyunliu@zju.edu.cn

## **Abstract**

Existing metabolomic clocks exhibit deficiencies in capturing the heterogeneous aging rates among individuals with the same chronological age. Yet, the modifiable and non-modifiable factors in metabolomic aging have not been systematically studied. Here, we leveraged metabolomic profiles of 239,291 UK Biobank participants for 10-year all-cause mortality prediction to generate and validate a new aging measure--MetaboAgeMort. The MetaboAgeMort showed significant associations with all-cause mortality, cause-specific mortality, and diverse incident diseases. Adding MetaboAgeMort to conventional risk factors model improved the predictive ability of 10-year mortality. We identified 99 modifiable factors for MetaboAgeMort, where 16 factors representing pulmonary function, body composition, socioeconomic status, dietary quality, smoking status, alcohol intake, and disease status showed quantitatively stronger associations. The genetic analyses revealed 99 genomic risk loci and 271 genes associated with MetaboAgeMort. Our study illuminates heterogeneous metabolomic aging across the same age, which provides avenues for developing anti-aging therapies and personalized interventions.

**Keywords:** aging, biological age, metabolomic, mortality, modifiable factor, genetic determinants.

## **Introduction**

Aging is the greatest common risk factor for most chronic diseases<sup>1</sup>. The rate at which individuals age is heterogeneous, engendering variations in the susceptibility and progression of diseases and mortality<sup>2</sup>. As the global demographic shifts towards an aging population, the ability to measure aging, identify individuals who age faster, and understand the factors that contribute to differential rates of aging are of utmost importance. These findings have significant implications for the development of targeted preventive programs and interventions. These efforts may alleviate the socioeconomic and healthcare burden of age-related diseases, thus promoting healthy aging and longevity.

Metabolomics offers a novel avenue for assessing the biological processes that underlie aging<sup>3</sup>. To date, multiple studies have attempted to elucidate how metabolomic profiles in various tissues (e.g., blood, urine, and cerebrospinal fluid) interact with aging, and a few metabolomic clocks have been proposed to measure biological aging<sup>4-7</sup>. The majority of these clocks are generated based on correlations between metabolomic profiles and chronological age; while chronological age is considered an imperfect surrogate for building aging measures as it does not fully capture the heterogeneity of individual aging rates<sup>8</sup>. In contrast, metabolomic aging measures based on health-related surrogate indicators (e.g., time to death) may better reflect an individual's health status and reveal intrinsic biological aging mechanisms<sup>9</sup>. Previous studies have utilized targeted or untargeted mass spectrometry and nuclear magnetic resonance (NMR) techniques to construct multivariable metabolite scores of all-cause mortality<sup>10-12</sup>. While these metabolite scores exhibit exceptional predictive accuracy even over conventional risk factor models, their applicability in risk stratification, especially across individuals of the same chronological age, remains constrained by their reliance on scaled biomarker values created independently for each cohort<sup>12</sup>. Metabolomic profiles manifest pronounced responsiveness to the confluence of endogenous genetic regulation and exogenous environmental exposures in each

individual cohort<sup>13</sup>. Thus, developing aging measures that can be calculated based on concentration units derived from individual-level data, may have more applicability in both clinical setting and research on the biology of aging.

Although certain behaviors (e.g., dietary quality) have displayed anti-aging properties in human and animal models<sup>14,15</sup>, their impact on metabolomic aging remains uncertain. Furthermore, the complex relationship between aging and a constellation of diverse factors (e.g., local environmental factors and socioeconomic status [SES]) has emerged as a burgeoning focus in the field of aging research. A thorough investigation of these modifiable factors in metabolomic aging could reveal new strategies for preventive interventions targeting the aging process. Moreover, prior investigations have indicated that aging measures may capture distinct aging domains influenced by varying genetic determinants<sup>16</sup>. However, the underlying mechanisms and pathways of metabolomic aging have not been elucidated, thereby restricting the identification of potential therapeutic targets<sup>17</sup>.

In this study, we leveraged large-scale metabolomic data from the UK Biobank (UKB), a prospective cohort study of over 500,000 participants. We first developed a novel aging measure, MetaboAgeMort, using all-cause mortality as a surrogate<sup>18</sup>. Next, we evaluated its applicability by examining its association with aging-related outcomes (i.e., morbidity and mortality), and comparing its performance to conventional risk factors and several previously trained metabolomic clocks. Finally, we systematically identified the modifiable factors and genetic determinants for MetaboAgeMort (Fig. 1).

## **Results**

### **Population characteristics**

As shown in Supplementary Fig. 1, 239,291 participants with complete data on plasma metabolomics and covariates at baseline were included. These participants had a median age of

58.3 years (interquartile range [IQR]: 50.6, 63.7), and the majority were female (53.0%), and white ethnicity (95.6%). To develop the MetaboAgeMort model, the 239,291 participants were randomly split into a training (n = 167,506) and a testing set (n = 71,785), with a 7 to 3 ratio. No significant differences were observed in the sociodemographic characteristics of participants between the training set and the testing set. Detailed characteristics of the total participants and by datasets are presented in Supplementary Table 1.

### **Development of MetaboAgeMort and MetaboAgeMort Acceleration (MetaboAgeMortAccel)**

During a median follow-up of 13.9 years, we documented 20,447 deaths among 239,291 participants. After adjustment for potential confounders and accounting for multiple testing, a total of 185 metabolic biomarkers, encompassing amino acids, glycolysis-related metabolites, ketone bodies, fatty acids, lipids, and lipoprotein subclasses, demonstrated significant correlations with all-cause mortality ( $P < 0.05/249$ ) (Supplementary Table 2).

To further select variables for inclusion in the MetaboAgeMort model, we applied a Cox regression model with least absolute shrinkage and selection operator (LASSO) penalization--where the hazard of all-cause mortality was regressed on the 185 metabolic biomarkers and chronological age--in the training set. Finally, chronological age and 35 metabolic biomarkers, including average diameter for very low-density lipoprotein (VLDL) particles, linoleic acid, ratio of omega-3 fatty acids to total fatty acids, ratio of monounsaturated fatty acids to total fatty acids, ratio of linoleic acid to total fatty acids, alanine, histidine, leucine, valine, phenylalanine, tyrosine, glucose, pyruvate, citrate, 3-hydroxybutyrate, acetate, acetoacetate, acetone, creatinine, albumin, glycoprotein acetyls, triglycerides in very large VLDL, free cholesterol in very large high-density lipoprotein (HDL), total lipids in small HDL, cholesteryl esters in small HDL, triglycerides to total lipids ratio in large VLDL, phospholipids to total lipids ratio in very small VLDL, phospholipids to total lipids ratio in intermediate density lipoprotein (IDL), cholesteryl esters to total lipids ratio in IDL, triglycerides to total lipids ratio in large low-density lipoprotein

(LDL), triglycerides to total lipids ratio in medium LDL, phospholipids to total lipids ratio in small LDL, free cholesterol to total lipids ratio in small LDL, cholesteryl esters to total lipids ratio in very large HDL, free Cholesterol to total lipids ratio in small HDL, were selected (Supplementary Table 3).

Next, MetaboAgeMort was developed using the methods previously proposed by Levine et al. in the training set<sup>18</sup>. For more information about the MetaboAgeMort estimator, refer to Supplementary Table 4. A profiling of the MetaboAgeMort performance was carried out in the testing set (n = 71,785). MetaboAgeMort ranged from 27.82 to 104.06 years, with a mean and median value of 55.81 (standard deviation [SD] = 9.18) and 56.10 (IQR: 48.99, 62.50) years. As shown in Fig. 2a, MetaboAgeMort was highly correlated with chronological age across all participants (r = 0.85) and within each sex subgroup (female: r = 0.85; male: r = 0.86).

We also calculated a metric, MetaboAgeMortAccel, following the methods by Liu et al.<sup>19</sup> MetaboAgeMortAccel represents the divergence of MetaboAgeMort from chronological age (i.e., whether a person appears younger [values < 0] or older [values > 0] than expected, based on his/her chronological age). The MetaboAgeMortAccel displayed a range of -16.43 to 41.29 years, with a mean and median value of 0 (SD = 4.78) and -0.41 (IQR: -3.29, 2.79) years in the testing set.

