Determinants of accelerated metabolomic and epigenetic ageing in a UK cohort

Oliver Robinson[1], Marc Chadeau Hyam[1], Ibrahim Karaman[1], Rui Climaco Pinto[1], Giovanni Fiorito[2], He Gao[1], Andy Heard[1], Marjo-Riitta Jarvelin[1], Mathew Lewis[3], Raha Pazoki[1], Silvia Polidoro[2], Ioanna Tzoulaki[1], Matthias Wielscher[1], Paul Elliott[1], Paolo Vineis[1, 2]

1. MRC- PHE Centre for Environment and Health, Imperial College London, St. Mary's Campus Paddington, W2 1PG, London, United Kingdom.
2. Italian Institute for Genomic Medicine (IIGM, former HuGeF), Via Nizza 52 –, 10126, Turin, Italy.
3. Division of Computational and Systems Medicine, Department of Surgery and Cancer. Imperial College London. South Kensington Campus, SW7 2AZ, London UK

Abstract

Markers of biological ageing have potential utility in primary care and public health. We developed an elastic net regression model of age based on untargeted metabolic profiling across multiple platforms, including nuclear magnetic resonance spectroscopy and liquid chromatography-mass spectrometry in urine and serum (almost 100,000 features assayed), within a large sample (N= 2,239) from the UK occupational Airwave cohort. We investigated the determinants of accelerated ageing, including genetic, lifestyle and psychological risk factors for premature mortality. The metabolomic age model was well correlated with chronological age (r = 0.85 in independent test set) and DNA methylation age. Increased metabolomic age acceleration (p < 0.05) was associated with high alcohol use, overweight or obesity, low income, and depression. We also observed increases in DNA methylation age acceleration associated with anxiety, post-traumatic stress disorder and low income that were of a greater size than for metabolomic age acceleration.

Introduction

Ageing can be defined as the "time-dependent decline of functional capacity and stress resistance, associated with increased risk of morbidity and mortality" [1]. Environmental stressors, including social adversity [2,3], psychological disorders [4,5], and genetic factors[6] may influence the ageing process , leading to differing ageing rates. Traditionally, quantitative assessment of "the rate of ageing" relies on the analysis of mortality curves of populations. However, at the level of a living individual, this method does not allow assessment of the state of ageing (i.e. the state of the functional decline) and a prediction of the risk of morbidity and remaining life expectancy. Therefore, markers of 'biological age' (the ageing state typical of one's chronological age) that can be assessed at any point in the lifespan therefore, may have enormous potential in both personalised medicine and public health. Since ageing is a process that affects almost all tissues and organs of the body and involves cross-talk between multiple physiological systems, there has been increased research into composite markers of ageing, involving multiple parameters [7]. Levine [8] employed 10 biomarkers representing multiple systems to develop a biological age score, that could better

predict mortality than chronological age. Belsky *et al.* [9] used a similar selection of biomarkers measured longitudinally in young adults to develop a biological age score and found that increased pace of ageing was associated with measures of functional decline such as cognitive ability. However, a synthetic indicator of biological age is still lacking. Modern 'omics' platforms have provided new opportunities for the systematic assessment of biological ageing. For example, Horvath [10] and Hannum *et al.* [11] employed genome-wide DNA methylation to develop highly predictive models of age based on multiple methylated CpG loci. Furthermore, it has been shown that 'age acceleration', defined as having a greater DNA methylation age than chronological age, is associated with multiple risk factors of mortality such as low social class, smoking, and alcohol use [3] and is predictive of mortality [12,13]. Agnostic metabolomics is a promising candidate technology to develop biomarkers of ageing. Several metabolomic studies have found strong associations between numerous metabolites and age, although in a limited sample size [14,15] or through employing targeted analyses that give limited coverage of the full metabolome [16] [17,18]. Only the study of Hertel *et al.* [18] combined a small set of markers to provide an overall assessment of biological ageing, observing that the predicted metabolomic age was associated with time to death, after adjustment for chronological age and other risk factors.

In the present study, we have employed untargeted metabolomics across multiple analytical platforms, providing unprecedented metabolome coverage (almost 100,000 features assayed), to develop a predictive model of age, within a large sample from the UK occupational Airwave cohort. A second cohort was used for longitudinal validation of selected metabolic age predictors. We explore the relationship between metabolomic age and DNA methylation age and lifespan associated genetic factors. Furthermore, we investigate the determinants of accelerated ageing, focussing on risk factors of premature mortality, and show that psycho-social risk factors including depression and low income are associated with accelerated metabolomic ageing.

Results

Building and validation of the metabolomic age model

The study population included 2,238 participants of the AIRWAVE cohort that had full metabolomic data. 60.5% of participants were male and mean age was 41.24 years (SD: 9.1, range: 19.2 – 65.2 years). Most participants (97.5%) were of white British ethnicity and 27.8% of participants were educated to degree level. The demographic characteristics of this sample are representative of the wider cohort (Elliott et al 2014). Further covariate information is provided in table 1. Metabolomic data were acquired from both urine and serum samples using multiple Nuclear Magnetic Resonance Spectroscopy (NMR) and Ultra-Performance Liquid Chromatography - Mass Spectrometry (UPLC-MS) platforms, providing in total nine different metabolomic data types (table s1). For purposes of constructing the main predictive model of age through elastic net regression, these data types were combined into one metabolomic dataset, giving a total of 98,824 metabolic features.

In the first stage of model building, an analysis by metabolomic platform (sequentially leaving on platform out each time) indicated that predictive performance (minimisation of mean squared error in 10-fold cross validation) was improved through using only the four following platforms (figure s1): Bruker IVDr Lipoprotein Subclass Analysis derived from NMR in serum ("sBiLISA"), lipid-targeted reverse-phase UPLC-MS in positive mode in serum ("sLPOS"), reverse-phase UPLC-MS in positive mode in urine ("uRPOS") and hydrophilic interaction UPLC-MS in positive mode in urine ("uHPOS") to give a total of 28941 metabolic features. The final predictive model selected 525 predictors from across this set (see table s2 for list of predictors along with table s3 annotation information), including 8 lipoprotein subclasses

from sBiLISA and 219, 104 and 194 features (retention time-m/z pairs) from the sLPOS, uHPOS and uRPOS platforms respectively. The model predicted age with high accuracy (mean absolute error, (MAE) = 1.47 years) in the building data set (80% of data n= 1790), with a correlation between chronological age and predicted age of 0.96 (figure 1a). When this model was applied to the independent validation dataset, consisting of the remaining 20% of study participants (N = 448), the MAE was 3.80 years and the correlation between predicted age and chronological age was 0.85 (figure 1a).

Pathway enrichment analysis, using the *Mummichog* algorithm performed across the UPLC-MS derived model predictors, identified enrichment ($p < 0.05$) in eleven metabolic pathways (table 2): Vitamin E metabolism, Tryptophan metabolism, CoA Catabolism, Urea cycle/amino group metabolism, Lysine metabolism, Carnitine shuttle, Vitamin B5 - CoA biosynthesis from pantothenate, Biopterin metabolism, Drug metabolism - cytochrome P450, Tyrosine metabolism, and Aspartate and asparagine metabolism.

We examined concentration changes of nine metabolites included in our age prediction model, that were available in an independent cohort, the Northern Finnish Birth Cohort 1966, that had serum NMR metabolomic data measured at two ages, 31 and 46 yrs, among 2144 individuals. Eight of these metabolites (89%) changed significantly with age, in the same direction as predicted in the metabolomic age model (table 3).

Metabolomic age, DNA methylation and genetic predictors of longevity

DNA methylation age was assessed for 1102 participants. Demographic characteristics for this sample were similar to those for participants with metabolomic age available (table s4). DNA methylation age predicted chronological age with a MAE of 4.37 years. DNA methylation age was strongly correlated with chronological age (r=0.91, figure 2a) and metabolomic age (n = 837, r = 0.85, figure 2b). Age acceleration scores were derived for both DNA methylation age acceleration (DNAmAA) and metabolomic age acceleration (mAA), as the difference, at a given age, between actual and predicted age. However, no correlation was observed between DNAmAA and mAA (r = 0.02, figure 2c).

Table 4 shows mean age acceleration scores by genotype for 11 single nucleotide polymorphisms (SNPs), that have robust and replicated associations with lifespan [6]. Directions of association between genotype and mAA that were consistent with effects on lifespan (e.g. age acceleration increases with number of effect alleles associated with shorter lifespan or visa versa) were observed for SNPs in the *BSND*, *TRAIP*, *FTO* and *APOC1* genes. Only SNPs in the *FTO* and *APOC1* genes were nominally significant (p= 0.05 for both). For DNAmAA, directions of association with each genotype that were consistent with effects on lifespan were noted for SNPs in genes *BSND* and *APOC1,* with a nominally significant association observed for *APOC1* (p= 0.03).

