Metabolomics (2025) 21:9
https://doi.org/10.1007/s11306-024-02199-8

ORIGINAL ARTICLE

Metabolomic heterogeneity of ageing with ethnic diversity: a step closer to healthy ageing

Dakshat Trivedi[1,2] · Katherine A. Hollywood[3] · Yun Xu[1] · Fredrick C. W. Wu[4] · Drupad K. Trivedi[3] · Royston Goodacre[1]

Received: 25 June 2024 / Accepted: 10 November 2024
© The Author(s) 2024

Abstract

Introduction Outside of case–control settings, ethnicity specific changes in the human metabolome are understudied especially in community dwelling, ageing men. Characterising serum for age and ethnicity specific features can enable tailored therapeutics research and improve our understanding of the interplay between age, ethnicity, and metabolism in global populations.

Objective A metabolomics approach was adopted to profile serum metabolomes in middle-aged and elderly men of different ethnicities from the Northwest of England, UK.

Methods Serum samples from 572 men of White European (WE), South Asian (SA), and African-Caribbean (AC) ethnicities, ranging between 40 and 86 years were analysed. A combination of liquid chromatography (LC) and gas chromatography (GC) coupled to high-resolution mass spectrometry (MS) was used to generate the metabolomic profiles. Partial Least Squares Discriminant Analysis (PLS-DA) based classification models were built and validated using resampling via bootstrap analysis and permutation testing. Features were putatively annotated using public Human Metabolome Database (HMDB) and Golm Metabolite Database (GMD). Variable Importance in Projection (VIP) scores were used to determine features of interest, after which pathway enrichment analysis was performed.

Results Using profiles from our analysis we classify subjects by their ethnicity with an average correct classification rate (CCR) of 90.53% (LC–MS data) and 85.58% (GC–MS data). Similar classification by age (<60 vs. ≥ 60 years) returned CCRs of 90.20% (LC–MS) and 71.13% (GC–MS). VIP scores driven feature selection revealed important compounds from putatively annotated lipids (subclasses including fatty acids and carboxylic acids, glycerophospholipids, steroids), organic acids, amino acid derivatives as key contributors to the classifications. Pathway enrichment analysis using these features revealed statistically significant perturbations in energy metabolism (TCA cycle), *N*-Glycan and unsaturated fatty acid biosynthesis linked pathways amongst others.

Conclusion We report metabolic differences measured in serum that can be attributed to ethnicity and age in healthy population. These results strongly emphasise the need to consider confounding effects of inherent metabolic variations driven by ethnicity of participants in population-based metabolic profiling studies. Interpretation of energy metabolism, *N*-Glycan and fatty acid biosynthesis should be carefully decoupled from the underlying differences in ethnicity of participants.

Keywords Ethnicity · Healthy ageing · Metabolomics · Mass spectrometry · Serum

✉ Drupad K. Trivedi
drupad.trivedi@manchester.ac.uk

✉ Royston Goodacre
roy.goodacre@liverpool.ac.uk

[1] Centre for Metabolomics Research (CMR), Department of Biochemistry, Cell, and Systems Biology, Institute of Systems Molecular and Integrative Biology, University of Liverpool, Liverpool, UK

[2] Clinical Metabolomics Unit (CMU), Human Development and Health, Institute of Developmental Sciences, University of Southampton, Southampton, UK

[3] Manchester Institute of Biotechnology (MIB), School of Chemistry, University of Manchester, Manchester, UK

[4] Andrology Research Unit (ARU), Division of Endocrinology, Diabetes and Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Published online: 15 December 2024

1 Introduction

The global population is rapidly ageing, and man is expected to live longer than ever before. Countries globally will face increased old age dependency ratios (OADR) with the proportion of 60 years old set to double by 2050 (WHO, 2022). The question: 'What is ageing?'—is yet to be fully answered (Kondoh et al., 2020). Despite substantial advances in understanding the mechanisms of aging, the translation of these insights from research to clinical application has been markedly constrained. This constraint reflects our limited understanding of the inherent variability in aging, especially in the context of the diverse environmental exposures, known as the exposome, that uniquely affect each individual. (Orešič et al., 2020). The heterogeneity impacts at all stages of life right from infancy to old age. For instance, rates of infant and maternal mortality (Knight et al., 2009), incidences of cardiovascular disease (CVD) (Lip et al., 2007), and the risk of type 2-diabetes (T2DM) (Mikhail et al., 2021) are all known to be higher in Black and South Asian populations globally. In comparison White counterparts suffer a higher mortality from cancer (Delon et al., 2022), dementia and Alzheimer incidences (Mukadam et al., 2023). From historical reports of selective immune response to fatal diseases such as Tuberculosis (TB), acquired immune deficiency syndrome (AIDS) (Winkler et al., 2004) and malaria (Arama et al., 2015) to more recent COVID-19 pandemic these disparities have been exemplified across ethnicities globally through evidently disproportionate impact and higher mortality rates (Tai et al., 2022; Yaya et al., 2020). Some studies refer to these as being a consequence of 'community-level deprivation' (Lo et al., 2021) whilst other propose 'randomness' to account for it (Smith, 2011). These disparities could be a result of genetics, environment, or an interaction between the two through epigenetic mechanisms. Independent of the cause, these disparities in health and disease must be investigated and understood to enable our populations to remain healthier, happier, and functional for longer. This becomes pivotal as we transition from the era of omics-based discoveries to precision medicine-led treatments and prognosis (Trivedi et al., 2017).

Ageing and ethnicity are two distinct entities with a common facet of being nonmodifiable and irreversible factors. Moreover, these are impacted by cultural, genetic, environmental, and socio-economic contributors. Although ethnicity can be considered a greater and more significant factor of ageing given its role within a population. Diverse ethnicities can exhibit varying preferences and awareness towards diet and nutrition which can reflect in their metabolome (Clarke et al., 2023) and can be associated to diet related metabolic diseases (Stratakis et al., 2022). Additionally, contributions from non-nutrient components like their choice of medical approach, i.e., choosing traditional, herbal, homeopathic, or naturopathic practices over modern medicine, this too can determine their metabolic makeup. There is evidence for geographical location and migration to play a role in cluster genetic variations within a population (Tishkoff & Kidd, 2004). Thus, understanding ageing through ethnic phenotype could be a way forward. Despite ageing being independent of gender distinctions, our knowledge of hormonal and biochemical processes at the onset of ageing in men is relatively understudied (Gava et al., 2011; Lee et al., 2009) in contrast to that in females (Huddleston et al., 2011). Whilst both ethnicity and age have been studied as factors in such studies, their influence on changes in metabolome in an ageing, multi-ethnic population remains unexplored.

In recent years, metabolomics has emerged as a powerful investigation tool in simplifying complex interplay between genetics, environment, and health. These phenotyping approaches have been used to study mammalian systems and changes occurring at the cellular level due to diseases, drug dosage and other factors like nutrition or environmental perturbations. Metabolomics which does not target predefined sets or lists of metabolites has immense potential when probing a system to generate a hypothesis besides its clinical utility in diagnosis, phenotypic stratification and as a precursor to the development of targeted metabolic therapies. Mass spectrometry-based methods have been tested for phenotyping large populations (Dunn et al., 2011, 2015; Hajjar et al., 2023; Psychogios et al., 2011) given the high-resolution and high throughput capabilities. The findings of these studies form a baseline reference of serum metabolite levels in a relatively healthy population and, to some extent, explore the complex relationships among demographic factors, environmental influences, and disease-associated heterogeneity and metabolomics.