### **Association of MetaboAgeMort with mortality**

The associations of MetaboAgeMort with all-cause and cause-specific mortality in the testing set are demonstrated in Fig 2b. After adjustment for potential confounders, each additional year in MetaboAgeMort corresponded to an 7% rise in the risk of all-cause mortality (Hazard Ratio [HR]: 1.07, 95% confidence interval [CI]: 1.06, 1.08) (Supplementary Table 5). Our finding remains consistent when: (1) stratified by chronological age, sex, ethnicity, education level, smoking status, alcohol intake frequency, regular exercise, healthy diet, and body mass index (BMI) category (Supplementary Table 6), (2) excluding participants who died within five years of

follow-up (Fig. 2b and Supplementary Table 5), and (3) restricting the sample to diseases-free participants (Supplementary Table 7). In addition, MetaboAgeMort was significantly positively associated with cause-specific mortality, including cancer (HR: 1.06, 95% CI: 1.05, 1.07), cardiovascular disease (CVD, HR: 1.08, 95% CI: 1.07, 1.10), respiratory disease (HR: 1.08, 95% CI: 1.06, 1.10), digestive disease (HR: 1.12, 95% CI: 1.09, 1.14), neurodegenerative disease (HR: 1.03, 95% CI: 1.01, 1.04), and other causes (HR: 1.09, 95% CI: 1.07, 1.10) (Fig. 2b and Supplementary Table 5).

The Kaplan-Meier survival curves indicated that individuals in the highest quartile group (Q4) of MetaboAgeMortAccel had significantly increased risks of all-cause and cause-specific mortality when compared to those in the lowest quartile (Q1) (Fig. 3a and Supplementary Table 8). The all-cause mortality rates of the highest quartile group (Q4) were found to be comparable, or in certain instances higher, than those of the lowest quartile group (Q1), despite the latter being 10 years older chronologically (Supplementary Fig. 4).

#### **Associations of MetaboAgeMort with diseases incidence**

The associations of MetaboAgeMort with the risk of multiple diseases in the testing set are depicted in Fig. 2c. After adjustment for potential confounders, each 1-year increment in MetaboAgeMort was significantly associated with higher risks of cancer (HR: 1.02, 95% CI: 1.02, 1.03), CVD (HR: 1.03, 95% CI: 1.03, 1.04), dementia (HR: 1.04, 95% CI: 1.03, 1.05), liver disease (HR: 1.06, 95% CI: 1.05, 1.07), respiratory disease (HR: 1.07, 95% CI: 1.06, 1.07), chronic kidney disease (HR: 1.09, 95% CI: 1.09, 1.10), hypertension (HR: 1.05, 95% CI: 1.04, 1.06), type 2 diabetes mellitus (HR: 1.07, 95% CI: 1.06, 1.08), eyes disease (HR: 1.02, 95% CI: 1.01, 1.02), depression (HR: 1.03, 95% CI: 1.02, 1.03), and anxiety (HR: 1.02, 95% CI: 1.01, 1.02), with an exception of osteoarthritis (HR: 1.00, 95% CI: 1.00, 1.01) (Supplementary Table 9).

The Kaplan-Meier survival curves exhibited discernible trajectories among the quartile groups of MetaboAgeMortAccel (Fig. 3b and Supplementary Fig. 3b). The group in the highest

quartile (Q4) showed a significant association with increased risks of multiple disease incidences compared to the lowest quartile (Q1) (Supplementary Table 10 and 11).

### **Discriminative improvements beyond clinical predictors**

Based on the findings depicted in Fig. 2d, it is evident that the area under the curve (AUC) of MetaboAgeMort displayed a considerable improvement relative to chronological age, aligning it more closely with the AUC of the conventional risk factors model. MetaboAgeMort added predictive utility of 10-year mortality beyond conventional risk factors (i.e., chronological age, sex, alcohol intake frequency, smoking status, BMI, systolic blood pressure, triglycerides, creatinine, total cholesterol, HDL cholesterol, and prevalent diabetes, CVD and cancer)<sup>11</sup>. Compared with the conventional risk factors model, the combined model including MetaboAgeMort had better discrimination ability, as demonstrated by significantly increased C-statistics (0.017,  $P < 0.001$ ) (Supplementary Table 12). The superior performance of MetaboAgeMort was further confirmed through substantial enhancements in reclassification, as evaluated by integrated discrimination improvement (IDI: 0.018, 95% CI: 0.004, 0.021) (Supplementary Table 12), suggesting that MetaboAgeMort captures something above and beyond what can be explained for mortality risk by conventional risk factors.

### **The comparison of MetaboAgeMort with MetaboAge and MetaboHealth score**

In addition, we calculated two pre-existing well-known multi-metabolite scores (i.e., MetaboAge<sup>5</sup> and MetaboHealth score<sup>12</sup>), and assessed their correlations with the mortality risk and multiple diseases incidence. The MetaboAge score predicts chronological age (in years) directly. Two versions of MetaboAge were calculated: MetaboAge\_LM (generated through linear regression), and MetaboAge\_EN (generated through ElasticNET regression)<sup>20</sup>. The MetaboHealth score is a multivariate model predicting all-cause mortality. It was calculated as the weighted sum of 14 log-transformed and cohort-scaled metabolites. The distributions of these multi-metabolite scores in the testing set are shown in Supplementary Fig. 5a. As shown in Supplementary Table 13, after



adjustment for potential confounders, the highest quartile group (Q4) of the MetaboHealth score has significantly increased risks of all-cause mortality, cause-specific mortality, and multiple incident diseases, compared to the lowest quartile group (Q1). The MetaboAge\_EN showed significant association with all-cause mortality, CVD mortality, and several cardiometabolic diseases (i.e., CVD and T2DM). Conversely, MetaboAge\_LM did not demonstrate any appreciable relationship with these same clinical endpoints. As shown in Supplementary Table 14, even after mutual adjustment, both MetaboAgeMort and MetaboHealth score maintained significant correlations with all-cause mortality. Notably, the Akaike Information Criterion (AIC) for the model integrating MetaboAgeMort indicated a relatively superior fit, implying that MetaboAgeMort may have a higher predictive capacity or enhanced explanatory power in forecasting all-cause mortality when compared to the model integrating the MetaboHealth score.

Moreover, we compared MetaboAgeMort with these multi-metabolite scores, in terms of their predictive utility for 10-year mortality risk and multiple diseases incidence. As shown in Supplementary Fig.5b, MetaboAgeMort displayed a significantly higher AUC than that of the MetaboAge and MetaboHealth score for 10-year all-cause mortality prediction ( $P < 0.001$ ). Combining MetaboAgeMort with conventional risk factors resulted in an AUC that outperformed models integrating MetaboAge (MetaboAge\_EN and MetaboAge\_LM, both  $P < 0.001$ ) or MetaboHealth score ( $P = 0.021$ ) with the same conventional risk factors. When combining MetaboAgeMort with MetaboHealth score and conventional risk factors, the resulting AUC was higher than that of the model integrating only MetaboHealth score with conventional risk factors, but this difference was not significant when compared to the model integrating only MetaboAgeMort with conventional risk factors. This finding suggests that MetaboAgeMort may capture something beyond what can be explained for mortality risk by MetaboHealth score, when taking into account conventional risk factors. MetaboAgeMort also exhibited a higher predictive utility for multiple diseases (except for cancer incidence) compared to the MetaboHealth score

(Supplementary Fig. 5c). After accounting for chronological age and sex, the AUC of the model integrating both MetaboAgeMort and MetaboHealth score was higher than that of the model with MetaboHealth score alone, but did not show a significant improvement over the model with MetaboAgeMort alone. This finding suggests that MetaboAgeMort may also capture something beyond what can be explained for diseases incidence by MetaboHealth score.