Risk factors of age acceleration

In bivariate analyses (table 1) we observed increased mAA ($p < 0.05$) among participants who were diabetic, heavy drinkers, overweight or obese, former smokers, or were suffering from depression, anxiety or PTSD. Clinical biomarkers associated with mAA included creatinine, total cholesterol, γ-glutamyl transferase (GGT) apolipoprotein B and glycated haemoglobin (%HBa1C). Regarding dietary intake in the week prior to sampling, those who reported high fish consumption and those in the second or fourth quintiles of the DASH score (compared to the first quintile, the least healthy dietary pattern) also had increased mAA. In bivariate analyses with DNAmAA (table s4), sex was associated at p<0.05, with an increase in DNAmAA of 0.89 (interpretable as years of increase in DNA methylation age, 95% CI: 0.47, 1.30) in men compared to women. Clinical biomarkers associated with

DNAmAA in bivariate analyses included creatinine, high density lipoproteins, GGT and apolipoprotein A.

Table 5 shows adjusted associations with mAA and DNAmAA for non-communicable disease and psychological risk factors (adjusted for sex, ethnicity, study centre, income, hypertension, diabetes, BMI, smoking, alcohol intake, physical activity, DASH score and fish consumption). We observed significant increases ($p<0.05$) in mAA with overweight, obesity, heavy drinking, lower income, depressive symptoms and depression, ranging from 0.35 (95% confidence interval (CI): 0.01, 0.69) for low income compared to those with high income, to 0.97 (95% CI: 0.57, 1.37) for obesity compared to those of normal weight. Significant increases in DNAmAA were observed with heavy drinking, anxiety and PTSD, ranging 0.92 (95% CI: 0.03, 1.80) for anxiety compared to those without anxiety symptoms, to 2.15 (95% CI: 0.31, 4.00) for PTSD compared to those who had not experienced trauma in the past six months.

Discussion

In an important proof of principle study, we have demonstrated in a large nationwide cohort study of working age adults that metabolomic profiling may be used to predict chronological age with high accuracy. We employed a wide range of metabolomic platforms to provide the broadest metabolome coverage yet presented in population based studies. We found that metabolomic age acceleration, defined as having a greater predicted metabolomic age than chronological age, was associated with NCD risk factors including low income, overweight and obesity and high alcohol intake. Mental well-being, particularly when assessed through reported symptoms of depression, was also strongly associated with metabolomic age acceleration. We did not observe an association between epigenetic age acceleration and metabolomic age acceleration, suggesting these measures capture separate aspects of the ageing process. We observed a different pattern of risk factors associated with epigenetic accelerated ageing including being male, heavy drinking, anxiety and PTSD.

The correlation between chronological and predicted age, of our measure of metabolomic ageing (r= 0.86 in the validation dataset), was somewhat lower than that of the Hannum epigenetic age clock in our cohort (r= 0.91) but greater than reported for other biological ageing markers, including the measure based on urinary NMR data [18](r = 0.53 in men and 0.61 in women in validation dataset), the blood transcriptomic clock [19] (r = 0.35- 0.74 depending on cohort) and telomere length (r ~ 0.3, [20]). Biological ageing markers aim to better capture the body's rate of decline or physiological breakdown than chronological age itself, and should therefore also be more predictive of mortality and age-related disease. The associations we observed between accelerated metabolomic ageing and factors known to increase risk of mortality, suggest that metabolomic age may capture this physiological decline.

Strong associations with mAA were observed with overweight and obesity. These conditions are forms of metabolic dysregulation and their additional metabolic burden may increase the rate of decline of the metabolic systems of the body. Genetic predisposition to longevity is associated with low levels of abdominal visceral fat [21] and many different conditions that prolong lifespan in animal models also improve obesity-related conditions. Furthermore, obesity has been linked to telomere shortening, and drastic measures to combat morbid obesity like bariatric surgery can actually cause a recovery in telomere length [22]. Much is now known about the ageing process at the molecular level primarily from experimental work. López-Otín et al.[23] proposed nine 'hallmarks of ageing' that may all be expected to have detectable effects on the metabolome and overlap significantly with the effects of metabolic disorders [24]. For instance, the hallmark 'deregulated nutrient signalling' refers to

pathways that sense and respond to nutrient availability such as "insulin and IGF1 signalling" (IIS) pathway, which is altered among diabetics.

We observed multiple metabolomic pathways enriched among the predictors of our metabolomic clock that reflect fundamental metabolic processes and are closely related to these hallmarks. We observed enrichment of the pathway related to metabolism of Vitamin E, a potent anti-oxidant and anti-inflammatory agent that protects cell membranes from oxidative damage that can induce genome instability [25]. As a primary hallmark, genome instability has far-reaching and complex consequences including altered nutrient sensing, energy metabolism and redox balance [26]. Many human progerias are disorders accompanied by the hyperactivation of DNA repair machinery dependent on nicotinamide adenine dinucleotide (NAD+). This hyperactivation leads to NAD+ depletion, resulting in inhibition of the NAD+-dependent nutrient sensor sirtuin 1 (SIRT1) [27]. Levels of NAD+ are also affected by other factors including circadian rhythm disruption, chronic inflammation [28] and tryptophan metabolism (also enriched among the metabolomic clock predictors). The functional impairment of SIRT1 [29], which limits expression of nuclear genes encoding mitochondrial proteins, leads directly to the hallmark mitochondrial dysfunction. We observed enrichment among the metabolomic clock predictors of the pathways CoA catabolism, Vitamin B5 - CoA biosynthesis from pantothenate and lysine metabolism, which all maintain acetyl-coA levels necessary for mitochondrial reactions, and carnitine shuttle, which is required for the transport of fatty acids for beta-oxidation in the mitochondria. The enrichment of these pathways suggests the importance of the mitochondrial dysfunction hallmark in our metabolomic ageing model. SIRT1 also contributes to regulating the circadian oscillation of acetyl-coA levels [30] which has been linked to the ageing process [31] and epigenetic alterations through acetylation [32]. Mitochondrial fitness further has impact on other ageing hallmarks [24], including genomic instability (dysfunctional mitochondria are major sources of genotoxic ROS), altered intercellular communication (ROS overgeneration is connected to the secretion of inflammatory mediators) and stem cell exhaustion (which are particularly sensitive to ROS [33]). The observed enrichment of the urea cycle and aspartate and asparagine metabolism pathways will also result from perturbation to the Krebs and urea cycles following changes in mitochondrial fitness.

The enrichment of the tryptophan, tyrosine and biopterin metabolic pathways appear to relate to the hallmark 'altered intercellular communication'. Tyrosine is required for signal transduction through incorporation into protein kinases, while tryptophan and biopterin are necessary for synthesis of neurotransmitters including dopamine, norepinephrine, epinepherine, serotonin and melatonin. Alterations to neurotransmitter levels may underlie the associations we observed between mAA with depressive symptoms and depression. Both psychological distress and major depression had similar hazard ratios for mortality in a recent prospective study [4], which would be consistent with the observed increases in mAA for both depressive symptoms and depression. Anxiety was also associated with increased mAA, albeit a smaller increase than observed for depression. This is again consistent with the relative hazard ratios for mortality observed for anxiety and depression in a nation-wide prospective cohort, even after taking suicides and accidental deaths into account [34]. While in this cross-sectional study we cannot disentangle the causal direction between depression and mAA, a study of biological ageing among elderly people found that accelerated biological age was associated with depressive symptoms at baseline and was also predictive of depressive symptoms at follow-up [35]. Consistent evidence demonstrates a bi-directional association between depression and so-called metabolic syndrome, suggesting common pathological roots [36]. Proposed pathophysiological commonalities include abnormal activation of the hypothalamic–pituitary–adrenal (HPA) axis and altered levels of circulating leptin and ghrelin, two peripheral hormones that are classically implicated in the homeostatic

control of food intake. A large body of research has investigated the concept of 'allostatic load' whereby repeated activation of the HPA axis leads to biological 'wear and tear' or physiological decline of downstream metabolic, immune and cardiovascular systems [37]. Many studies have demonstrated a link between social adversity [38,39] and allostatic load and it is theorised that chronic stress associated with low socio-economic position leads to prolonged activation of the HPA axis. We observed that lower income was associated with increased mAA, which may similarly be considered to capture physiological decline.

We observed increases in DNAmAA associated with anxiety, PTSD and low income that were generally of greater size than for mAA. Meta-analyses have shown that both PTSD [5] and low socio-economic position [3] to be associated with increases in DNAmAA. We did not observe any evidence for an association between depression and DNAmAA, suggesting the two ageing measures may be sensitive to separate dimensions of mental health. The DNA methylation clock has been shown to perform well as marker of biological age since it is predictive of all-cause mortality, even after adjusting for chronological age and a variety of known risk factors, and is associated with physical measures of ageing such as frailty and cognitive decline [40]. However, other biological ageing markers may add value in capturing different aspects of the ageing process. Peters *et al.* [19] reported that transcriptomic age was only moderately correlated with DNA methylation age and the different measures were associated with different ageing phenotypes. Similarly, Belsky *et al.* [41] report only weak correlations between telomere length, DNA methylation age, and a composite biomarker-based measure of biological ageing among young adults. While metabolomic and DNA methylation age were correlated in our study, there was no association between mAA and DNAmAA. DNAmAA has been shown to be predictive of cancer related mortality but not CVD [13,40] while the risk factors associated with mAA suggest it may be predictive of cardio-metabolic related disease. Accelerated transcriptomic age was found to be similarly associated with CVD risk factors, although it was not related to mental health [19]. Further research into biological ageing may consider combining markers at different levels of biological organisation to provide a more complete picture of the ageing process.