Despite the perceived sample preparation and chromatographic challenges, LC–MS has been the versatile choice over the years for high-resolution data acquisition (Roberts et al., 2022; Trivedi, 2012; Vuckovic 2012). The addition of GC–MS extends the detection coverage to small polar molecules which are useful as these represent central carbon and nitrogen metabolism (McGarrah et al., 2018; Misra, 2021; Mojsak et al., 2021; Zeki et al., 2020). In this study we use a combination of reversed phase LC–MS and GC–MS platforms to generate metabolomic profiles from human serum for comprehensive coverage (Zeki et al., 2020).

2 Materials and methods

2.1 Study population

The study population consisted of middle-aged and older men (mean ± SD age: 59 ± 12 years) of White European, South Asian, and African Caribbean ethnic origins. These men were originally recruited from the Manchester area for the Human Serum Metabolome project (Dunn et al., 2015) and the European Male Ageing Study (EMAS). Briefly, fasting blood samples were collected along with clinical parameters (including age, gender, BMI, smoking status), health questionnaire and anthropometric measures. Details including sample collection and background have been outlined by Lee et al. (2009) and Dunn et al. (2015). From 709 samples available those with missing metadata such as age, Body Mass Index (BMI), comorbidities or with 'missing' or 'other' ethnicities were excluded (see details in Table S1). A subset of 572/709 samples analysed in this study are detailed in Table 1. Individuals were matched for age and BMI. Individuals with more than three co-morbidities reported were excluded to avoid confounded effects due to multimorbidity (Salisbury et al., 2011) in the profiles generated.

2.2 Chemicals and reagents

For *LC–MS:* Optima® LC–MS grade water, methanol, and formic acid (Fisher Scientific) and for *GC–MS:* Optima® LC–MS grade water, anhydrous pyridine, *N*-methyl-*N*-trimethylsilyl tri fluoroacetamide (MSTFA), and methoxamine (Fisher Scientific). For retention index (RI) reagents used were decane, dodecane, pentadecane, nonadecane, docosane and hexane. Internal standard (IS) mix consisted of lysine-d_4, succinic-d_4 acid, glycine-d_5, and benzoic acid-d_5 (Sigma Aldrich). HPLC grade absolute ethanol (99.8%, *v/v*) diluted in water was used for apparatus cleaning and general maintenance.

Table 1 Distribution and demographics of subjects in this study

Parameters	White European (WE)	South Asian (SA)	African Caribbean (AC)
N	287	143	142
Age (years)	63.66 ± 10.68	56.97 ± 11.12	54.01 ± 10.63
BMI (kg/m^2)*	27.47 ± 3.89	27.75 ± 3.70	27.80 ± 4.86
Alcohol intake	74: 141 (0.34)	141:2 (70.5)	128:14 (9.14)
Smoking status	21: 266 (0.07)	22: 120 (0.18)	21: 121 (0.17)
Morbid: healthy**	150: 137 (1.09)	86: 57 (1.5)	60: 82 (0.73)

*BMI and age are expressed as mean ± standard deviation**.** Alcohol intake is expressed as a ratio of drinkers: non-drinkers (72 of the 287 had missing data for alcohol consumption). Smoking status is expressed as a ratio of smokers: non-smokers. **Data were collected on morbidities including high blood pressure, heart conditions, liver conditions, kidney conditions, thyroid disease, prostate disease, diabetes, bronchitis, asthma and incidences of any cancers and stroke. Subjects with 2 or more comorbidities of these were classed as morbid, whilst those with 0 or 1 of these conditions reported were classed as healthy

2.3 Sample preparation

Samples were stored at − 80 °C until preparation. Samples were randomly picked and further re-randomised prior to data acquisition to ensure no systematic biases remained. Samples were passively thawed on ice during preparation. Two × 50 µL aliquots per sample, one for each LC–MS and GC–MS analysis were collected into a 1.5 mL Eppendorf tubes. For GC–MS analysis 50 µL of IS was added to all samples at this stage. Serum was deproteinised by adding 200 µL ice-cold methanol. Samples for LC–MS analyses were vortexed for 10 s followed by centrifugation at 30 °C, 13,500 *g,* for 15 min. The supernatant was lyophilised in a vacuum centrifuge overnight. Lyophilised pellets were reconstituted in 100 µL methanol. Samples were vortexed, transferred to glass vials for LC–MS analysis.

GC–MS derivatisation: Post-deproteinisation samples were derivatised with addition of 25 µL *O*-methoxyamine-HCl in anhydrous pyridine followed by vortex mixing for 10 s and placing onto a heat block at 65 °C for 40 min. Following that, 25 µL of MSTFA was added to each sample, vortex mixed (10 s) and placed on to a heating block at 65 °C for 40 min. 10 µL of 1 mg/mL RI solution (Dunn et al., 2011) was added to each vial before vortex mixing for 10 s and centrifuging at 30 °C, 17,500 *g* for 15 min. 50 µL of supernatant was transferred to a sterile GC glass vial capped with rubber sealed caps.

Quality control: A fixed amount of serum was aliquoted from all samples to prepare a pooled quality control (QC) and system conditioning QCs as described in (Broadhurst et al., 2018). Extraction blanks were prepared alongside samples by substituting serum volume (50 µL) with water. Solvent blanks were also prepared to assess data quality and system performance but not used in statistical analysis. Pooled QCs were used during acquisition and preprocessing steps to assess instrument and data quality.

2.4 Metabolic profiling and data acquisition

LC–MS analysis: Samples were analysed with an Ultimate 3000 UHPLC (Thermo Scientific) coupled to a Q-Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer. Hypersil GOLD C_{18} column (1.9 µm, 2.1 mm × 100 mm), heated and held at 55 °C was setup for chromatographic separation of analytes. 5 µL extracted serum was injected for each sample. Metabolites were separated in samples at a

flow rate of 0.5 mL/min with a controlled 15-min gradient of LC–MS grade mobile phases: (A) H_2O with added (0.1% (*v/v*) formic acid) and (B) methanol (MeOH) with (0.1% (*v/v*) formic acid). The gradient started at 5% (B) for 2 min, increased linearly to 95% (B) over 8 min, and was held for 2 min. It then returned to 5% (B) at 12.1 min, held for 2 min and finally re-equilibrated at 14.1 min to initial conditions.

Data acquisition was conducted in full MS mode in the scan range of 70–1050 m/*z* with a resolution of 70,000 FWHM, an AGC target of 3.10^6 and a maximum injection time of 200 ms. The samples were analysed in positive mode. For MS, the spray voltage was set to 3.5 kV. The capillary temperature was 260 °C, sheath gas 50, aux gas 12.5, max spray current 100 and the probe heater temperature was 425 °C for the analysis. Tuning, calibration, and instrument maintenance was carried out as per the manufacturer's recommendations.