### **Modifiable factors for MetaboAgeMort**

Then we investigated the modifiable factors of MetaboAgeMort. We considered a total of 107 potentially modifiable factors from the UKB baseline survey. After Bonferroni correction, 99 factors showed a significant association with MetaboAgeMort ( $P < 4.67 \times 10^{-4}$ , Fig. 4a and Supplementary Table 13). For 16 factors across four categories, the associations were relatively more substantial ( $\geq 2$  years of change in MetaboAgeMort) per 1-SD change in the factor (Supplementary Fig. 6), such as forced vital capacity (FVC) (tertile 3 vs. tertile 1,  $\beta = -2.22$ , 95% CI: -2.28, -2.15), body fat percentage (tertile 3 vs. tertile 1,  $\beta = 3.55$ , 95% CI: 3.49, 3.61), dietary index (tertile 3 vs. tertile 1,  $\beta = -2.04$ , 95% CI: -2.09, -1.98), and average total household income before tax (greater than 100,000 vs. less than 18,000,  $\beta = -3.34$ , 95% CI: -3.44, -3.24). Comparable correlations were noted when stratifying by chronological age and sex (Supplementary Table 13).

### **Genetic determinants for MetaboAgeMort**

To better understand the genetic mechanisms underlying metabolomic aging, we performed a genome-wide association study (GWAS) analysis of MetaboAgeMort. In the GWAS analysis, 11,688 SNPs significantly associated with MetaboAgeMort were identified ( $P < 5 \times 10^{-8}$ ) (Supplementary Fig. 7). The SNP-derived heritability of MetaboAgeMort was 38.26% ( $P = 2.23 \times 10^{-86}$ ). Using the Functional Mapping and Annotation (FUMA) online platform (v1.5.2), we pinpointed 1,068 independent significant SNPs, 319 lead SNPs, and 99 genomic risk loci (Supplementary Table 14-17). Furthermore, 585 prioritized genes that may be involved in the

genetic etiology of MetaboAgeMort were identified by positional mapping (Supplementary Table 18). The leading SNP of the most significant locus (rs174575, locus 57) was positioned near or within *FADS1* and *FADS2* on chromosome 11. The leading SNP of the second most significant locus (rs217184, locus 78) was in *TXNLAB*, *HPR*, and *HP* on chromosome 16 (Supplementary Table 19). Summary results per genomic risk locus are shown in Fig. 4b.

We also performed analyses pertaining to genes and tissue enrichment using Multi-Marker Analysis of GenoMic Annotation (MAGMA) v1.08 within FUMA. In genome-wide gene-based association analysis (GWGAS), 310 genes were determined to be genome-wide significant after applying Bonferroni correction (Supplementary Table 20), where 271 genes were also positionally mapped to significant loci from the SNP-based analysis above. The gene-set analysis identified 13 significant gene sets after Bonferroni correction, including lipid-related biological process (e.g., reverse cholesterol transport, cholesterol metabolism, and phospholipid homeostasis), CYP2E1 reactions, and liver specific genes (Supplementary Table 21). The tissue-enrichment analysis indicated that liver displayed significant specificity in gene expression for the MetaboAgeMort-associated genes (Fig. 4c and Supplementary Table 22-23).

## **Discussion**

Leveraging the large-scale metabolomics data, we have formulated a groundbreaking aging measure, MetaboAgeMort, based on 10-year all-cause mortality risk prediction. The MetaboAgeMort demonstrated remarkable predictive utility of mortality risk across a wide range of demographic and socioeconomic stratifications, as well as health behavior factors and causes of death. Significantly, the conventional risk factors were augmented by MetaboAgeMort in terms of predictive accuracy for 10-year mortality. Meanwhile, our study has revealed compelling associations between accelerated metabolomic aging within the same age group and an increased likelihood of various health-related outcomes. This implies that it could serve as a comprehensive

indicator for health and mortality risk stratification in a clinical setting. Next, we identified 99 modifiable factors across seven categories for MetaboAgeMort, highlighting the crucial role played by body composition, healthy diet, SES, and pulmonary function in the process of metabolomic aging. The genetic analyses ultimately revealed 99 significant genomic risk loci and 271 genes linked to MetaboAgeMort, thus offering new insights into the genetic architecture of metabolomic aging.

Previous studies have generated metabolomic clocks while using chronological age as a surrogate<sup>21</sup>. In comparison, our MetaboAgeMort further incorporates information on all-cause mortality risk, which is considered a more reliable surrogate for biological aging than chronological age, through a sophisticated modeling method. As one would expect from a measure of aging, MetaboAgeMort not only assesses the risks of all-cause and cause-specific mortality, but also the risks of various incident diseases, highlighting its considerable potential in the early detection of individuals at risk and facilitating timely and effective interventions. It is worth mentioning that we have substantiated the contribution of MetaboAgeMort in enhancing the predictive capacity for 10-year all-cause mortality risk in addition to conventional risk factors. These findings translate into potential clinical application of MetaboAgeMort as an additional source of discriminatory information to refine comprehensive risk assessments for death and diseases. To date, numerous studies have endeavored to identify metabolite predictors of mortality risk and have successfully developed multivariate metabolite scores (e.g., MetaboHealth score) that exhibited significant association with all-cause mortality<sup>22</sup>. Nonetheless, the utilization of these metabolite scores is constrained due to the reliance on cohort-specific scaled biomarker values, thereby impeding its clinical application in identifying individuals with a propensity for accelerated aging and comparison across different populations. MetaboAgeMort, directly calculated from individual-level data, more intuitively shows the heterogeneity in mortality risk among individuals of the same chronological age, exhibiting enhanced

generalization.

The 35 metabolic biomarkers employed in our study to develop MetaboAgeMort are implicated in diverse processes, including lipoprotein and fatty acid metabolism, fluid balance, and inflammation, indicating that aging is an intricate multidimensional phenomenon. Previous studies have sought to elucidate the interaction between metabolites and aging. In line with the discoveries made by Deelen et al.<sup>12</sup>, our study revealed that average diameter for VLDL particles, histidine, leucine, phenylalanine, valine, glucose, acetoacetate, albumin, glycoprotein acetyls, the ratio of polyunsaturated fatty acids to total fatty acids (i.e., omega-3 fatty acids to total fatty acids percentage, and linoleic acid to total fatty acids percentage), and total lipids in small HDL were essential independent indicators for mortality. In addition, we observed associations for several other metabolic biomarkers, such as ketone bodies and relative lipoprotein lipid concentrations. KBs are endogenous fuels generated by the liver in response to metabolic stress<sup>23</sup>. In a healthy community-based population, higher elevated endogenous KBs have shown to be positively associated with all-cause mortality<sup>24</sup>. Relative lipoprotein lipid concentrations play a role in lipid homeostasis and their associations with mortality may be partially attributed to their regulatory effect on plasma triglyceride levels, a critical mortality risk factor<sup>25</sup>. Collectively, our study contributes to the comprehension of metabolic alterations that underlie the process of aging.

The primary focus of aging research has been on the development of strategies to combat aging. Measures like MetaboAgeMort, which capture future morbidity and mortality risk, could facilitate evaluation of intervention efficacy while eliminating the requirement for extended follow-up periods. A recent study in the UK Airwave cohort has discovered a correlation between metabolomic aging and several factors such as overweight, obesity, heavy drinking, diabetes, depressive symptoms, depression, anxiety, and post-traumatic stress disorder<sup>7</sup>. Using a larger-scale population-based cohort, our study meticulously investigated the modifiable factors associated with metabolomic aging. Consequently, we identified a total of 99 potential factors

across seven distinct categories: SES, early life and sexual health, medical history, physical measures, psychosocial factors, local environment, and lifestyle. Stronger associations ( $\beta > 2$ ) were quantitatively observed for 16 factors related to pulmonary function, body composition, SES, dietary quality, smoking status, alcohol intake, and disease status, all of which have previously been reported to be associated with aging<sup>26-29</sup>. Our study provides a metabolomic insight into the mechanisms linking these factors to aging process. Certain local environmental factors, such as fine particulate matter (PM<sub>2.5</sub>) and nitrogen dioxide (NO<sub>2</sub>), have been evidenced to potentiate oxidative stress responses within biological systems, impair mitochondrial function, and subsequently lead to metabolic irregularities<sup>30,31</sup>. The correlation between these factors and MetaboAgeMort underscores the significance of enacting urban planning and environmental preservation policies to slow down the aging process. Furthermore, MetaboAgeMort demonstrated responsiveness to specific early-life exposure factors, suggesting that targeted interventions during the critical developmental period, such as optimizing maternal nutrition and improving the developmental environment for children, could potentially contribute to the promotion of healthy aging and reduction of risks associated with late-life mortality and morbidity<sup>32</sup>.