To provide further external validation of the biological ageing markers we explored associations with genetic factors, selected from SNPs found to be robustly associated with lifespan [6]. The strongest associations were observed for the *FTO* and *APOC1* genes with mAA and the *APOC1* gene for DNAmAA. *APOC1* encodes a member of the apolipoprotein C1 family and plays a key role in lipoprotein metabolism and has been associated with multiple are-related disordered including cognitive decline [42], Alzheimer's disease [43] and heart disease [44]. *FTO* encodes an alpha-ketoglutarate-dependent hydroxylase and is best known for its associations with BMI [45] and obesity [46]. While these findings suggest the role of genetic predisposition to longevity in our biological age markers, they require confirmation in larger, independent populations.

This study had some important limitations. The study was cross-sectional, based on a single biological sampling from participants at a wide range of ages over a single sampling period. It is therefore difficult to separate processes relating to the ageing itself from cohort effects associated with the different environment of people at different ages. This is particularly a problem for analyses of the metabolome which contains both endogenous metabolites related to physiological processes such as ageing and short-lived exogenous metabolites related to factors such as diet and medication. Indeed, we observed that fish consumption, which is associated with reduced risk of mortality [47], actually increased mAA, likely due to the confounding of our model by cohort effects. We addressed these points in two ways: Firstly, we validated some of the metabolomic age predictors that were available in an independent cohort at two timepoints (15 years apart) in the early adult life of the same

individuals. We found that there were highly significant changes in levels of the majority of metabolites we checked, in the same direction as predicted by the metabolomic age model. Secondly, we adjusted associations with mAA for diet that had been assessed through a food diary in the week prior to sampling. Pathways that were enriched in our model were generally related to physiological processes known to be related to ageing, with the possible exception of the drug metabolism pathway. However, medication history was unavailable in this study.

The second limitation was the use of untargeted metabolomics. This limits the potential to apply the full model in separate metabolomic datasets due to differences in retention time and mass accuracy in different runs of spectral acquisition. Furthermore, full laboratory annotation of all predictors was outside the scope of the present study, and may not even be possible for some predictors without current database matches. However, the aim of the study was to develop an overall predictive model to assess metabolic ageing rather than identify individual predictors. Indeed, the nature of the variable selection method used means that an equally valid predictive model can be built on different sets of predictors. We used the *Mummichog* pathway analysis tool to extract information at the pathway level, as the algorithm bypasses laboratory annotation based on the assumption that misidentification will apply equally both to the feature set (metabolites included in the age prediction model) and the reference set (metabolites not selected into the model). The tool has been validated in separate datasets that have also undergone full laboratory annotation [48].

The main strengths of this study also relate to its use of untargeted metabolomics. We incorporated a range of MS platforms able to detect both lipophilic and hydrophilic molecules at low concentrations and NMR platforms able to detect larger structures such as lipoproteins that would be destroyed during MS acquisition. We also analysed both serum and urine that contain different sets of metabolites – more lipophilic molecules in serum and more polar molecules that are present at higher concentrations in urine. Together, we were able to assay a large portion of the metabolome that would not be possible with current targeted methods. Other strengths include the incorporation of genomic and DNA methylation data, the wide age range of participants including those in early adult life where ageing interventions may be most effective [49], and the use of validated psychological instruments. Future work will assess the effects of mAA on functional ageing measures and other health endpoints and assess metabolomic age in longitudinal, repeat samples.

In conclusion, we have developed a predictive indicator of aging based on broad metabolomic analysis among working age adults. We found that while mAA, the difference between metabolomic and chronological age, was not related to DNAmAA, it was associated with mortality risk factors including obesity, diabetes, heavy alcohol use and psycho-social factors including depression, anxiety and lower income. Biological age acceleration may be an important mechanism linking psycho-social stress to age-related disease. Advances in life expectancies have led to an increased prevalence of age-related morbidities. Targeting the process of ageing itself, through changes in living conditions, behaviours or therapeutic interventions, may help more people experience healthy ageing.

Methods

Cohort and covariate information

The Airwave Health Monitoring Study is an occupational cohort of employees of 28 police forces from across Great Britain. Full details of the cohort and methods are available in Elliott et al [50]. The study started recruitment in 2006 and now contains 53,280 participants. The study received ethical approval from the National Health Service Multi-Site Research

Ethics Committee (MREC/13/NW/0588). At the baseline health screening, participants underwent health examination, self-completed a computer questionnaire and provided urine and blood samples. Blood samples were spun at the health clinic and the biological samples were stored in a Thermoporter (LaminarMedica) and sent overnight from the clinics for next-day analysis of standard clinical chemistry tests or were frozen at -80 °C long term storage. DNA samples and plasma for metabolomic analysis were extracted from blood collected in EDTA tubes.

Important covariates in the analysis were categorised from self-report or clinical data as follows: Ethnicity was defined as 'white' or otherwise. Marital status was defined as living with partner or otherwise. Income was defined as low, medium or high, based on terciles of total net household income after adjustment for the number of dependant household members. Education was defined as low (completed GCSEs or equivalent only), medium (completed 'A' levels or equivalent only) or high (completed university or higher degree). Alcohol use was classed as non-drinker, moderate drinker (≤ 14 alcohol units/week for women and ≤ 21 alcohol units/week for men) or heavy drinker (> 14 alcohol units/week for women and >21 alcohol units/week for men) . Hypertension was defined as ether reported diagnosis or systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg. Diabetic status was defined as normal (no diagnosis and HbA1c < 6.5%), or diabetic (diagnosis or HbA1c ≥ 6.5%). Physical activity was defined as low, moderate or high based on the scoring protocol of the International Physical Activity Questionnaire [51] .

Psychological instruments

The Patient Health Questionnaire – 9 depression questionnaire was used to define participants as "normal (i.e. no depression)", "minimal symptoms of depression" or as a "depression case" [52]. The Hospital Anxiety and Depression Scale questionnaire was used to assess anxiety levels as "normal (i.e. no anxiety)", "borderline" and "anxiety case" [53]. Participants were defined as job strained if they had levels of 'job demand' above the median of the whole cohort and levels of 'job latitude' below the median of the cohort based of a subset of six items from the Job Content Questionnaire [54]. Subjective feelings of chronic fatigue were assessed as "low", "medium" or "high" according to the questionnaire of [55]. Participants were asked if they had experienced a work-related traumatic incident in the previous six months. Those who reported a traumatic incident were then asked to complete a brief screening instrument for post-traumatic stress disorder (PTSD) [56]. Participants were thus classed into three categories: "not experienced traumatic incident in past 6 months", "experienced traumatic incident in past 6 months without leading to PTSD", and "experienced traumatic incident in past 6 months leading to PTSD",

Assessment of Diet

Dietary intake was measured using validated 7-day estimated weight food diaries as fully described previously [57]. Nutritional intake was calculated using Dietplan6.7 software (Forestfield Software, Horsham, UK) which is based on the McCance and Widdowson's 6th Edition Composition of Foods UK Nutritional Data set (UKN) by a team of trained coders trained to match food and drink items to the UKN database code and a portion size.

Energy adjusted average consumption of fruit, vegetables, red meat, processed meat, wholegrain and dairy over the week was categorised into tertiles. For fish, consumption was divided into none, medium (below median consumption among consumers) and high (above median consumption). Total average energy consumption was categorised as low, medium and high based on separate tertiles for men and women. Two overall dietary scores were calculated: The Dietary Approaches to Stop Hypertension (DASH) diet score divided into quintiles [58] and the Mediterranean Diet score as a continuous measure [59].

Metabolomic data acquisition

Metabolomic analysis of serum and urine was performed at the National Phenome Centre, based at Imperial College London. Samples were randomly sorted into batches of 80 and thawed to 4°C, centrifuged to remove particulate matter, and the supernatant dispensed across dedicated 96-well plates for each assay. Study-Reference (SR) samples, a pool of all samples for each matrix in the study, and Long-Term Reference (LTR) samples, a pool of samples external to study, were included in each analytical run to allow for quantification and correction of technical variation. Samples are prepared and analysed daily in batches of 80 study samples with the addition of 4 quality controls (2 SR and 2 LTR). Samples were maintained at 4°C during preparation for, and while awaiting, acquisition.