GC–MS analysis: Data were acquired using an Agilent 7890B GC paired with an Agilent 5977B MSD and operated with an Agilent 7693 autosampler. The sample (1 µL) was injected onto a VF-5 ms (inert, 5% phenylmethyl polysiloxane) column (30 m × 250 µm × 0.25 µm; Agilent Technologies) with an inlet temperature of 280 °C and a split ratio of 25:1. Helium was used as the carrier gas with a flow rate of 1.5 mL/min and a pressure of 14.1 psi. The chromatography was programmed to begin at 70 °C with a hold time of 4 min, followed by an increase to 300 °C at a rate of 14 °C/min and a final hold time of 4 min. The total run time per analysis was 14.42 min. The MS was equipped with an electron ion source using 70 eV ionisation and a fixed emission of 35 µA. The mass spectrum was collected for the range of 50–550 *m/z*.

2.5 Data processing and statistical analysis

Data pre-processing: All raw data files were centroided and converted to.*mzXML* using ProteoWizard 3.0 (Chambers et al., 2012). Data were deconvolved and aligned using Progenesis QI for LC, and *eRAH* package using R for GC combined with in-house data pre-processing toolbox [available at BioSpec Github (https://github.com/Biospec/)]. For LC–MS, parameters were setup within data pre-processing method for all 30,265 features detected. A filter was applied such that if a feature was not present in at least 10% of QCs, it was removed, and remainder of features were kept only if relative standard deviation (%) was less than coefficient of variance (%CV) of 40%. This was followed by blank signal intensity ratio filter, which removed 28,507 features. The resulting output had 1758 features. Similar strategy for GC-MS reduced 2074 features to 1253. Pooled QCs were used to algin signal drifts within batches and to perform LOESS (locally estimated scatterplot smoothing) normalisation between batches. The data were $\log_{10}$-transformed, Pareto scaled, and missing values replaced with cubic spline interpolation.

Clustering analysis: Data trends were tested with the unsupervised multivariate method of Principal Component Analysis (PCA) using MetaboAnalyst 4.0 (Chong et al., 2019).

Discriminant analysis and classification modelling*:* To test the classification and prediction ability of the LC–MS and GC–MS profiles generated we employed supervised approach with PLS-DA based classification models for: (i) age with the four intervals 40–49.99, 50–59.00, 60–69.99 and ≥ 70 years of age; (ii) age < 60 and ≥ 60 years; and (iii) the three ethnicities (South Asian, White European and African Caribbean). These models were validated with bootstrap resampling with 1000 iterations. In addition, 1000 null models were also generated where the output class variable was randomly permuted (Gromski et al., 2015). PLS-DA was performed in MATLAB 2019a (MathWorks), and the results are presented as correct classification rates (CCRs) for the 1000 test sets only.

Feature selection and ranking: To identify the significant contributors to the PLS-DA classifications above the variable importance in projection (VIP) scores were calculated using MATLAB 2019a for all three models. The top features from LC–MS analysis were putatively annotated using accurate mass matching to HMDB (https://hmdb.ca). For GC–MS, the features were assigned putative identifications by matching fragment spectra to compound spectra in the GMD (http://gmd.mpimp-golm.mpg.de/). Metabolomics Standards Initiative (MSI) guidelines were adhered to for annotation of all compounds (Sumner et al., 2007) and so LC–MS were at Level 3 at best and GC–MS to MSI Level 2.

Pathway enrichment and functional analysis: Mummichog was employed within MetaboAnalyst 4.0 (https://www.metaboanalyst.ca) (Pang et al., 2022) to assess collective functional impact of significant metabolic features (Li et al., 2013). For mummichog analysis all *m/z* features were ranked by Mann-Whitney (MW) U-test and Kruskal Wallis (KW) ANOVA scores, and corresponding *p*-values. These features were then mapped onto a combination of known human metabolic models, including Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000), Biochemical Genetic and Genomic knowledgebase (BiGG) (King et al., 2016) and the Edinburgh Model (Ma et al., 2007).

Correlation and confounder analysis: To identify associations and guide downstream statistical analysis, a Spearman's correlation test was performed. We aimed to investigate association of age and ethnicity to the metadata available for these subjects. The metadata included anthropometric measures (height, BMI, waist circumference, waist-to-hip and waist to height ratios), clinical measurements (glucose, insulin, homa_ir (homeostatic model assessment of insulin resistance), homa_s (homeostatic model assessment for

insulin sensitivity), shbg (sex hormone binding globulin), cholesterol, LDL (low density lipoprotein) and HDL [high density lipoprotein) cholesterol, triglycerides, HDL (high density lipoprotein) triglycerides, high_bp (blood pressure)] and data on comorbidities and health conditions (bronchitis, asthma, liver condition, kidney condition, thyroid, heart condition, diabetes, prostate condition, cancer and stroke). A correlation matrix was calculated for these in R and visualised as a correlation squares plot (Fig. S8). To evaluate confounding effects of age in ethnicity classification and vice versa we built PLS-DA models that were age-adjusted in ethnicity or ethnicity-adjusted in age. Additionally, models were created using subsets of our subject population i.e., drinkers versus non-drinkers to test for influence of alcohol intake in our classification models. Using the same subsets pathway enrichment analysis was also performed to determine confounding effect of alcohol intake in age specific pathway impact. To determine if there was a confounding effect of ethnicity on citric acid expression in our cohort, box whiskers plots and pairwise *t*-test was performed on normalised citrate intensities of these subjects.

3 Results

The metabolic features which passed data filtering thresholds, QC normalisation, alignment checks and appeared consistently for all samples were deemed reproducible. We retained 1758 such features from LC–MS and 1253 such features from GC–MS for statistical analysis. PCA showed no clear clustering or trends in the data separating these samples by age or ethnicity, however, QCs showed good clustering (Fig. S1). No confounding effects were seen on the age and ethnicity models constructed. No significant correlation was observed between age, ethnicity, and subject metadata from Spearman's correlation tests. Figure S9 displays insignificant difference in citrate level across ethnicities.

3.1 Classification of subjects by *ethnicity*

Samples were divided into three categorical classes based on subject's ethnic origin. Classes consisted of $n = 143$ SAs, $n = 142$ ACs and $n = 287$ WEs. It is known that models generated using unbalanced sample numbers may introduce bias towards a single class i.e., WEs ethnicity outnumbering SAs and ACs approximately 2–1. To avoid this we applied the Synthetic Minority Oversampling Technique (SMOTE) previously reported (Chawla et al., 2002) to our dataset. We built and compared PLS-DA models built with and without SMOTE applied data. Models with SMOTE applied showed marginal but not statistically significant improvement in their averaged correct classification rate (data not shown), indicating that SMOTE did not bias these analyses to the smaller group. A series of PLS-DA models were built and rather than show PLS scores, which in the literature are at best generated from only the training data and do not represent the outputs of the models [in terms of the *Y*-variable that is used as the target output (Westerhuis et al., 2008)], we show our results in terms of the ability for PLS-DA to classify the test set data. Therefore, Fig. 1 represents a histogram of the test set outputs from 1000 bootstrap validated PLS-DA models built to classify subjects by their ethnicity using metabolic features in their serum. The figure shows observed versus null distributions (from 1000 permutation tests) and reports CCRs alongside prediction accuracy, sensitivity, and specificity of these models (the same process was used for Figs. 2 and 3).