Genetics play a substantial role in determining individual biological aging rates<sup>33</sup>. Based on our current understanding, this study offers the initial evidence regarding the genetic determinants of metabolomic aging. By utilizing genotyping data, we pinpointed 99 genomic risk loci and 271 genes associated with MetaboAgeMort. The most significant SNPs were identified within the *FADS* cluster (*FADS1*, *FADS2*) on chromosome 11. These genes have emerged as significant genes in prior studies on serum omega-3 fatty acid<sup>34</sup>, a crucial type of polyunsaturated fatty acids that have demonstrated favorable impacts on age-related diseases (e.g., CVD and metabolic diseases)<sup>35</sup>. We also identified several genome-wide significant genes on chromosome 16, such as *TXNLAB*, *HPR*, and *HP*. These genes play integral roles in modulating the core

biological mechanisms, including cell cycle progression, oxidative balance maintenance, protein conformational dynamics and stability, which all exhibit profound interdependencies with the aging process<sup>36,37</sup>. Moreover, the MAGMA gene-set analysis unveiled the critical involvement of lipid metabolism, CYP2E1-mediated reactions, and liver-specific genes in metabolomic aging, providing valuable insights into the underlying molecular mechanisms and potential therapeutic targets pertinent to this process. The tissue-enrichment analysis further emphasizes the importance of liver in metabolomic aging. This finding implies that strategies focused on preserving or restoring liver health, such as modulating key metabolic pathways, enhancing antioxidant defenses, and stem cell therapy, may have far-reaching systemic benefits in countering metabolomic aging process. Further research is needed to refine these potential strategies and evaluate their efficacy in promoting healthy aging and preventing age-related diseases. Notably, the pathways identified for MetaboAgeMort were distinct from those enriched by genes associated with PhenoAgeAccel or BioAgeAccel<sup>16</sup>, thus reaffirming that the heterogeneous aging patterns observed among individuals might be partly attributed to varying genetic susceptibilities.

Some limitations in this study should also be noted. First, the number of biomarkers captured by the targeted NMR platform is only a fraction of the metabolites in the human plasma. Nevertheless, NMR has the ability to offer highly accurate quantification at a minimal expense, thus facilitating the straightforward implementation of metabolomic clocks in population health. Second, the association between modifiable risk factors and MetaboAgeMort is cross-sectional, and further causal inferences are needed. Third, even though we have taken into account numerous modifiable risk factors, it is conceivable that certain factors may have been unintentionally neglected. Fourth, the majority of participants in the UKB were White British and tended to be healthier and wealthier<sup>38</sup>, and thus, our sample was less representative of the overall UK adult population. It is worth mentioning that the generalizability of our findings to

populations in developing countries or other contexts may be limited.

In conclusion, we have successfully developed a novel metabolomic-based measure of aging known as MetaboAgeMort. This measure has proven to be highly predictive of mortality and various diseases. Of particular significance is the fact that MetaboAgeMort can enhance the predictive accuracy of 10-year mortality beyond conventional risk factors. These findings imply that it has remarkable potential as a comprehensive measure of overall health and risk of mortality in a clinical setting. The potential of body composition, healthy diet, SES, pulmonary function, smoking status, and alcohol intake were highlighted as possible factors for delaying metabolomic aging. Our research has led us to the identification of 99 significant genomic risk loci and 271 genes linked to MetaboAgeMort. This breakthrough sheds new light on the complex genetic underpinnings governing metabolomic aging processes.

## **Methods**

### **Study participants**

The UKB is a large prospective cohort study that comprised over 500,000 participants aged 37-73 years at the time of baseline assessment (2006-2010). Information was collected via touch-screen questionnaires, biological samples, physical measurements, and linked medical or death register records. Detailed study design and methodology were described elsewhere<sup>39</sup>. Ethics for the UKB was approved by the North West Multicenter Research Ethics Committee, and all participants have provided signed informed consent.

### **Plasma metabolomics**

A total of 251 metabolic biomarkers for EDTA plasma samples from a randomly selected subset of approximately 280,000 UKB participants were measured between June 2019 and April 2020 (Phase 1) and April 2020 and June 2022 (Phase 2) using a high-throughput NMR metabolomics platform developed by Nightingale Health Ltd. The metabolic biomarkers span multiple



metabolic pathways, including fatty acids, fatty acid compositions, and lipoprotein lipids in 14 subclasses, as well as various low-molecular weight metabolites, such as ketone bodies, amino acids, and glycolysis metabolites. Detailed protocols for sample collection and methodology for the Nightingale NMR pipeline were described elsewhere<sup>40,41</sup>. The present study considered 249 available metabolic biomarkers (except for glucose-lactate and spectrometer-corrected alanine) and the values of each metabolic biomarker were transformed using natural logarithmic transformation ( $\ln[x+1]$ ) followed by Z-normalisation prior to analysis.

### **Development of MetaboAgeMort and MetaboAgeMortAccel**

Two sequential steps were undertaken in order to create the MetaboAgeMort model. In the first step, we identified aging-related metabolic biomarkers. Initially, We evaluated the associations of each metabolic biomarker (per 1-SD increment) with all-cause mortality using multivariable Cox regression models, with adjustment for chronological age, sex, ethnicity, education level, Townsend deprivation index (TDI), alcohol intake frequency, smoking status, regular exercise, healthy diet, BMI, cholesterol-lowering medication, anti-hypertensive medication, anti-diabetes medication, and prevalent diseases at baseline (i.e., cancer, CVD, hypertension, diabetes mellitus, and chronic obstructive pulmonary disease [COPD]) among all participants ( $n = 239,291$ ), and 185 metabolic biomarkers were selected with Bonferroni correction. Then, we employed a Cox regression model with LASSO penalization--where the hazard of all-cause mortality was regressed on 185 metabolic biomarkers and chronological age--in the training set. Finally, an optimal  $\lambda$  of 0.00097 were selected via ten-fold cross-validation, and 36 variables, including chronological age were assigned nonzero coefficients.

In the second step, we constructed the MetaboAgeMort in the training set by adopting the methodology previously proposed by Levine et al.<sup>18</sup>. We fitted two proportional hazards regression models based on the parametric Gompertz distribution: one used 36 variables selected above as predictors, and the other used only chronological age as a predictor. Based on the two

models, we predicted the 10-year all-cause mortality risk using the cumulative distribution function, respectively. We then converted the mortality risk into units of years by equating the risk from the two models and solving for age, thus obtaining MetaboAgeMort. In general, an individual's MetaboAgeMort represents the chronological age within the general population corresponding to that individual's mortality risk. For example, two individuals are chronologically 40 years old, but one may have a MetaboAgeMort of 45 years and the other a MetaboAgeMort of 35 years, indicating that they have the average mortality risk of someone who is chronologically 45 or 35 years old, respectively. In addition, we calculated a metric, MetaboAgeMortAccel, following the methods by Liu et al<sup>42</sup>. MetaboAgeMortAccel represents the divergence of MetaboAgeMort from chronological age, defined as the residual resulting from a least-squares linear model when regressing MetaboAgeMort on chronological age.

### **Health-related outcomes**

Information on date and cause of death were obtained through the linkage to national death registries. The all-cause and cause-specific mortality (i.e., cancer, CVD, respiratory disease, neurodegenerative disease, digestive disease, and other causes) were determined using the International Classification of Disease (ICD)-10 codes (Supplementary Table 24)<sup>43</sup>. Follow-up time was calculated from the date of baseline assessment to the date of death, loss to follow-up, or end of follow-up (Dec 31, 2022), whichever came first.