Acquisition of Nuclear Magnetic Resonance Spectroscopy (NMR) profiles (the NOESY experiment in urine and the CPMG experiment in serum) was conducted as described in [60]. Lipoprotein parameters were generated by the Bruker B.I.-LISA (Bruker IVDr Lipoprotein Subclass Analysis platform, derived from NMR of serum. Spectra were acquired at 600 MHz with Bruker Ascend 600 magnets and Avance III HD consoles configured to the Bruker IVDr specification (Bruker Corporation, Billerica, MA, USA).

Ultra-Performance Liquid Chromatography - Mass Spectrometry (UPLC-MS) acquisitions were conducted in batches of up to 1000 study-samples, interleaved with alternating SR and LTR samples every five injections (16 per 80 samples), each batch was flanked by a serial dilution of the SR sample to assess linearity of response. Multiple analytical experiments were performed to increase metabolomic coverage. Hydrophilic interaction chromatography was performed in both urine and serum as described in [61]. Reversed-phase chromatography was performed on urine samples in both positive and negative modes as described in [61]. Lipid-targeted reverse-phase chromatography was applied in serum ionised in both positive and negative modes as described in [62]. All UPLC-MS profiling assays were acquired on Waters G2-S ToF mass spectrometers, with Acquity UPLC chromatography systems (Waters Corporation, Milford, MA, USA).

Metabolomic data processing

NMR spectra were automatically processed in TopSpin 3.2, followed by a suite of in house scripts [60]. Each spectrum was automatically checked, before all spectra were aligned to a common reference scale. Analytical quality was further assessed manually on four factors: Line width of less than 0.9 Hz, quality of water-suppression, even baseline signal and accurate chemical shift referencing. Urine samples were referenced to an internal spiked standard 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) at 0 ppm. Plasma samples were referenced to the α-anomeric glucose doublet at 5.233 ppm. Spectra were aligned to a common reference scale, running from 10 to -1 ppm, and interpolated onto a common 20,000 point grid. Lipoprotein parameters were validated according to Bruker's B.I.-LISA protocols.

Chromatograms and mass spectra instrument raw files were imported into Progenesis QI (Waters Corp. Milford, MA, USA) for retention-time alignment and feature detection. Progenesis QI was configured to align retention time to the central LTR sample of the acquisition. Peak detection was configured with a minimum chromatographic peak width of 0.01 minutes, and automatic noise detection set to the minimum threshold of 1. Peaks arising from isotopes and chemical adducts were automatically resolved according to the observed m/z and chromatographic peak shape, and peaks areas integrated. Further processing and filtering of UPLC-MS profiling datasets was conducted with in-house scripts, and used to account for analytical run-order effects and remove noise from each dataset. Analytical run-order effects were accounted for with an adaption of the method described in

[63]. A robust LOWESS regression was generated per-feature, based on the SR samples, in run-order, with the window scaled to include 21 SR samples. The smoothed response values for each feature were then interpolated to the intermediate study sample injections using simple linear interpolation. Finally, the median intensity of each feature in each analytical batch was aligned. Extracted features spuriously arising from analytical noise were removed from the dataset by a pair of approaches, both applied on a per-feature basis. First, a serial dilution of the study reference sample was used to assess the linearity of responses of each feature. Detected features were correlated to their expected intensity in the dilution series, and those features showing a Pearson's r of less than 0.7 were excluded from further analysis. Second, the relative standard deviation (RSD) of each feature across the study reference samples was calculated, and those features where the RSD exceeded 30%, or the observed biological variance was less than 1.5 times the RSD, were excluded.

Metabolomic age model

Untargeted NMR datasets were glog-transformed [64], the quantified BiLISA data was log-transformed, and the UPLC-MS data were log transformed, following unit addition to every value to allow transformation of zero values. Data were then mean centred and scaled to unit variance.

A predictive model of metabolomic age was constructed using elastic net regression [65] in the "glmnet" package [66] in R. The model was fitted on metabolic features from across all metabolomic datasets, using a multi-step process on 80% of the data (the training dataset). The remaining 20% was reserved for assessment of the predictive ability (Pearson's correlation between predicted and chronological age) of the model in an independent dataset (the test dataset). The steps were as follows:

Step 1 Parameterisation: Elastic net model parameters, α (that defines mixing between lasso and ridge penalties) and λ (overall strength of penalty), were found following 10-fold cross validation. A line search across α, between 0 and 1 in 0.01 increments, was performed to find the minimum mean cross-validated error (MSE) using the optimal value of λ found using the 'cvfit' command for each α value.

Step 2 Leave platform out analysis: Due to potential redundancy between metabolomic datasets, we performed the parameterisation step above on data with one metabolomic platform left out each time. Platforms were removed from further analysis if model performed better (lower MSE) with their exclusion. We continued this process leaving further platforms out each time until no improvement in MSE was observed.

Step 3 Stability analysis: Using the selected metabolomic datasets, we repeated elastic net regression on 100 subsamples of the training dataset (a random selection of 80% each time). The metabolic features selected in each model was stored for each iteration.

Step 4 Metabolomic data restriction: On the same subsample for 101 iterations, the number of metabolic features available to build an elastic net model was restricted by the percentage of iterations in step 3 that a feature was selected, moving from 100% to 0%, in 1% decrements for each subsequent iteration. The correlation between predicted and chronological age in remaining 20% of training set was stored for each iteration and the percentage restriction value that gave the best correlation, was chosen for the final metabolic feature restriction in step 5.

Step 5 Final model building: On the complete training dataset, a final elastic net model was constructed using metabolic features restricted to those present in a set percentage of models, as found in step 4.

Metabolomic age acceleration (metAA) was defined as the difference between chronological age and predicted age, adjusted on actual age as previously defined for DNA methylation age acceleration [10]. That is, we define mAA as the residuals of a linear regression between the chronological age and predicted age difference, with chronological age itself.

Metabolic feature and pathway annotation

Tentative annotations were provided for mass-spectrometry based metabolic features bases on m/z searches across the Human metabolome database [67], for the ion forms M+2H, M+H+NH4, M+NH4, M+H, M+ACN+H, M+CH3OH+H, M+Na, M+K, 2M+H at ± 8 ppm mass tolerance.

For five UPLC-MS based metabolic features that were both tentatively annotated by exact mass within our metabolomic age model and also available in repeat measurements within the Northern Finnish Birth Cohort dataset, we performed further annotation procedures. Two of these annotations, for citrate (as in-source fragmentation product) and leucine (M+Na ionic form), were supported by matching retention times and accurate mass to an internal reference standard database.

Significantly enriched metabolic pathways were predicted using the mummichog program [48]. The algorithm searches tentative compound lists from metabolite reference databases against an integrated model of human metabolism to identify functional activity. Fisher's exact tests and permutation are used to infer p-values for likelihood of pathway enrichment among significant features as compared to pathways identified among the entire compound set present in reference list (the entire metabolome dataset), considering the probability of mapping the significant m/z features to pathways. Mummichog parameters were set to match against ions included in the 'generic positive mode' setting at ± 8 ppm mass tolerance.

Metabolite validation in the Northern Finnish Birth Cohort 1966

The Northern Finnish Birth Cohort 1966 is a prospective birth cohort that sampled 12,058 live births in 1966, including 96.3% of all births in the regions of Oulu and Lapland in Finland [68]. Fasting blood samples were collected at follow-up of participants at ages 31 and 46 yrs and stored at -80 °C for subsequent biomarker profiling. A high-throughput NMR metabolomics platform was used for the analysis of 87 metabolic measures [69]. This metabolomics platform provides simultaneous quantification of routine lipids and lipid concentrations of 14 lipoprotein subclasses and major sub-fractions, and further quantifies abundant fatty acids, amino acids, ketone bodies and gluconeogenesis-related metabolites in absolute concentration units.

We assessed changes of nine metabolites, that were available in this dataset and also included in our predictive model, between these two sampling points using 1-tailed t-tests.

DNA methylation analysis

For the microarray, bisulphite conversion of 500 ng of each DNA sample was performed using the EZ DNA Methylation-Lightning™ Kit according to the manufacturer's protocol (Zymo Research, Orange, CA). Then, bisulfite-converted DNA was used for hybridization on the Infinium HumanMethylation EPIC BeadChip, following the Illumina Infinium HD Methylation protocol. Briefly, a whole genome amplification step was followed by enzymatic end-point fragmentation and hybridization to HumanMethylation EPIC BeadChips at 48°C for 17 h, followed by single nucleotide extension. The incorporated nucleotides were labelled with biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP). After the extension step and staining, the BeadChip was washed and scanned using the Illumina

HiScan SQ scanner. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-sample data using different internal controls that are present on the HumanMethylation EPIC BeadChip and internal background probes. The methylation score for each CpG was represented as a β-value according to the fluorescent intensity ratio representing any value between 0 (unmethylated) and 1 (completely methylated).