3.2 Classification of subjects by *age*

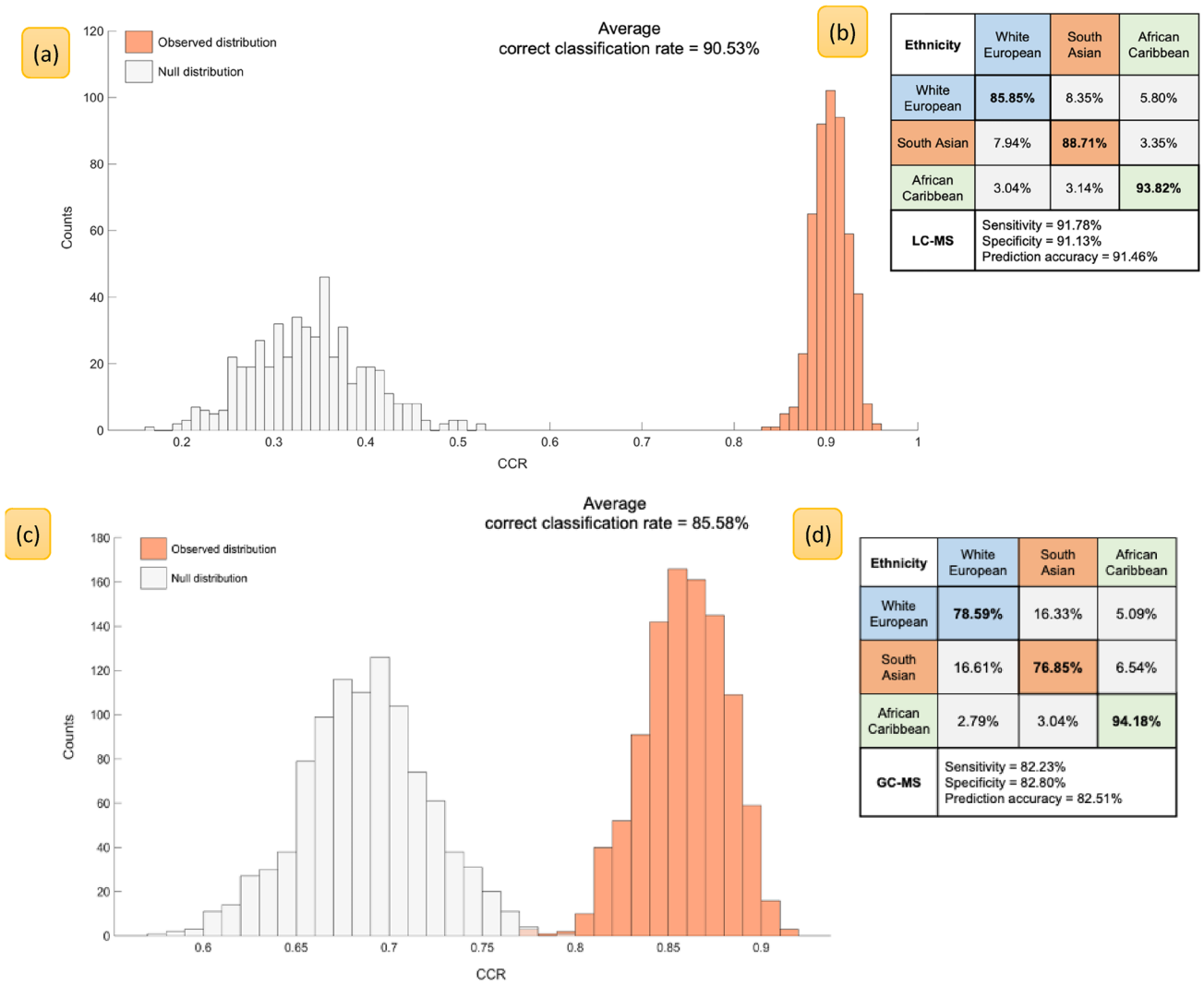
For classification by age, two models were built, each using different class distributions. In model 1, subjects were stratified into four 10-year bands between 40 and 86 years i.e., 40–49.99 years ($n = 145$), 50–59.99 years ($n = 174$), 60–69.99 years ($n = 121$) or ≥ 70 years ($n = 140$). In model 2, the subjects were split into two broader classes: < 60 years ($n = 319$) versus ≥ 60 years ($n = 261$). Figure 2 shows PLS-DA output and permutation testing distributions for model 1. Classifying subjects into tighter 10-year age bands, resulted in decreased CCRs of 73.13% (LC–MS) and 36.08% (GC–MS), indicating lesser metabolic variation in serum profiles. Alternatively, this may be due to the reduced number of samples within four groups compared to two group comparison. Both classification rates and prediction accuracies were seen to improve significantly for model 2 when just two age brackets are used, with almost no overlap between the actual models and the null distributions (Fig. 3).

3.3 Determining confounding effects of age and ethnicity

The models created to test confounding effects of age in ethnicity and vice versa reported average correct classification rates of $< 30\%$ (for both instances). This indicated no obvious confounding effects were present Fig. S2. Additionally, PLS-DA classification for ethnicity showed no significant contribution from 'age' when included as a variable.

3.4 Feature annotation and metabolite identification

Having established a clear and good classification of these men based on their metabolomic profiles, by age and ethnicity, the next stage was to attempt to understand what biochemical processes may be driving these models. We note here that we are aware of the published MSI guidelines