Information on diagnoses and medical conditions of the participants were obtained through the linked hospital inpatient record data, self-reported data, and primary care data from the UK National Health Services. The incident diseases were ascertained by the ICD-9 and ICD-10 codes (Supplementary Table 25). Follow-up time was calculated from the date of baseline assessment to the date of first diagnosis of the disease, death, loss to follow-up, or end of follow-up (Oct 31, 2022), whichever came first.

### **MetaboAge and MetaboHealth score**

The MetaboAge score predicts CA (in years) directly<sup>5</sup>. Two different models were utilized: a linear regression model (MetaboAge\_LM) and an ElasticNET regression model (MetaboAge\_EN). The model weights/coefficients were obtained from the most recent publication by Akker et al.<sup>20</sup> In quality control, we excluded three samples with one or more zero values per sample, as well as one or more concentrations that were more than 5 times the SD away from the overall mean of the feature. The MetaboHealth score is a multivariate model predicting all-cause mortality<sup>12</sup>. It was calculated as the weighted sum of 14 log-transformed and cohort-scaled metabolites, using the R-package MiMIR<sup>44</sup>. To avoid infinite values after log-transformation, a value of 1 was added to the metabolites containing any zero.

### **Modifiable factors**

We have taken into account a total of 107 potentially modifiable factors from the UKB baseline survey. Details of processing the factors are presented in Supplementary Table 26. These factors were classified into seven categories: local environment (e.g., greenspace percentage, buffer 1000m), psychosocial (e.g., nervous feelings), SES (e.g., TDI), medical history (e.g., prevalent CVD at baseline), early life and sexual health (e.g., breastfed as a baby), physical measures (e.g., handgrip strength), and lifestyle (e.g., healthy diet).

### **Genome-wide association analysis**

To better understand the genetic mechanisms underlying MetaboAgeMort, we performed a GWAS analysis using the data from UKB v3 genotyping release. The SNPs were excluded if meeting any of the following criteria: (1) minor allele frequency < 0.01, (2) Hardy-Weinberg equilibrium test  $P < 1.0 \times 10^{-6}$ , (3) missing minor allele frequency, or results of Hardy-Weinberg equilibrium test. This particular aspect of the analysis was constrained to individuals of White British descent (n= 226,937). The fastGWA-MLM analysis in the Genome-wide complex trait analysis (GCTA) software (version 1.94.1) was used to perform GWAS analysis<sup>45</sup>. Models

included chronological age, sex, genotype array, and the top 10 principal components as the covariates. The genome-based restricted maximum likelihood (GREML) method in GCTA was used to estimate the SNP-based heritability (variance explained by all the SNPs)<sup>46</sup>.

### **Functional mapping and annotation**

The FUMA online platform (v1.5.2) was applied for functional mapping and annotation of GWAS results (default parameters were used unless explicitly stated otherwise), with annotations derived from the human genome assembly GRCh37 (hg19)<sup>47</sup>. To identify independent genomic risk loci (defined by  $r^2 > 0.6$ ) and variants in linkage disequilibrium (LD) with lead SNPs, the SNP2Gene module was applied using the genetic data of European populations in 1000G phase3 as the reference<sup>48</sup>. Positional mapping with a 10 kilobase (kb) window size was employed to map risk loci to neighboring protein-coding genes.

### **Gene-based association, gene-set, and gene-property analyses with MAGMA**

We performed analyses pertaining to genes and tissue enrichment (i.e., GWGAS, gene-set, and tissue expression analyses) using MAGMA v1.08 within FUMA<sup>49</sup>. For each of the 18,955 protein-coding genes, GWGAS accessed the joint effect of all variants within the gene. Bonferroni correction was used to establish the genome-wide significance threshold ( $P < 0.05/18,955 = 2.64 \times 10^{-6}$ ). Gene-set analysis was further performed using hypergeometric tests for curated gene sets and Gene Ontology (GO) terms from MsigDB v7.0 to discern whether specific biological pathways or cellular functions were implicated in the genetic etiology of MetaboAgeMort, with Bonferroni correction also being utilized. Simultaneously, tissue enrichment analysis was carried out with 30 general and 53 specific tissue types from GTEx v8<sup>50</sup>.

### **Covariates**

Information on chronological age, sex (female or male), ethnicity (White or Non-white), education level (high, intermediate, or low), alcohol intake frequency (never or special occasions only, one to three times per month, one to four times per week, or daily or almost daily), smoking

status (never, previous, or current smoker), regular exercise (yes or no), healthy diet (yes or no), and medication information were collected through questionnaire interview. The TDI was assigned by participants' postcodes, representing SES levels<sup>51</sup>. BMI ( $\text{kg}/\text{m}^2$ ) was calculated as measured weight/height<sup>2</sup>.

### **Statistical analysis**

Baseline characteristics were described using median (interquartile range, IQR) or count (percentage). Two-sample Wilcoxon test for continuous variables and chi-squared test for categorical variables were used to test the differences between the training and the testing set.

The study's roadmap is illustrated in Fig. 1. The Cox proportional hazard model was used to evaluate the association between MetaboAgeMort and all-cause mortality, with adjustment for chronological age, sex, ethnicity, education level, TDI, alcohol intake frequency, smoking status, regular exercise, healthy diet, BMI, medication use, and prevalent diseases at baseline. To further assess robustness, we repeated the analyses by (1) stratified by several demographic, socioeconomic, as well as health behavior factors, (2) excluding participants who died within 5 years of follow-up to reduce the influence of end-of-life metabolomic status, and (3) only including participants who were free of prevalent diseases at baseline to minimize the influence of reverse causality. Proportional hazards of the associations were tested using Schoenfeld's residuals. We then grouped the participants into quartiles of MetaboAgeMortAccel. The Kaplan-Meier plots were drawn to visualise survival curves.

The evaluation of MetaboAgeMort associations with cause-specific mortality and incident diseases was conducted using Fine and Gray's competing risk models<sup>52</sup>. When constructing models evaluating the associations of MetaboAgeMort with incident diseases, we excluded the participants with a specific diagnosis before or at the time of recruitment from the models. Chronological age, sex, ethnicity, education level, TDI, alcohol intake frequency, smoking status, regular exercise, healthy diet, and BMI were used as covariates. The associations of

MetaboAgeMortAccel with mortality and incident diseases were evaluated using same method as that for MetaboAgeMort. The second and the third MetaboAgeMortAccel quartiles (Q2 and Q3) were set as the reference.

Next, receiver operating characteristic (ROC) curves were used to evaluate the utility of MetaboAgeMort for 10-year all-cause mortality risk prediction beyond conventional risk factors. C-statistic and IDI were calculated, in comparison to that of the conventional risk factors model. In addition, we used ROC to compare its predictive utility for 10-year all-cause mortality risk and multiple diseases incidence with several previously trained multi-metabolite scores (i.e., MetaboAge and MetaboHealth score). Moreover, we tested the associations of these multi-metabolite scores with cause-specific mortality and diverse age-related diseases using the same methods employed for MetaboAgeMort.

Multivariable linear regression models were applied to test the response of MetaboAgeMort (response variable) for each modifiable factor (independent variable), with a Bonferroni-corrected significance threshold for identifying top hits ( $P < 0.05/107 = 4.67 \times 10^{-4}$ ). In these analyses, continuous variables were Z-normalised, and results were shown as  $\beta$  coefficients per 1-SD increment in the corresponding factor (based on the availability of each individual factor). Moreover, we divided modifiable factors into tertiles and showed the results using the lowest tertile as a reference. The models were adjusted for chronological age, sex, and ethnicity. Stratified analyses according to chronological age ( $< 60$  and  $\geq 60$  years) and sex (male and female) were conducted, utilising Bonferroni-correction method to determine top hits.

Data analyses and visualisation were all performed in R version 4.2.2. A two-sided  $P$  of  $\leq 0.05$  was considered statistically significant.