DNA methylation (DNAm) data were pre-processed and normalized using in-house software written for the R statistical computing environment, including background and color bias correction, quantile normalization, and Beta MIxture Quantile dilation (BMIQ) procedure to remove type I/type II probes bias, as described elsewhere [3]. DNAm levels were expressed as the ratio of the intensities of methylated cytosines over the total intensities (β values). Cross-reactive and polymorphic probes - with minor allele frequency greater than 0.01 in Europeans [70] - were excluded. Methylation measures were set to missing if the detection p-value was greater than 0.01. Samples with the bisulfite conversion control fluorescence intensity lower than 10,000 for both type I and type II probes and those with total call rate lower than 95% were excluded. Finally, samples were excluded if the predicted sex (based on chromosome X methylation) did not match that self-reported.

DNA methylation age was computed according to the algorithm described by Hannum et al. [11] based on a set of 71 blood-specific age-associated CpG sites. We used this algorithm, rather than the algorithm of Hovarth, since it was developed specifically for blood samples and found to be the most predictive of mortality [12]. Age acceleration (AA) was defined as the difference between epigenetic and chronological age. Since AA could be correlated with chronological age and WBC percentage, we computed the so-called intrinsic epigenetic age acceleration [12], which is defined as the residuals from the linear regression of AA with chronological age and blood cell counts (measured using flow cytometry) for neutrophils, lymphocytes, monocytes and eosinophils.

Genotyping

Genotyping was performed on the Illumina Infinium HumanCoreExome-12v1-1 BeadChip and quality control filters including call rate (>=97%), heterozygosity rate (<=3SD from the mean) were applied on the samples. Duplicated and second-degree relatives were further excluded and 14,062 samples of European ancestry based on principle component analysis remained. Markers were removed for high missing rate (>2%), deviation from Hardy-Weinberg equilibrium (P<1E-5) or minor allele frequency below 1%, resulting in 254,027 high-quality and common markers. Imputation was performed using the Haplotype Reference Consortium (HRC) panel (version r1.1 2016).

We selected 11 SNPs, previously associated with lifespan [6], and tested their associations with both DNAmAA and mAA, in bivariate linear models. DNAmAA or mAA was used as the dependent variable and the dosage of the effect allele for each SNP (i.e. 0,1 or 2) was used as the independent variable.

Analysis of risk factors of biological age acceleration

We analysed associations between mortality risk factors, including psychosocial factors, and age acceleration scores in separate adjusted linear regression models. The adjustment set, chosen *a priori* and included in all models was: sex, ethnicity, study centre, income, hypertension, diabetes, BMI, smoking, alcohol intake, physical activity, DASH score and fish consumption.

Acknowledgements

References

1   Burkle, A. *et al.* MARK-AGE biomarkers of ageing. *Mechanisms of Ageing and Development* **151**, 2-12, doi:10.1016/j.mad.2015.03.006 (2015).

2   Stringhini, S. *et al.* Socioeconomic status and the 25 x 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. *Lancet (London, England)* **389**, 1229-1237, doi:10.1016/s0140-6736(16)32380-7 (2017).

3   Fiorito, G. *et al.* Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. *Scientific reports* **7**, 16266, doi:10.1038/s41598-017-16391-5 (2017).

4   Chiu, M. *et al.* Mortality risk associated with psychological distress and major depression: A population-based cohort study. *Journal of Affective Disorders* **234**, 117-123, doi:10.1016/j.jad.2018.02.075 (2018).

5   Wolf, E. J. & Morrison, F. G. Vol. 19   (Current Psychiatry Reports, 2017).

6   McDaid, A. F. *et al.* Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nature Communications* **8**, doi:10.1038/ncomms15842 (2017).

7   Jylhävä, J., Pedersen, N. L. & Hägg, S. Vol. 21   29-36 (The Authors, 2017).

8   Levine, M. E. Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* **68**, 667-674, doi:10.1093/gerona/gls233 (2013).

9   Belsky, D. W. *et al.* Quantification of biological aging in young adults. *Proc.Natl.Acad.Sci.U.S.A* **112**, E4104-E4110, doi:10.1073/pnas.1506264112 (2015).

10   Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14**, R115-R115, doi:10.1186/gb-2013-14-10-r115 (2013).

11   Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).

12   Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

13   Dugue, P. A. *et al.* Association of DNA Methylation-Based Biological Age With Health Risk Factors and Overall and Cause-Specific Mortality. *American journal of epidemiology* **187**, 529-538, doi:10.1093/aje/kwx291 (2018).

14   Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. & Yanagida, M. Individual variability in human blood metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 4252-4259, doi:10.1073/pnas.1603023113 (2016).

15   Rist, M. J. *et al.* Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS ONE* **12**, 1-21, doi:10.1371/journal.pone.0183228 (2017).

16   Auro, K. *et al.* A metabolic view on menopause and ageing. *Nature Communications* **5**, 1-11, doi:10.1038/ncomms5708 (2014).

17   Yu, Z. *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* **11**, 960-967, doi:10.1111/j.1474-9726.2012.00865.x (2012).

18   Hertel, J. *et al.* Measuring Biological Age via Metabonomics: The Metabolic Age Score. *Journal of Proteome Research* **15**, 400-410, doi:10.1021/acs.jproteome.5b00561 (2016).

19   Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nature Communications* **6**, doi:10.1038/ncomms9570 (2015).

20   Muezzinler, A., Zaineddin, A. K. & Brenner, H. A systematic review of leukocyte telomere length and age in adults. *Ageing research reviews* **12**, 509-519, doi:10.1016/j.arr.2013.01.003 (2013).

21   Sala, M. L. *et al.* Genetically determined prospect to become long-lived is associated with less abdominal fat and in particular less abdominal visceral fat in men. *Age and Ageing* **44**, 713-717, doi:10.1093/ageing/afv063 (2015).

22   Laimer, M. *et al.* Telomere length increase after weight loss induced by bariatric surgery: results from a 10 year prospective study. *International Journal Of Obesity* **40**, 773, doi:10.1038/ijo.2015.238 (2015).

23   López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Vol. 153 (2013).

24   López-Otín, C., Galluzzi, L., Freije, J. M. P., Madeo, F. & Kroemer, G. Vol. 166 802-821 (2016).

25   Claycombe, K. J. & Meydani, S. N. Vitamin E and genome stability. *Mutation research* **475**, 37-44 (2001).

26   Garinis, G. A., van der Horst, G. T. J., Vijg, J. & H.J. Hoeijmakers, J. DNA damage and ageing: new-age ideas for an age-old problem. *Nature Cell Biology* **10**, 1241, doi:10.1038/ncb1108-1241 (2008).

27   Fang, Evandro F. *et al.* Defective Mitophagy in XPA via PARP-1 Hyperactivation and NAD+/SIRT1 Reduction. *Cell* **157**, 882-896, doi:https://doi.org/10.1016/j.cell.2014.03.026 (2014).

28   Verdin, E. NAD+ in aging, metabolism, and neurodegeneration. *Science* **350**, 1208-1213, doi:10.1126/science.aac4854 (2015).

29   Gomes, Ana P. *et al.* Declining NAD+ Induces a Pseudohypoxic State Disrupting Nuclear-Mitochondrial Communication during Aging. *Cell* **155**, 1624-1638, doi:https://doi.org/10.1016/j.cell.2013.11.037 (2013).

30   Sahar, S. *et al.* Circadian Control of Fatty Acid Elongation by SIRT1 Protein-mediated Deacetylation of Acetyl-coenzyme A Synthetase 1. *Journal of Biological Chemistry* **289**, 6091-6097, doi:10.1074/jbc.M113.537191 (2014).

31   Chang, H.-C. & Guarente, L. SIRT1 Mediates Central Circadian Control in the SCN by a Mechanism that Decays with Aging. *Cell* **153**, 1448-1460, doi:https://doi.org/10.1016/j.cell.2013.05.027 (2013).

32   Su, X., Wellen, K. E. & Rabinowitz, J. D. Metabolic control of methylation and acetylation. *Current Opinion in Chemical Biology* **30**, 52-60, doi:https://doi.org/10.1016/j.cbpa.2015.10.030 (2016).

33   Ito, K. *et al.* Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nature Medicine* **12**, 446, doi:10.1038/nm1388

https://www.nature.com/articles/nm1388#supplementary-information (2006).

34   Meier, S. M. *et al.* Increased mortality among people with anxiety disorders: total population study. *The British Journal of Psychiatry* **209**, 216-221, doi:10.1192/bjp.bp.115.171975 (2016).

35   Brown, P. J. *et al.* Biological Age, Not Chronological Age, is Associated with Late Life Depression. *The Journals of Gerontology: Series A* **00**, 1-7, doi:10.1093/gerona/glx162 (2017).

36   Marazziti, D., Rutigliano, G., Baroni, S., Landi, P. & Dell'Osso, L. Vol. 19   293-304 (2014).

37   McEwen, B. S. & Seeman, T. Protective and damaging effects of mediators of stress. Elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences* **896**, 30-47 (1999).

38   Castagne, R. *et al.* Allostatic load and subsequent all-cause mortality: which biological markers drive the relationship? Findings from a UK birth cohort. *European journal of epidemiology*, doi:10.1007/s10654-018-0364-1 (2018).