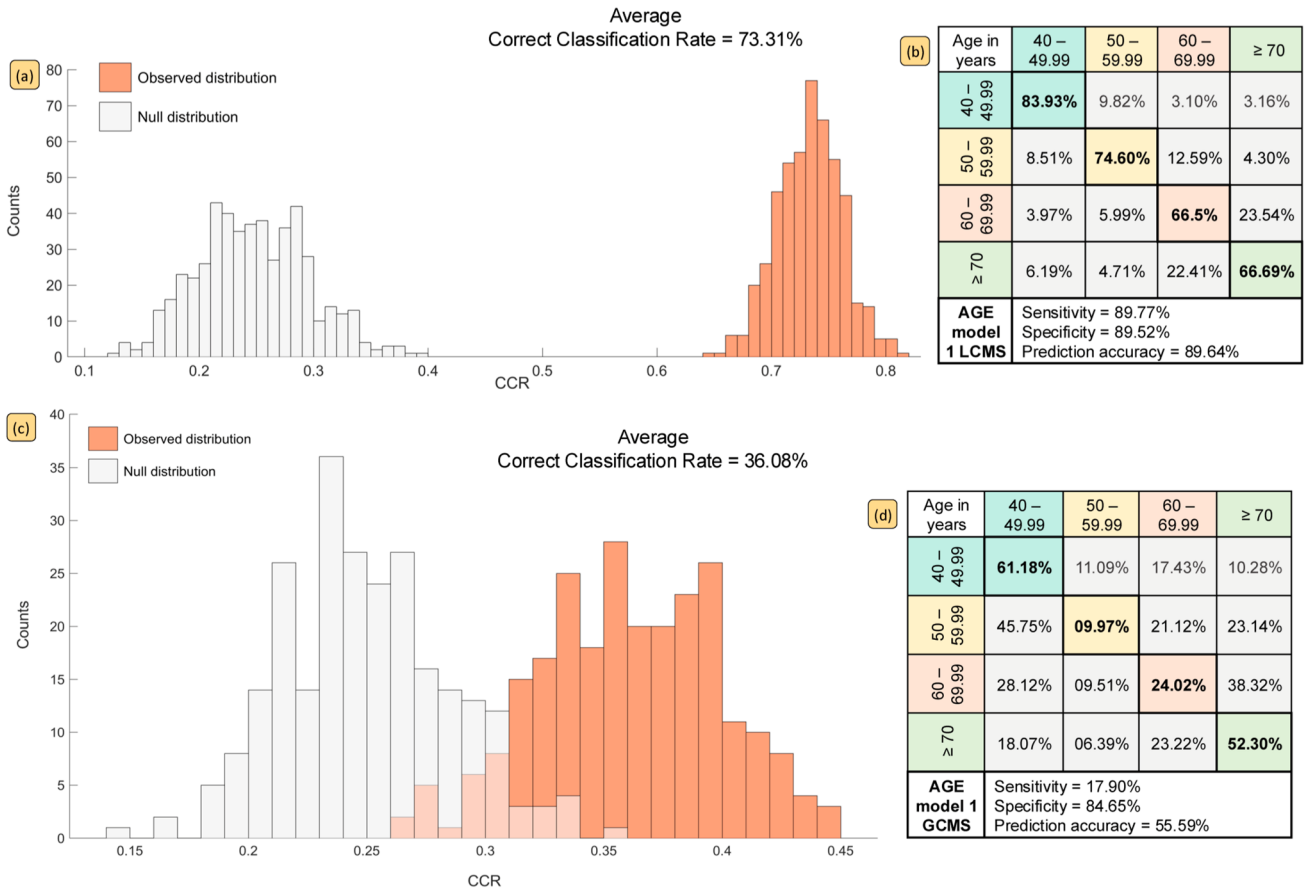
Ethnicity	White European	South Asian	African Caribbean
White European	**85.85%**	8.35%	5.80%
South Asian	7.94%	**88.71%**	3.35%
African Caribbean	3.04%	3.14%	**93.82%**
LC-MS	Sensitivity = 91.78% Specificity = 91.13% Prediction accuracy = 91.46%		

Ethnicity	White European	South Asian	African Caribbean
White European	**78.59%**	16.33%	5.09%
South Asian	16.61%	**76.85%**	6.54%
African Caribbean	2.79%	3.04%	**94.18%**
GC-MS	Sensitivity = 82.23% Specificity = 82.80% Prediction accuracy = 82.51%		

Fig. 1 PLS-DA models built to classify subjects based on their ethnicity (White European vs. South Asian vs. African Caribbean) using metabolic features discovered with LC–MS (**a**, **b**) and GC–MS (**c**, **d**). Histograms **a** and **c** show the null distribution (grey bars) versus observed distribution (orange bars) the averaged correct classification and graphics **b** and **d** report class-wise classification rates, in addition to model's overall sensitivity, specificity and prediction accuracy for each bootstrapped PLS-DA model

(Sumner et al., 2007), where two orthogonal properties are used for annotation and identification (if a standard is available). For our analyses the two different physicochemical properties are retention time—relating to polarity in LC–MS and volatility in GC–MS—and mass spectrometry information, where for LC–MS this is solely accurate mass (due to the high mass resolution mass analyser used) and for GC–MS fragmentation patterns due to high energy (viz. 70 eV) electron ionization. Thus, for GC–MS our identification against external libraries (e.g., GMD) is at best MSI Level 2, and for LC–MS as we are using accurate mass only, this is less confident. In addition, we did fact check our annotations to make sure they are not spurious following suggestions published recently (Theodoridis et al., 2023). If an identification of analyte had low biological plausibility (i.e. a plant specific metabolite seen in serum) or low plausibility in terms of chromatographic retention (i.e. a lipid peak was seen within first two minutes of a gradient analysis using C_{18} column), we annotated the feature as 'unknown/unidentified'.

To select discriminatory variables, we used input variables from PLS-DA models with VIP score > 1 from both ethnicity and the two age PLS-DA models. Features were ranked by descending VIP scores. The top 10% variables from each model were annotated with *m/z* matching to HMDB (LC–MS) and accurate mass and spectral fragmentation match to GMD database (GC–MS). In PLS-DA modelling a clear ranking-based feature selection may not always be possible when too many variables contribute equally towards classification of the response variable, therefore data



(b) Age in years	40 – 49.99	50 – 59.99	60 – 69.99	≥ 70
40 – 49.99	**83.93%**	9.82%	3.10%	3.16%
50 – 59.99	8.51%	**74.60%**	12.59%	4.30%
60 – 69.99	3.97%	5.99%	**66.5%**	23.54%
≥ 70	6.19%	4.71%	22.41%	**66.69%**
AGE model 1 LCMS	Sensitivity = 89.77% Specificity = 89.52% Prediction accuracy = 89.64%			

(d) Age in years	40 – 49.99	50 – 59.99	60 – 69.99	≥ 70
40 – 49.99	**61.18%**	11.09%	17.43%	10.28%
50 – 59.99	45.75%	**09.97%**	21.12%	23.14%
60 – 69.99	28.12%	09.51%	**24.02%**	38.32%
≥ 70	18.07%	06.39%	23.22%	**52.30%**
AGE model 1 GCMS	Sensitivity = 17.90% Specificity = 84.65% Prediction accuracy = 55.59%			

Fig. 2 PLS-DA models built to classify subjects based on age in four brackets (viz. 40–49.99 years, 50–59.99 years, 60–69.99 years, or ≥ 70 years) using metabolic features discovered with LC–MS (**a**, **b**) and GC–MS (**c**, **d**). Histograms **a** and **c** show the null distribution (grey bars) versus observed distribution (orange bars) the averaged correct classification and graphics **b** and **d** report class-wise classification rates, in addition to overall sensitivity, specificity and prediction accuracy for each bootstrapped PLS-DA model

specific cut-off should be calculated for accurate ranking of features (Akarachantachote et al., 2014). We chose the top 10% annotated variables (Tables S2–S5) ranked by their VIP scores for visualisation (Figs. S3–S7).

3.5 Pathway enrichment analysis

Pathway enrichment analysis was performed using the most important *m*/*z* features, determined by their VIP scores being higher than 1 from PLS-DA models used to predict age and ethnicity. Features selection returned 1574 features for LC–MS and 561 features for GC-MS with VIP scores > 1. Figure 4 shows the output in form of a bubble plot with each pathway represented with a circular node. For simpler interpretation, each filled circle corresponds to a metabolic pathway or a biological process, its size represents importance of that pathway in the context of analysis and its colour profile (yellow to red) infer the statistical significance (low to high) of these computed enrichments. This allowed us to prioritise selection of relevant enriched pathway for further interpretation, with the published literature. Typically, a large and red circle tends to be a pathway of high priority given its enrichment and significance. Pathways with significant hits have been highlighted. Note, not all the named pathways had statistically significant *p*-values, i.e., *p*-value below the set threshold value of 0.05 post Benjamini–Hochberg false discovery rate (BH-FDR) correction. Only those truly significant (i.e., FDR adjusted *p*-value ≤ 0.05) are discussed below.