## **Funding**

This research was supported by Grants from Research Center of Prevention and Treatment of Senescence Syndrome, School of Medicine Zhejiang University (2022010002), “Pioneer” and “Leading Goose” R&D Programs of Zhejiang Province (2023C03163), the National Natural Science Foundation of China (72374180), the Fundamental Research Funds for the Central Universities, Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province (2020E10004), and Zhejiang University Global Partnership Fund (188170-11103). The funders had no role in the study design; data collection, analysis, or interpretation; in the writing of the report; or in the decision to submit the article for publication.

## **Acknowledgments**

This research was conducted using the UK Biobank resource under application number 61856. We wish to acknowledge the UK Biobank participants who provided the sample that made the data available.

## **Competing interests**

The authors have no conflicts of interest related to this manuscript to declare.

## **Ethics approval and consent to participate**

UK Biobank has approval from the North West Multi-Centre Research Ethics Committee as a Research Tissue Bank approval in 2011 and is renewed every 5 years, which allowed researchers to use data from UK Biobank without an additional ethical clearance. All participants have provided signed informed consent.

## Reference:

1. Kennedy, B. K. *et al.* Geroscience: linking aging to chronic disease. *Cell* **159**, 709–713 (2014).
2. Hamczyk, M. R., Nevado, R. M., Baretino, A., Fuster, V. & Andrés, V. Biological Versus Chronological Aging: JACC Focus Seminar. *J Am Coll Cardiol* **75**, 919–930 (2020).
3. Panyard, D. J., Yu, B. & Snyder, M. P. The metabolomics of human aging: Advances, challenges, and opportunities. *Sci Adv* **8**, eadd6155 (2022).
4. Hwangbo, N. *et al.* A Metabolomic Aging Clock Using Human Cerebrospinal Fluid. *J Gerontol A Biol Sci Med Sci* **77**, 744–754 (2022).
5. van den Akker, E. B. *et al.* Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. *Circ Genom Precis Med* **13**, 541–547 (2020).
6. Hertel, J. *et al.* Measuring Biological Age via Metabonomics: The Metabolic Age Score. *J Proteome Res* **15**, 400–410 (2016).
7. Robinson, O. *et al.* Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell* **19**, e13149 (2020).
8. Finkel, D., Whitfield, K. & McGue, M. Genetic and environmental influences on functional age: a twin study. *J Gerontol B Psychol Sci Soc Sci* **50**, P104-113 (1995).
9. Bernabeu, E. *et al.* Refining epigenetic prediction of chronological and biological age. *Genome Med* **15**, 12 (2023).
10. Wang, F. *et al.* Plasma metabolomic profiles associated with mortality and longevity in a prospective analysis of 13,512 individuals. *Nat Commun* **14**, 5744 (2023).
11. Fischer, K. *et al.* Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* **11**, e1001606 (2014).
12. Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun* **10**, 3346 (2019).
13. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451–459 (2016).
14. Neuffer, P. D. *et al.* Understanding the Cellular and Molecular Mechanisms of Physical Activity-Induced Health Benefits. *Cell Metab* **22**, 4–11 (2015).
15. Fontana, L. & Partridge, L. Promoting health and longevity through diet: from model organisms to humans. *Cell* **161**, 106–118 (2015).
16. Kuo, C.-L., Pilling, L. C., Liu, Z., Atkins, J. L. & Levine, M. E. Genetic associations for two biological age measures point to distinct aging phenotypes. *Aging Cell* **20**, e13376 (2021).
17. Melzer, D., Pilling, L. C. & Ferrucci, L. The genetics of human ageing. *Nat Rev Genet* **21**, 88–101 (2020).
18. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573–591 (2018).
19. Liu, Z. *et al.* A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. *PLoS Med* **15**, e1002718 (2018).
20. Bizzarri, D. *et al.* Technical Report: A Comprehensive Comparison between Different Quantification Versions of



- Nightingale Health's 1H-NMR Metabolomics Platform. *Metabolites* **13**, 1181 (2023).
21. Han, J.-D. J. The ticking of aging clocks. *Trends Endocrinol Metab* **35**, 11–22 (2024).
  22. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat Rev Genet* **23**, 715–727 (2022).
  23. Newman, J. C. & Verdin, E. Ketone bodies as signaling metabolites. *Trends Endocrinol Metab* **25**, 42–52 (2014).
  24. Shemesh, E. *et al.* Circulating ketone bodies and cardiovascular outcomes: the MESA study. *Eur Heart J* **44**, 1636–1646 (2023).
  25. Liu, J. *et al.* Effects of blood triglycerides on cardiovascular and all-cause mortality: a systematic review and meta-analysis of 61 prospective studies. *Lipids Health Dis* **12**, 159 (2013).
  26. Kim, Y. *et al.* Higher diet quality relates to decelerated epigenetic aging. *Am J Clin Nutr* **115**, 163–170 (2022).
  27. Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E. & Boks, M. P. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res Rev* **69**, 101348 (2021).
  28. Liu, W.-S. *et al.* Association of biological age with health outcomes and its modifiable factors. *Aging Cell* **22**, e13995 (2023).
  29. Robertson, T. *et al.* Is socioeconomic status associated with biological aging as measured by telomere length? *Epidemiol Rev* **35**, 98–111 (2013).
  30. Li, H., Han, M., Guo, L., Li, G. & Sang, N. Oxidative stress, endothelial dysfunction and inflammatory response in rat heart to NO<sub>2</sub> inhalation exposure. *Chemosphere* **82**, 1589–1596 (2011).
  31. Thangavel, P., Park, D. & Lee, Y.-C. Recent Insights into Particulate Matter (PM<sub>2.5</sub>)-Mediated Toxicity in Humans: An Overview. *Int J Environ Res Public Health* **19**, 7511 (2022).
  32. Britton, A., Shipley, M., Singh-Manoux, A. & Marmot, M. G. Successful aging: the contribution of early-life and midlife risk factors. *J Am Geriatr Soc* **56**, 1098–1105 (2008).
  33. Gialluisi, A. *et al.* Exploring domains, clinical implications and environmental associations of a deep learning marker of biological ageing. *Eur J Epidemiol* **37**, 35–48 (2022).
  34. Coltell, O. *et al.* Genome-Wide Association Study for Serum Omega-3 and Omega-6 Polyunsaturated Fatty Acids: Exploratory Analysis of the Sex-Specific Effects and Dietary Modulation in Mediterranean Subjects with Metabolic Syndrome. *Nutrients* **12**, 310 (2020).
  35. Djuricic, I. & Calder, P. C. Beneficial Outcomes of Omega-6 and Omega-3 Polyunsaturated Fatty Acids on Human Health: An Update for 2021. *Nutrients* **13**, 2421 (2021).
  36. Dowds, B. C. & Hoch, J. A. Regulation of the oxidative stress response by the hpr gene in *Bacillus subtilis*. *J Gen Microbiol* **137**, 1121–1125 (1991).
  37. Ju, Z. *et al.* TXNL4B regulates radioresistance by controlling the PRP3-mediated alternative splicing of FANCI. *MedComm (2020)* **4**, e258 (2023).
  38. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026–1034 (2017).
  39. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).

40. Würtz, P. *et al.* Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* **186**, 1084–1096 (2017).
41. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* **8**, 192–206 (2015).
42. Liu, Z. *et al.* A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. *PLoS Med* **15**, e1002718 (2018).
43. Han, H. *et al.* Association of a Healthy Lifestyle With All-Cause and Cause-Specific Mortality Among Individuals With Type 2 Diabetes: A Prospective Study in UK Biobank. *Diabetes Care* **45**, 319–329 (2022).
44. Bizzarri, D., Reinders, M. J. T., Beekman, M., Slagboom, P. E. & van den Akker, E. B. MiMIR: R-shiny application to infer risk factors and endpoints from Nightingale Health's 1H-NMR metabolomics data. *Bioinformatics* **38**, 3847–3849 (2022).
45. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**, 1749–1755 (2019).
46. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565–569 (2010).
47. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
48. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
49. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).
50. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. 787903 Preprint at <https://doi.org/10.1101/787903> (2019).
51. Blane, D., Townsend, P., Phillimore, P. & Beattie, A. Health and Deprivation: Inequality and the North. *The British Journal of Sociology* **40**, 344 (1989).
52. Fine, J. P. & Gray, R. J. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* **94**, 496–509 (1999).