39   Dowd, J. B., Simanek, A. M. & Aiello, A. E. Socio-economic status, cortisol and allostatic load: a review of the literature. *International journal of epidemiology* **38**, 1297-1309, doi:10.1093/ije/dyp277 (2009).

40   Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biology* **17**, 0-22, doi:10.1186/s13059-016-1030-0 (2016).

41   Belsky, D. W. *et al.* Telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: Do they measure the same thing? *bioRxiv*, 071373-071373, doi:10.1101/071373 (2016).

42   De Jager, P. L. *et al.* A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiology of aging* **33**, 1017.e1011-1015, doi:10.1016/j.neurobiolaging.2011.09.033 (2012).

43   Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Translational psychiatry* **8**, 99, doi:10.1038/s41398-018-0150-6 (2018).

44   van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation research* **122**, 433-443, doi:10.1161/circresaha.117.312086 (2018).

45   Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature genetics* **49**, 1458-1467, doi:10.1038/ng.3951 (2017).

46   Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics* **45**, 501-512, doi:10.1038/ng.2606 (2013).

47   Zhao, L. G. *et al.* Fish consumption and all-cause mortality: a meta-analysis of cohort studies. *European journal of clinical nutrition* **70**, 155-161, doi:10.1038/ejcn.2015.72 (2016).

48   Li, S. *et al.* Predicting Network Activity from High Throughput Metabolomics. *PLOS Computational Biology* **9**, e1003123, doi:10.1371/journal.pcbi.1003123 (2013).

49   Moffitt, T. E., Belsky, D. W., Danese, A., Poulton, R. & Caspi, A. The Longitudinal Study of Aging in Human Young Adults: Knowledge Gaps and Research Agenda. *The journals of gerontology. Series A, Biological sciences and medical sciences* **72**, 210-215, doi:10.1093/gerona/glw191 (2017).

50   Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in Great Britain: Rationale, design and methods. *Environmental research* **134C**, 280-285, doi:10.1016/j.envres.2014.07.025 (2014).

51   The IPAQ group. *International Physical Activity Questionnaire*, <http://www.ipaq.ki.se/> (2016).

52   Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* **16**, 606-613 (2001).

53   Zigmond, A. S. & Snaith, R. P. The hospital anxiety and depression scale. *Acta psychiatrica Scandinavica* **67**, 361-370 (1983).

54   Karasek, R. *et al.* The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of occupational health psychology* **3**, 322-355 (1998).

55   Vercoulen, J. H. *et al.* Dimensional assessment of chronic fatigue syndrome. *Journal of psychosomatic research* **38**, 383-392 (1994).

56   Brewin, C. R. *et al.* Brief screening instrument for post-traumatic stress disorder. *The British journal of psychiatry : the journal of mental science* **181**, 158-162 (2002).

57   Gibson, R. *et al.* Dietary assessment of British police force employees: a description of diet record coding procedures and cross-sectional evaluation of dietary energy intake reporting (The Airwave Health Monitoring Study). *BMJ open* **7**, e012927, doi:10.1136/bmjopen-2016-012927 (2017).

58   Fung, T. T. *et al.* Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Archives of internal medicine* **168**, 713-720, doi:10.1001/archinte.168.7.713 (2008).

59   Trichopoulou, A., Costacou, T., Bamia, C. & Trichopoulos, D. Adherence to a Mediterranean diet and survival in a Greek population. *The New England journal of medicine* **348**, 2599-2608, doi:10.1056/NEJMoa025039 (2003).

60   Dona, A. C. *et al.* Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* **86**, 9887-9894, doi:10.1021/ac5025039 (2014).

61   Lewis, M. R. *et al.* Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. *Analytical Chemistry* **88**, 9004-9013, doi:10.1021/acs.analchem.6b01481 (2016).

62   Sarafian, M. H. *et al.* Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry. *Anal Chem* **86**, 5766-5774, doi:10.1021/ac500317c (2014).

63   Zelena, E. *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal Chem* **81**, 1357-1364, doi:10.1021/ac8019366 (2009).

64   Parsons, H. M., Ludwig, C., Gunther, U. L. & Viant, M. R. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC bioinformatics* **8**, 234, doi:10.1186/1471-2105-8-234 (2007).

65   Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320, doi:doi:10.1111/j.1467-9868.2005.00503.x (2005).

66   Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* **33**, 1-22 (2010).

67   Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* **46**, D608-d617, doi:10.1093/nar/gkx1089 (2018).

68   Rantakallio, P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatric and perinatal epidemiology* **2**, 59-88 (1988).

69   Soininen, P., Kangas, A. J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation. Cardiovascular genetics* **8**, 192-206, doi:10.1161/circgenetics.114.000216 (2015).

70   Chen, Y. A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203-209, doi:10.4161/epi.23470 (2013).

Tables

Table 1: Demographic and covariate information of participants with metabolomic data, and bivariate associations with metabolomic age acceleration

		N (%) or Mean (SD)	BIVARIATE ASSOCIATION WITH METABOLOMIC AGE ACCELERATION	
			β (95% CI)	p value
DEMOGRAPHIC AND NCD RISK FACTORS				
AGE	years	41.24 (9.1)	-	-
SEX	Female	884 (39.5)	-	-
	Male	1354 (60.5)	-0.01 (-0.26, 0.24)	0.94
MARITAL STATUS	Married/cohabiting	1751 (80.2)	-	-
	other	431 (19.8)	0.11 (-0.2, 0.41)	0.5
ETHNICITY	White	2181 (97.5)	-	-
	other	56 (2.5)	-0.37 (-1.14, 0.4)	0.35
BODY MASS INDEX	<25	731 (32.7)	-	-
	>=25 & < 30 (overweight)	1041 (46.5)	0.5 (0.23, 0.77)	0.00035
	>=30 (obese)	466 (20.8)	1.01 (0.67, 1.34)	4.6E-09
DIABETIC STATUS	Normal	2157 (96.4)	-	-
	Diabetic	80 (3.6)	0.78 (0.13, 1.42)	0.019
HYPERTENSION	No	1540 (68.8)	-	-
	Yes	697 (31.2)	0.2 (-0.06, 0.46)	0.13
INCOME	High	879 (39.4)	-	-
	Medium	812 (36.4)	0.26 (-0.01, 0.54)	0.062
	Low	538 (24.1)	0.2 (-0.11, 0.51)	0.2
EDUCATION LEVEL	High	622 (27.8)	-	-
	Medium	710 (31.7)	0.22 (-0.09, 0.53)	0.16
	Low	906 (40.5)	0.12 (-0.18, 0.41)	0.44
ALCOHOL CONSUMPTION	None	174 (7.8)	-	-
	Moderate	1876 (83.9)	0.13 (-0.32, 0.58)	0.57
	Heavy	187 (8.4)	1.02 (0.42, 1.62)	0.00085
SMOKING	Non-smoker	1539 (68.8)	-	-
	Former smoker	477 (21.3)	0.41 (0.11, 0.7)	0.0077
	Current smoker	221 (9.9)	-0.23 (-0.64, 0.18)	0.28
PHYSICAL ACTIVITY	High	1305 (58.3)	-	-
	Moderate	585 (26.1)	-0.21 (-0.5, 0.07)	0.14
	Low	348 (15.5)	0.21 (-0.13, 0.56)	0.23
PSYCHOLOGICAL FACTORS				
DEPRESSION DIAGNOSIS	Normal	1545 (69.1)	-	-
	Minimal symptoms	501 (22.4)	0.45 (0.16, 0.74)	0.0023
	Depression	191 (8.5)	0.91 (0.47, 1.34)	0.000045
ANXIETY DIAGNOSIS	Normal	1733 (79)	-	-
	Borderline	274 (12.5)	0.2 (-0.17, 0.56)	0.3
	Anxiety case	188 (8.6)	0.52 (0.09, 0.96)	0.019