4 Discussion

Literature reviews reveal that the study of ethnic difference in men's metabolic profiles has to date received very little attention. As noted by Vasishta et al. (2022), many metabolomics studies examining the metabolome from an ethnic perspective yield inconclusive result. This lack of clarity often stems from factors such as insufficient

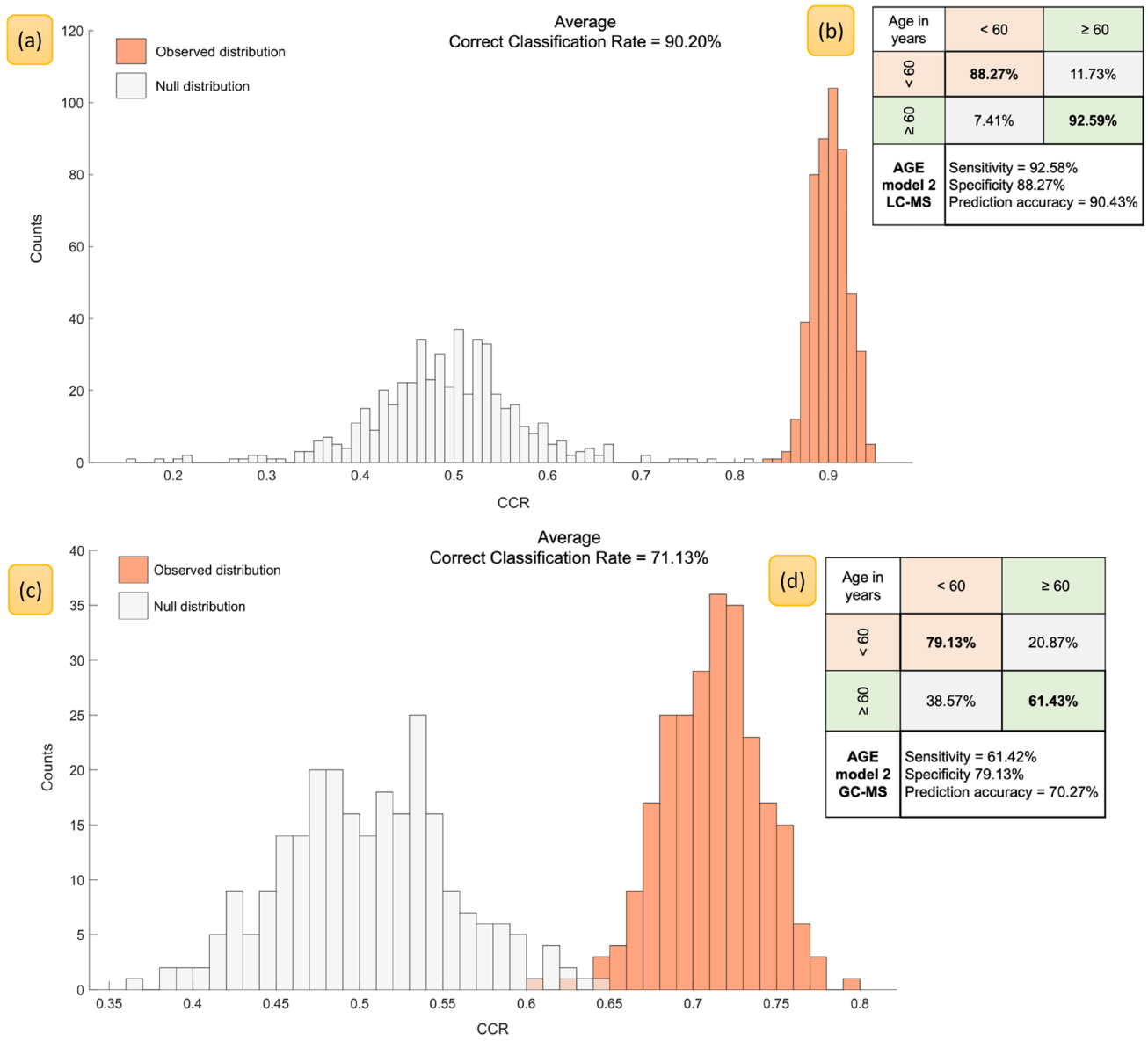


Fig. 3 PLS-DA models built to classify subjects based on two classes of age (viz. < 60 years vs. ≥ 60 years old) using metabolic features discovered with LC–MS (**a**, **b**) and GC–MS (**c**, **d**). Histograms **a** and **c** show the null distribution (grey bars) versus observed distribution (orange bars) the averaged correct classification and graphics **b** and **d** report class-wise classification rates, in addition to overall sensitivity, specificity and prediction accuracy for each bootstrapped PLS-DA model

sample sizes, issues with metadata (e.g., gender, BMI, waist-to-hip ratio) either being mismatched or self-reported, or the constraints of limited case–control study designs (Hu et al., 2022). Our study follows recommendations from the community (Trivedi et al., 2017) and had 572 well-matched, fasted serum samples from community dwelling subjects with at least a 113 samples in each class for any single statistical analysis performed. These results are therefore a good representation of the ethnic phenotypes in the UK (White European, South Asian and African Caribbean) that have been characterised by metabolomics approaches. We were able to obtain metabolomic profiles of these men using LC–MS and GC–MS and could classify accurately these men using their metabolic signatures into ethnic phenotypes and age groups. To achieve this, we leveraged PLS-DA and validated this supervised learning method using bootstrapping and permutation testing. Finally, in this process the key contributing input variables (metabolites) were ranked and used for enriched metabolic pathway analysis using MetaboAnalyst. Below are discussions of the main pathways that