# Development of MetaboAgeMort model

## Feature selection

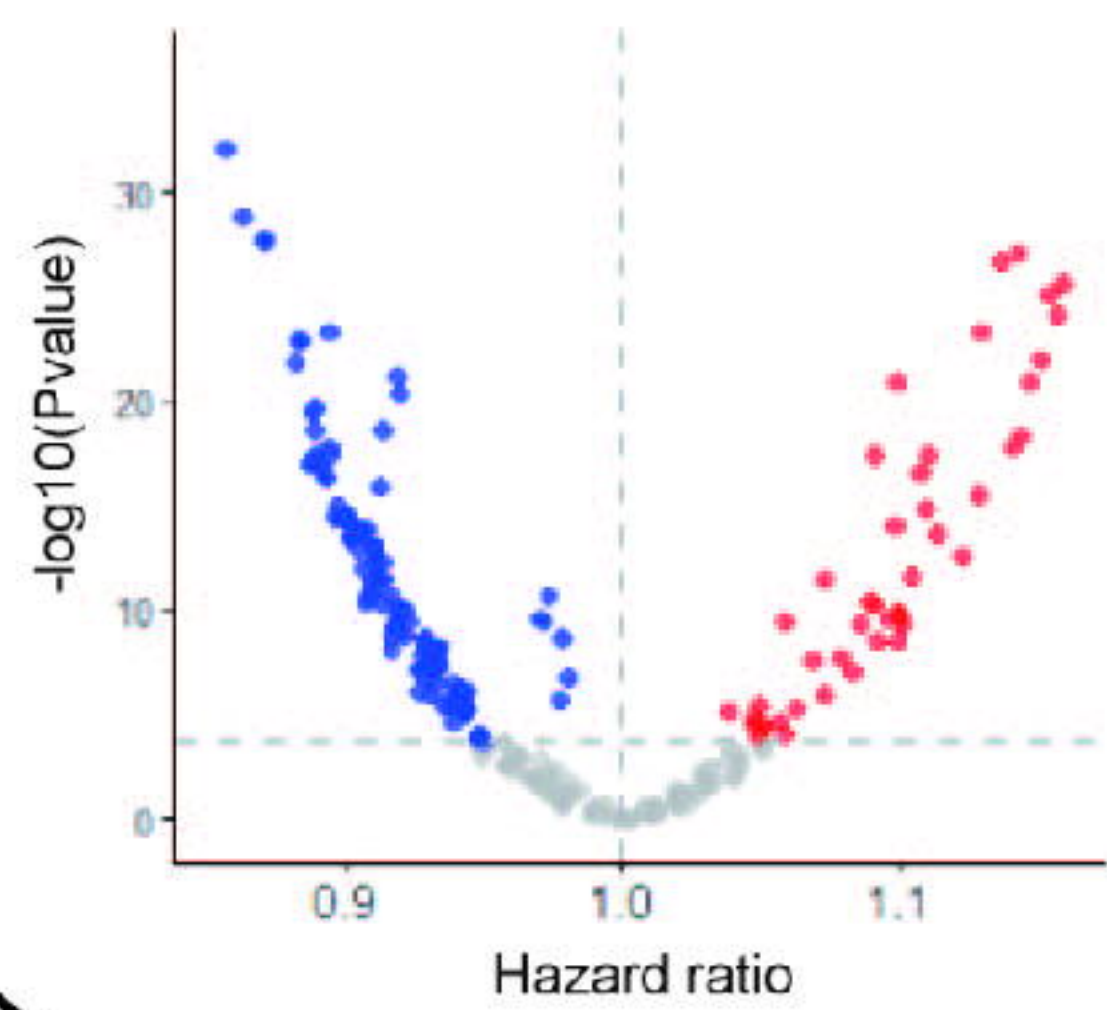
## Model construction

Metabolic biomarkers associated with mortality

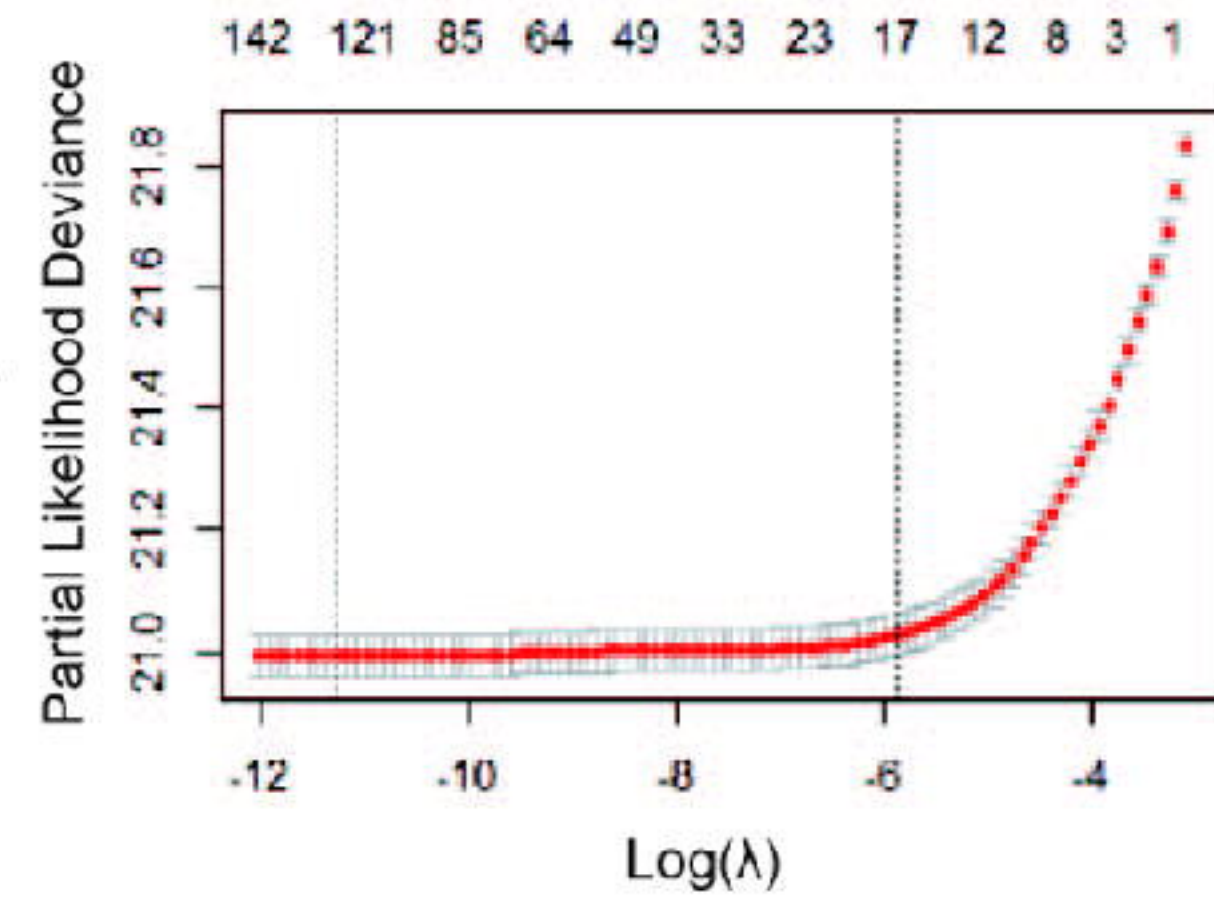
LASSO regression (10-fold cross-validation)