SUBJECTIVE FATIGUE	Low	620 (27.7)	-	-
	Medium	685 (30.6)	0.1 (-0.21, 0.42)	0.52
	High	930 (41.6)	0 (-0.29, 0.3)	0.99
JOB STRAIN	No	1701 (76)	-	-
	Yes	536 (24)	-0.04 (-0.33, 0.24)	0.76
EXPERIENCED TRAUMA IN LAST 6 MONTHS	No	1896 (84.7)	-	-
	Yes, without PTSD	261 (11.7)	0.14 (-0.24, 0.51)	0.47
	Yes, with PTSD	81 (3.6)	0.42 (-0.22, 1.07)	0.2
DIETARY VARIABLES				
FISH CONSUMPTION	None	541 (28.6)	-	-
	Medium	697 (36.8)	0.25 (-0.08, 0.58)	0.14
	High	654 (34.6)	0.51 (0.17, 0.84)	0.0029
FRUIT CONSUMPTION	Low	634 (33.5)	-	-
	Medium	637 (33.7)	0.11 (-0.22, 0.43)	0.51
	High	621 (32.8)	0.11 (-0.21, 0.44)	0.5
RED MEAT CONSUMPTION	Low	580 (30.7)	-	-
	Medium	704 (37.2)	0.21 (-0.11, 0.54)	0.2
	High	608 (32.1)	0.12 (-0.22, 0.45)	0.49
VEGETABLE CONSUMPTION	Low	23 (1.2)	-	-
	Medium	627 (33.1)	0.17 (-0.16, 0.5)	0.31
	High	617 (32.6)	0.08 (-0.24, 0.41)	0.61
WHOLE GRAIN CONSUMPTION	Low	648 (34.2)	-	-
	Medium	604 (31.9)	0.23 (-0.1, 0.56)	0.17
	High	661 (34.9)	0.17 (-0.15, 0.49)	0.3
DASH SCORE	<20 (least healthy)	344 (18.2)	-	-
	≥20 and <23	357 (18.9)	0.54 (0.1, 0.97)	0.015
	≥23 and <25	297 (15.7)	0.18 (-0.28, 0.63)	0.45
	≥25 and <28	396 (20.9)	0.67 (0.24, 1.09)	0.002
	> 28 (most healthy)	498 (26.3)	0.23 (-0.17, 0.64)	0.26
MEDITERRANEAN DIET SCORE	continuous (1-10)	4.73 (1.82)	-0.03 (-0.11, 0.04)	0.35
CLINICAL BIOMARKERS				
SYSTOLIC BLOOD PRESSURE	mmHg	130.85 (15.21)	0.01 (0, 0.01)	0.082
DIASTOLIC BLOOD PRESSURE	mmHg	79.67 (10.1)	0.01 (-0.01, 0.02)	0.28
PULSE	beats/minute	70.37 (11.44)	0 (-0.01, 0.01)	0.76
FIBRINOGEN	g/L	3.87 (0.88)	0.02 (-0.12, 0.16)	0.75
PROTHROMBIN TIME	seconds	13.64 (1.41)	-0.08 (-0.17, 0)	0.058
C-REACTIVE PROTEIN	mg/l	1.92 (3.1)	0.02 (-0.02, 0.06)	0.4
CREATININE	µmol / L	92.35 (12.73)	0.01 (0, 0.02)	0.017
TOTAL CHOLESTEROL	mmol/l	5.25 (1.01)	0.41 (0.29, 0.53)	9.6E-12
HIGH DENSITY LIPOPROTEIN	mmol/L	1.5 (0.39)	0.06 (-0.24, 0.37)	0.68

Γ-GLUTAMYL TRANSFERASE	U/L	31.35 (24.06)	0.01 (0.01, 0.02)	0.000017
APOLIPOPROTEIN A	g/L	1.35 (0.3)	-0.14 (-0.54, 0.25)	0.48
APOLIPOPROTEIN B	g/L	0.93 (0.23)	1.55 (1.04, 2.07)	3.9E-09
% GLYCATED HAEMOGLOBIN	%	5.63 (0.48)	0.32 (0.07, 0.57)	0.013
UREA	µmol / L	5.04 (1.17)	0.08 (-0.02, 0.18)	0.12

Table 2: Significantly enriched metabolic pathways among metabolomic age predictors

PATHWAYS	OVERLAP SIZE	PATHWAY SIZE	P-VALUE
VITAMIN E METABOLISM	15	37	0.00128
TRYPTOPHAN METABOLISM	23	69	0.00165
COA CATABOLISM	4	6	0.00255
UREA CYCLE/AMINO GROUP METABOLISM	17	52	0.0027
LYSINE METABOLISM	10	29	0.00452
CARNITINE SHUTTLE	11	33	0.00476
VITAMIN B5 - COA BIOSYNTHESIS FROM PANTOTHENATE	5	11	0.00492
BIOPTERIN METABOLISM	6	15	0.00563
DRUG METABOLISM - CYTOCHROME P450	15	52	0.00942
TYROSINE METABOLISM	24	91	0.0146
ASPARTATE AND ASPARAGINE METABOLISM	18	71	0.03215

Table 3: Validation of selected metabolomic age predictors in the Northern Finnish Birth Cohort

PLATFORM	FEATURE NAME	MOLECULE	SD CHANGE IN AGE MODEL IN AIRWAVE	MEAN CHANGE FROM 31 TO 46 YEARS IN NFBC (SD)	P VALUE IN NFBC
U_RPOS	1.49_154.0 844m/z	L-Leucine	-0.24	-0.05 (0.19)	2.20E-12
U_RPOS	1.10_147.0 285m/z	Citrate	0.20	0.11 (0.21)	2.20E-12
S_BILISA	IDPL	Phospoholipids in intermediate density lipoproteins	-0.20	0.09 (0.23)	1
S_BILISA	HDTG	Triglycerides in high density lipoproteins	0.20	0.01 (0.36)	0.08
S_BILISA	L3TG	Triglycerides in low density lipoproteins (medium size subclass)	0.12	0.01 (0.46)	2.20E-12
S_BILISA	V6TG	Triglycerides in very low density lipoproteins (smallest size 6 subclass)	-0.37	-0.02 (0.35)	0.004
S_BILISA	V6CH	Cholesterol in very low density lipoproteins (smallest size 6 subclass)	-0.44	-0.03 (0.23)	9.60E-10
S_BILISA	TPA1	Apo-A1 in total plasma	0.31	0.05 (0.15)	2.20E-12
S_BILISA	H1TG	Triglycerides in high density lipoproteins (largest size 1 subclass)	0.20	0.02 (0.46)	5.00E-02

Table 4: Associations between longevity-associated genotypes and age acceleration scores

RS NUMBER	GENE	UK BIOBANK EFFECT (MONTHS LIFE LOST)	GENO-TYPE*	METABOLOMIC AA			DNA METHYLATION AA		
				N (%)	Mean AA (SD)	p value	N (%)	Mean AA (SD)	p value
RS12117661	BSND	5.6	CC	847 (55.9)	0.04 (3)	0.77	587 (55.1)	0.09 (3.56)	0.49
			CG	582 (38.4)	0.02 (2.9)		413 (38.8)	-0.04 (3.16)	
			GG	87 (5.7)	-0.09 (3.02)		65 (6.1)	-0.13 (3.71)	
RS646776	CELSR2	-1.1	CC	72 (4.7)	-0.21 (3.48)	0.91	46 (4.3)	0.13 (3.38)	0.95
			CT	562 (37.1)	0.07 (3.03)		387 (36.3)	0.01 (3.42)	
			TT	882 (58.2)	0.01 (2.87)		632 (59.3)	0.03 (3.43)	
RS2352974	TRAIP	-3.2	CC	373 (24.6)	-0.05 (3.22)	0.53	265 (24.9)	0.09 (3.41)	0.08
			CT	783 (51.6)	0.03 (2.86)		544 (51.1)	-0.28 (3.46)	
			TT	360 (23.7)	0.08 (2.9)		256 (24)	0.61 (3.27)	
RS3936510	C5ORF67	-3.4	GG	977 (64.4)	0.02 (2.79)	0.98	667 (62.6)	-0.06 (3.47)	0.26
			GT	480 (31.7)	0.05 (3.3)		353 (33.1)	0.13 (3.36)	
			TT	59 (3.9)	-0.07 (2.77)		45 (4.2)	0.4 (3.11)	
RS6904450	FTH1PS	-6.9	AA	1033 (68.1)	0 (3.02)	1	713 (66.9)	0.17 (3.41)	0.05
			AT	439 (29)	0.14 (2.81)		324 (30.4)	-0.28 (3.43)	
			TT	44 (2.9)	-0.57 (3.11)		28 (2.6)	-0.27 (3.43)	
RS10455872	LPA	-6.1	AA	1289 (85)	0.02 (3.02)	0.95	909 (85.4)	0.03 (3.47)	0.59
			AG	220 (14.5)	0.14 (2.55)		151 (14.2)	0.05 (3.12)	
			GG	7 (0.5)	-1.98 (3.01)		5 (0.5)	-2.66 (3.03)	
RS1333045	CDKN2BAS	-3.7	TT	354 (23.4)	0 (3.39)	0.64	252 (23.7)	0.08 (3.47)	0.82
			CC	767 (50.6)	-0.01 (2.9)		551 (51.7)	-0.06 (3.44)	
			CC	395 (26.1)	0.1 (2.66)		262 (24.6)	0.15 (3.34)	
RS951266	CHRNA5	-5.7	GG	693 (45.7)	0.05 (2.86)	0.66	493 (46.3)	-0.02 (3.39)	0.31
			GA	657 (43.3)	-0.09 (3.09)		456 (42.8)	-0.05 (3.39)	
			AA	166 (10.9)	0.35 (2.84)		116 (10.9)	0.49 (3.65)	
RS2008514	NPIPB8	-1.1	GG	539 (35.6)	-0.06 (2.68)	0.67	391 (36.7)	0.03 (3.61)	0.86
			GA	711 (46.9)	0.11 (3.1)		497 (46.7)	-0.01 (3.35)	
			AA	266 (17.5)	-0.02 (3.12)		177 (16.6)	0.11 (3.2)	
RS9939973	FTO	-1.1	GG	501 (33)	-0.12 (2.88)	0.052	370 (34.7)	0.09 (3.11)	0.85
			GA	728 (48)	0.01 (2.9)		506 (47.5)	-0.04 (3.47)	
			AA	287 (18.9)	0.32 (3.22)		189 (17.7)	0.08 (3.87)	
RS4420638	APOC1	-4.8	AA	970 (64)	-0.04 (2.88)	0.054	681 (63.9)	-0.14 (3.23)	0.03
			AG	490 (32.3)	0.02 (2.86)		342 (32.1)	0.28 (3.8)	
			GG	56 (3.7)	1.16 (4.57)		42 (3.9)	0.59 (3.05)	