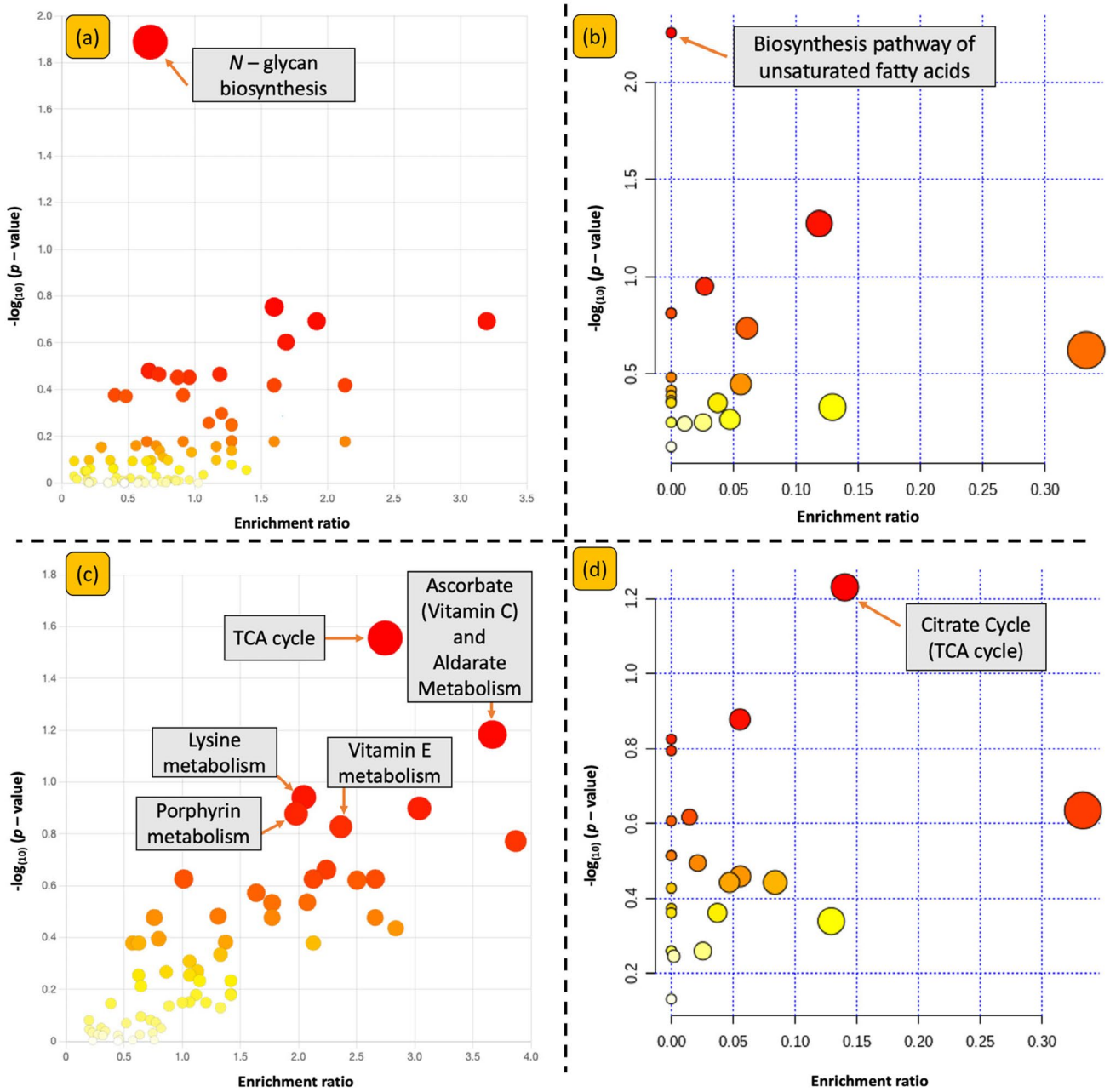


Fig. 4 Pathway enrichment analysis results from MetaboAnalyst 4.0. This functional analysis was performed using all *m/z* features, their corresponding MW *U*-test (age) and KW-ANOVA scores (ethnicity), and *p*-values. Subfigures (**a**) and (**b**) represent ethnicity-specific pathways derived from LC-MS and GC-MS data, respectively. Subfigures (**c**) and (**d**) display age-specific pathways from LC-MS and GC-MS data, respectively. The y-axis represents the $-\log_{(10)}$ *p*-values from pathway enrichment analysis, with *p*-value threshold set at 0.05. The x-axis shows pathway enrichment expressed as the ratio of observed hits to expected hits on a pathway, where expected hits are based on chance alone

we propose are biochemically distinct between ethnic and age phenotypes.

4.1 *N*-Glycan biosynthesis in ethnicity

With Fisher's exact test (FET) *p*-value of 0.02, *N*-glycan biosynthesis was the only highly significant impacted pathway by ethnicity specific discriminators (Fig. 4a). Gebrehiwot et al. (2018) defined the inter-ethnic variations of the

N-glycome in serum in healthy individuals of US origin including South Indian, Japanese, and Ethiopian populations. These authors demonstrated an ethnicity-specific *N*-glycan expression pattern with 13 out of 51 glycans exclusively detected in Ethiopians. Immunoglobulin G (IgG), the most abundant circulating immunoglobulin, is heavily glycosylated, the alteration of which imposes a dysregulated immune-inflammatory response leading to increased risk of cardiovascular disease development. IgG-associated *N*-glycan traits are differentially associated with the development of hypertension in a population-pooled analysis of 4757 participants of different ethnicities (Chinese Han, Croatian Korcula, Croatian Vis and Scottish Orkney) (Wang et al., 2016). In a similar study, at least 10 IgG *N*-glycan traits associated with incident hypertension in ethnic minorities of China (namely, Uygur, Kazak, Kirgiz, and Tajik) (Liu et al., 2018). Although previous studies have not established a clear link between stratified ethnicity and IgG *N*-glycosylation, *N*-glycan traits show differential correlations with ethnic groups. This information can be further investigated in future population-based studies across both genders to gain a clearer understanding of how *N*-glycans vary across ethnicities and whether these variations are significant.

4.2 Biosynthesis of unsaturated fatty acids and ethnicity

Ethnicity specific metabolic conversion of fatty acids such as linoleic acid (LA) and arachidonic acid (AA) have been reported previously (Abdelmagid et al., 2015). Recently, a study performed in a US based cohort identified 259 metabolites including γ-LA, an intermediate between LA and AA in the conversion chain to be differentiating Black and White women based on their ethnicity. The study linked these metabolites to the increased susceptibility of coronary heart disease observed in these women (Hu et al., 2022). Such changes can be attributed to ethnicity specific genetic variations for enzymes involved in long-chain fatty acid synthesis, such as delta 5-desaturase (D5D) and delta 6-desaturase. Polymorphisms of genes encoding these enzymes have been reported to alter in Caucasian and Asian adults, resulting in differential plasma fatty acid profiles (Merino et al., 2011). Omega-3 polyunsaturated fatty acids suggest that ethnic diversity affects cardiovascular outcomes through variations in lipid profiles and glucose homeostasis (Patel et al., 2010). Ethnic differences are also associated with distinct inflammatory phenotypes; for instance, several studies indicate that Black individuals have higher levels of inflammatory biomarkers compared to Hispanics, Whites, or Asians (Abdelmagid et al., 2015; Carroll et al., 2009; Stowe et al., 2010). Anti-inflammatory long-chain polyunsaturated fatty acids (LC-PUFAs) are often prescribed to mitigate subclinical inflammation associated with aging. A question remains as to whether dysregulation or heterogenicity of fatty acid metabolism biologically contributes to ethnicity-based differences in health outcomes, considering the age of the cohort. Whether this contribution is independent of other factors should also be investigated. However, to definitively address these questions, it is necessary to investigate clinical measurements of known disease precursors and hormonal data. Previous research (Eendebak et al., 2017) has shown that these factors have differential impacts by ethnicity in men from the same cohort as ours.

Despite being classified as an unmodifiable risk factor for poor cardiovascular outcomes, ethnicity is increasingly recognised to have a significant impact on patient outcomes. For instance, African-Caribbeans, Hispanics, and other ethnic minorities have a substantially increased prevalence of hypertension (HTN) and type 2 diabetes mellitus, which automatically imposes a higher baseline threshold for developing end-organ damage, including heart failure (HF) (Kim et al., 2010). In this cohort Spearman's correlation of the clinical and biochemical metadata (Fig. S8) revealed no strong correlation of age, clinical characteristics, morbidity, and physiology with ethnicity. There was a moderately strong negative correlation between ethnicity and age ($r=-0.47$). There was a weak negative correlation observed between ethnicity and waist-to-hip ratio ($r=-0.31$) and shbg ($r=-0.15$). Age showed moderate positive correlation to shbg ($r=0.36$), high blood pressure ($r=0.35$) and stroke ($r=0.22$) all of which are expected to vary with age. As such, in absence of any strong correlation there is very little indication of any significant confounding effect from comorbidities and lifestyle present in this cohort. It should be noted that association with age related ailments such as high blood pressure and heart condition or stroke is expected in a cohort consisting of middle aged and older men. An essential takeaway for epidemiological metabolomics analysts would be the strengthened requirement to account for confounding effects of both age and ethnicity in future human metabolic profiling studies using serum.