MetaboAgeMort model

MetaboAgeMortAccel



+CA  
→



Gompertz proportional hazard regression models

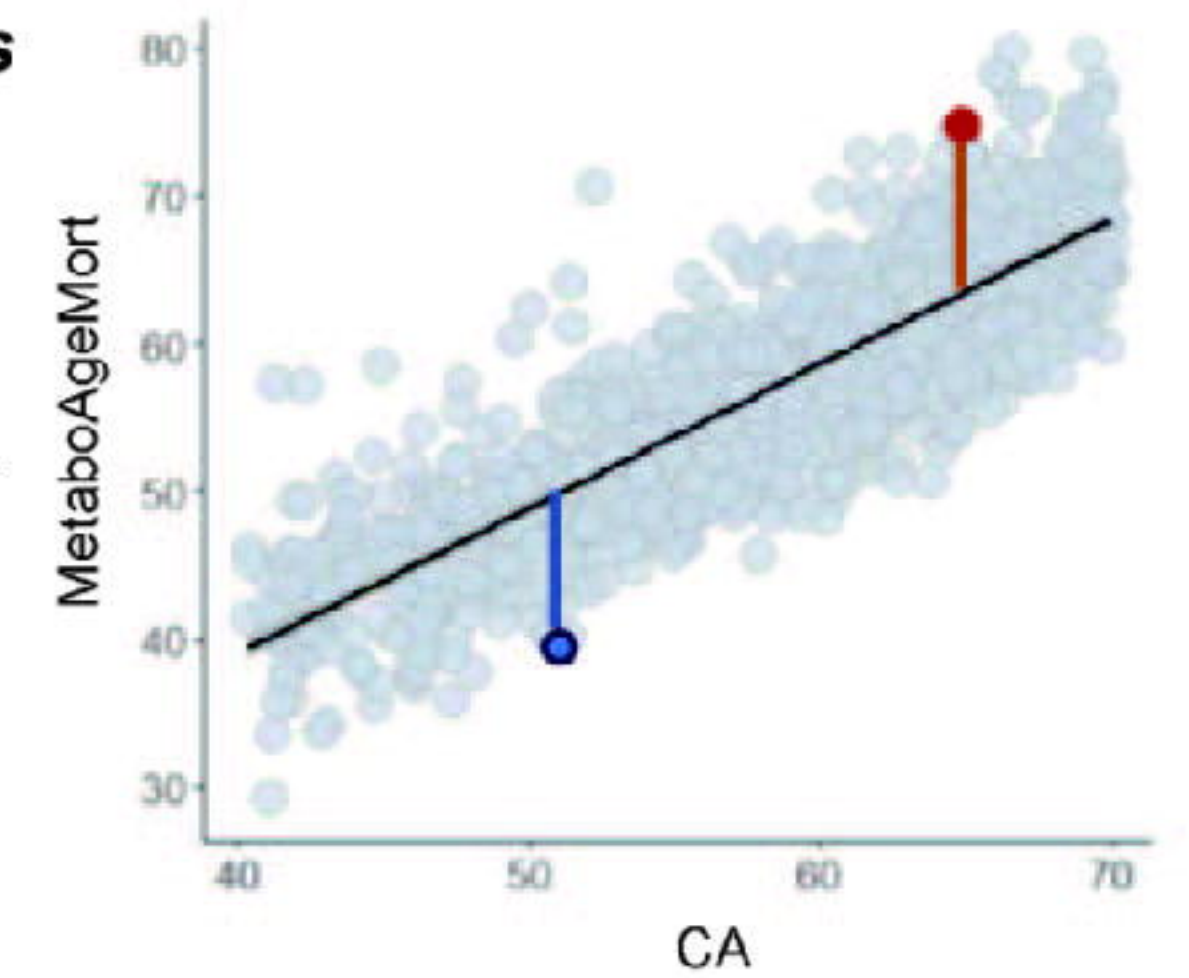
$$\text{CDF}(10, \text{age}) = \text{CDF}(10, x_b)$$

(CDF: cumulative distribution function)

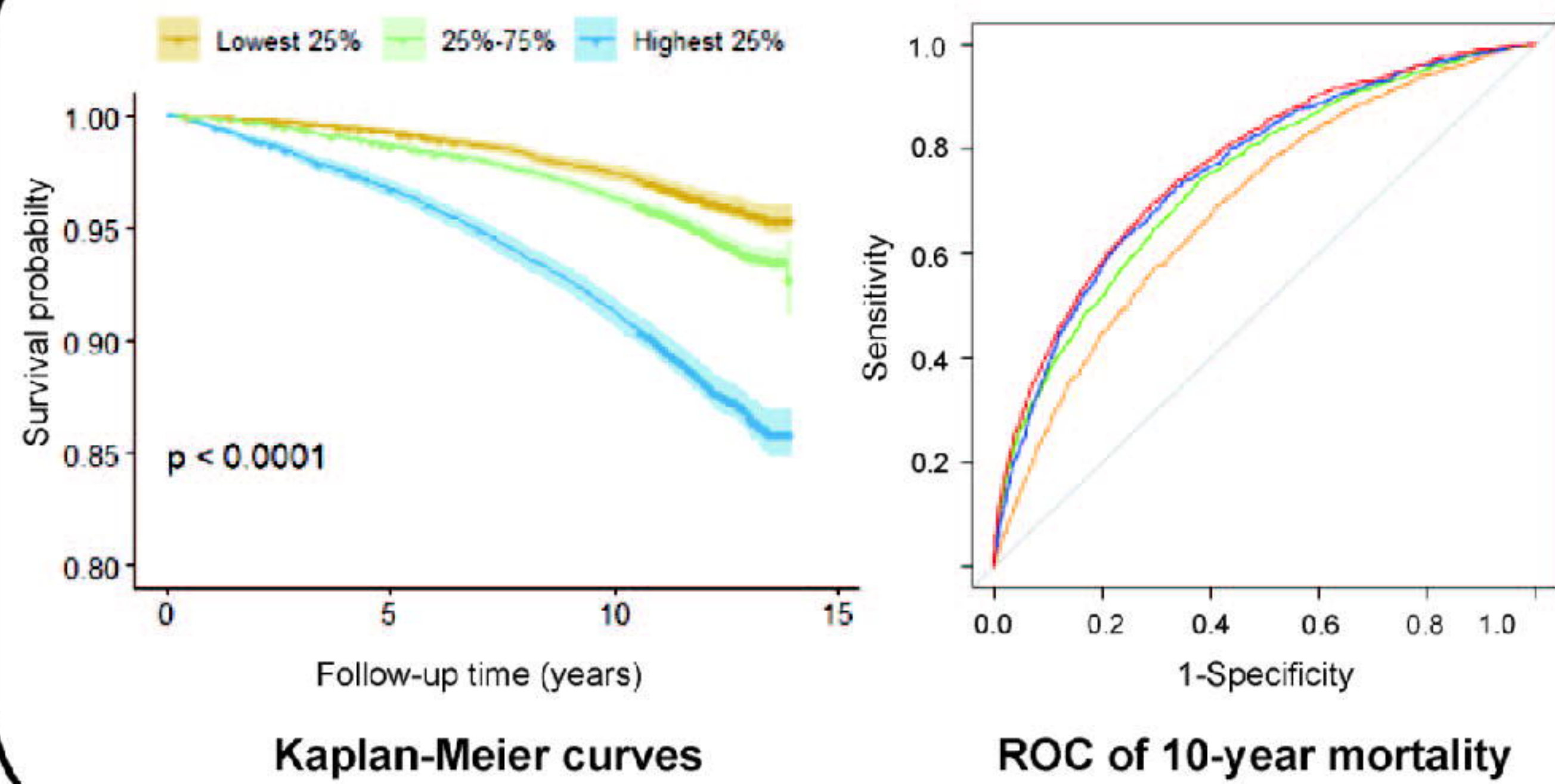
$$\text{MetaboAgeMort} = \frac{\ln\left(-\frac{\ln(1-M)}{\gamma_0^{-1}(e^{\gamma_0 t} - 1)}\right) - \beta_0}{\beta_1}$$

Where:

$$M = 1 - e^{-\left(e^{x_b \gamma_1^{-1}}(e^{\gamma_1 t} - 1)\right)}$$



## Associations with mortality



## Modifiable and unmodifiable factors



## Associations with incident diseases