*Ordered by increasing dosage of effect allele

Table 5: Adjusted associations between disease risk factors and age acceleration scores

|		ADJUSTED ASSOCIATIONS WITH METABOLOMIC AGE ACCELERATION		ADJUSTED ASSOCIATIONS WITH DNA METHYLATION AGE ACCELERATION	
		β (95% CI)	p value	β (95% CI)	p value
DIABETIC STATUS	Normal	-	-	-	-
	Diabetic	0.34 (-0.41, 1.1)	0.37	0.54 (-0.75, 1.83)	0.41
HYPERTENSION	No	-	-	-	-
	Yes	-0.05 (-0.36, 0.25)	0.73	-0.12 (-0.66, 0.42)	0.66
INCOME	High	-	-	-	-
	Medium	0.36 (0.06, 0.67)	0.019	-0.01 (-0.55, 0.54)	0.98
	Low	0.35 (0.01, 0.69)	0.042	0.50 (-0.10, 1.10)	0.10
ALCOHOL CONSUMPTION	None	-	-	-	-
	Moderate	0.08 (-0.42, 0.58)	0.75	0.67 (-0.28, 1.61)	0.17
	Heavy	0.91 (0.23, 1.58)	0.008	1.37 (0.09, 2.64)	0.036
BMI	<25	-	-	-	-
	>=25 & < 30 (overweight)	0.57 (0.26, 0.89)	0.00042	0.04 (-0.51, 0.6)	0.88
	>=30 (obese)	0.97 (0.57, 1.37)	2.1E-06	0.11 (-0.61, 0.82)	0.77
SMOKING	Non-smoker	-	-	-	-
	Former smoker	0.29 (-0.04, 0.62)	0.086	-0.13 (-0.69, 0.42)	0.64
	Current smoker	-0.32 (-0.77, 0.13)	0.17	-0.6 (-1.41, 0.2)	0.14
PHYSICAL ACTIVITY	High	-	-	-	-
	Moderate	-0.28 (-0.59, 0.03)	0.074	-0.28 (-0.83, 0.27)	0.32
	Low	0.22 (-0.16, 0.6)	0.26	-0.11 (-0.77, 0.56)	0.75
DEPRESSION DIAGNOSIS	Normal	-	-	-	-
	Minimal symptoms	0.38 (0.05, 0.7)	0.023	0.20 (-0.38, 0.78)	0.5
	Depression	0.74 (0.26, 1.22)	0.0024	0.14 (-0.71, 0.99)	0.75
ANXIETY DIAGNOSIS	Normal	-	-	-	-
	Borderline	0.11 (-0.29, 0.51)	0.59	0.03 (-0.67, 0.74)	0.92
	Anxiety case	0.49 (-0.01, 0.98)	0.055	0.92 (0.03, 1.8)	0.043
EXPERIENCED TRAUMA IN LAST 6 MONTHS	No	-	-	-	-
	Yes, without PTSD	0.29 (-0.11, 0.7)	0.16	-0.22 (-0.9, 0.45)	0.51
	Yes, with PTSD	0.59 (-0.13, 1.31)	0.11	2.15 (0.31, 4)	0.022
SUBJECTIVE FATIGUE	Low	-	-	-	-
	Medium	0.31 (-0.03, 0.66)	0.074	-0.28 (-0.89, 0.33)	0.37
	High	0.11 (-0.21, 0.44)	0.5	-0.08 (-0.66, 0.5)	0.78
JOB STRAIN	No	-	-	-	-
	Yes	0.05 (-0.26, 0.35)	0.77	0.39 (-0.16, 0.94)	0.17

*Models adjusted for sex, ethnicity, study centre, income, hypertension, diabetes, BMI, smoking, alcohol intake, physical activity, DASH score and fish consumption

Figures
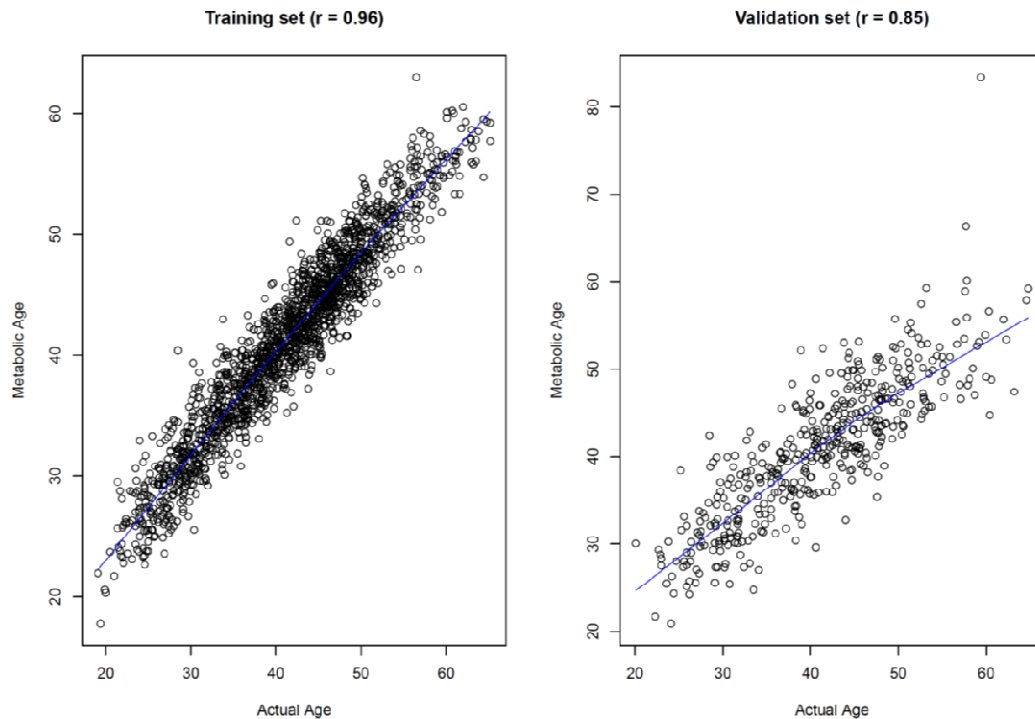
Figure 1: Summary of metabolomic age prediction model. A: Predicted age plotted against chronological age in training set. Pearson's correlation coefficient (r) is shown. B: Predicted age plotted against chronological age in test set. Pearson's correlation coefficient (r) is shown
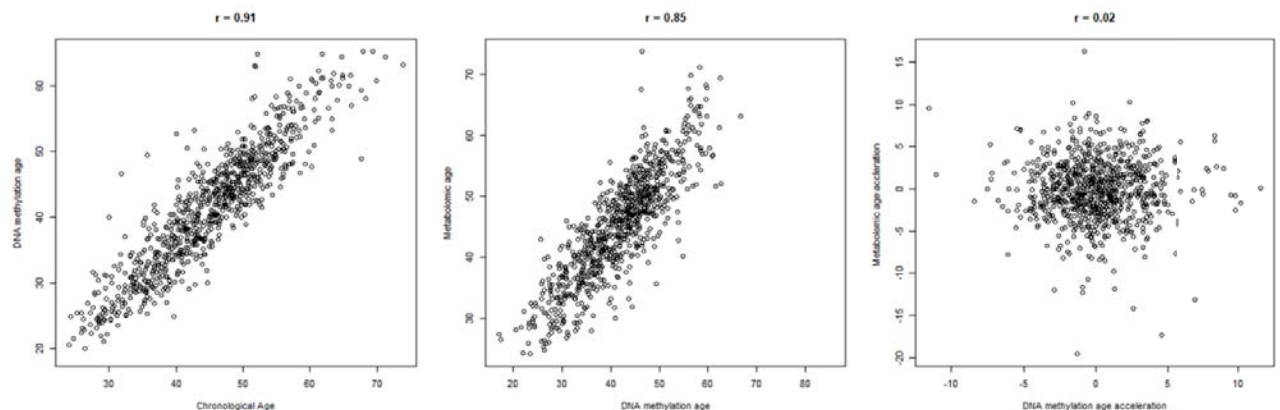
Figure 2: Relationships between different age measures. A. DNA methylation age plotted against chronological age. B: metabolic age plotted against DNA methylation age. C. MetAA plotted against DNAmethAA. Pearson's correlation coefficients (r) are shown