4.3 Citrate (TCA) cycle and ageing

In the present study citrate, as an indicator of the TCA cycle, was significantly impacted (FET p-value = 0.04) by age specific classifiers from our discriminant analysis. This finding that citric acid was upregulated in older age groups corroborates findings from a large-scale UK based phenotyping study in serum by Dunn et al. (2015). Moreover, Saito et al. (2016) also claim energy metabolites of citrate cycle to be significantly different between young and old men. Whilst the TCA cycle has been linked to ageing and age-associated diseases (Chin et al., 2014) it has not been co-studied for influence from ethnicity of individuals. To determine if there was a confounding effect of ethnicity on citric acid

expression in our cohort, box whiskers plots and pairwise *t*-test was performed which showed insignificant difference in citrate level across ethnicities (Fig. S9). Extracellular citrate accumulation in the senescent cells is considered to be a marker for chronological age independent of sex, BMI or telomere length (Mycielska et al., 2022) and abnormalities in the citrate cycle have been associated to declined intrinsic capacity in older individuals (Pan et al., 2022). In rodents (Chen et al., 2018), and in humans, citrate stored in bones is released into extracellular matrix in osteoporosis, a disease more prevalent in the later age.

The citrate cycle occurs within the mitochondria and involves a series of chemical reactions resulting in the production of diverse metabolites mediating several cellular functions, including epigenetic modifications, post-translational modifications, and diverse biosynthetic pathways (Martínez-Reyes & Chandel, 2020). There is growing evidence that links TCA metabolic dysregulation with ageing and age-associated diseases. A UK-based phenotyping study in a large populations ($n = 1200$) found citric acid is not gender-specific and increased with age in both males and females (Dunn et al., 2015) where ethnicity was not accounted for.

4.4 Porphyrin metabolism and other age-related pathways

The impact on porphyrin metabolism was noticeable with respect to age. There is not much information available to link porphyrin metabolism to age. Reports of alcohol consumption causing disturbances to porphyrin metabolism have been published (Doss et al., 2000). Alcohol is considered a porphyrinogenic agent and a cause for hepatic coproporphyrinuria later manifested as alcohol-induced liver damage. Among our subjects 13/572 report chronic liver-related disease or dysfunction. It is unlikely for this enrichment to have been contributed from these 13 men; this association between age and porphyrin metabolism should be further investigated. To investigate whether alcohol intake in this population influenced ageing in this population, we performed mummichog analysis on two subsets of our population i.e., drinkers and non-drinkers. Porphyrin metabolism pathway remained impacted (Fig. S10a, b) in both cases suggesting the alcohol intake in these subjects did not play a role in its enrichment.

Anthranilic acid, a metabolite produced in the L-tryptophan-kynurenine pathway, was observed to increase with age (Fig. S6). Previous studies report L-tryptophan-kynurenine pathway associated with ageing and age-related diseases. Dunn et al. (2015) reported decreasing tryptophan with male ageing which can be explained by increased muscle mass turnover. An increase in kynurenine causes tryptophan to decrease and this has been linked with age-specific cognitive decline (van der Goot & Nollen, 2013). Kynurenine is a downstream metabolite of tryptophan metabolism and a substrate for formation of anthranilic acid (Ramos-Chávez et al., 2018). Therefore, decrease in tryptophan can explain increasing anthranilic acid with progressing age. A recent study linked metabolic biomarkers of decreased intrinsic capacity in elderly to tryptophan metabolism among other key pathways reported (Pan et al., 2022).

We note that some of the metabolites (Figs. S6 and S7) may not be of endogenous origin yet differentially expressed. Metabolites like gallic acid (Kahkeshani et al., 2019) may be from anti-inflammatory, neuropsychological and metabolic supplements, caffeic acid from coffee intake (Bastianini et al., 2018) whereas potential metabolites from use of cosmetic and anti-ageing solutions such as L-rhamnose (Pageon et al., 2019) were also detected. We found kaempferol—a polyphenol administered as supportive treatment for diseases in some traditional medicinal approaches (Ren et al., 2019). This could be indicative of the cultural diversity within this population and their self-care habits. Detection of such metabolites is expected alongside endogenous metabolites in a global profiling experiment performed using highly sensitive, chemically specific MS-platforms.

4.5 Strengths, limitations, and outlook

There are some limitations to this study. Moreover, the same ethnic groups in different countries may have different serum metabolic profiles. These differences in serum metabolomes could be attributed to many external factors such as diet, environment, socio economic factors and access to healthcare in early age (Goodacre, 2007). Additional metadata including dietary intake, medication, significant lifestyle factors and any exposome influencers would enhance findings in such population-based metabolomics studies. Ethnicity in this population was determined using ethnic origins of one or both parent/s going back up to four generations. As robust as this method is, it is not a universal method to determine or record ethnicity in populations worldwide. Studies relying on any self-reported ethnicity data for e.g., Navarro et al. (2023) or ethnicity otherwise estimated or recorded, may not always be accurate or reliable. Therefore, the phenotype associated to such data may not be confidently defined; that is to say, there may not be a strong difference for these different groups. That said this is a large study where relatively large populations of individuals were measured: 287 White Europeans, 143 South Asians and 142 African Caribbeans.

Moreover, we address the inter-individual heterogeneity within our cohort, with respect to demographic data, different clinical parameters, and comorbidities. The study is not intended to identify metabolic biomarkers of certain disease entities, but rather to illustrate the global association of metabolic profiles with aging and ethnicity. We did not

therefore attempt to investigate whether ethnicity is directly or independently related to the identified metabolites or metabolic pathways.

Our study highlights the potential of metabolomics in routine clinical decision making, where currently despite a suite of administrable therapies available in hospitals for some cases, the decision-making is often a struggle due to poorly understood patient response owing to population heterogeneity (Caiazzo et al., 2022). Koeken et al. (2022) sets the premise to show how metabolomic profiles can be used to predict immune responses which in turn can enable delivery of effective and tailored vaccines and therapies in populations. We encourage further work in this area with added analytical platforms, inclusion of other genders, wider ethnic diversity, larger sample size, additional metadata, and a multi-center validation mechanism to accelerate research in healthy ageing to alleviate socio-economic burdens. This could pave way for personalised biomarkers in future geared towards predictive, diagnostic, and therapeutic interventions needed for inclusive and healthy ageing.

5 Conclusion

This work is our contribution towards the UN's 'decade of healthy ageing' declaration. With the novel ethnicity specific metabolome changes reported here, we are a step closer to understanding differential mechanisms of ageing across these ethnic groups. We report that men of White European, South Asian, and African Caribbean origin have distinct metabolic signatures in serum with prediction potential of their ethnicity and age group. We report changes in energy and fatty acid metabolism driven by age and ethnicity of these men. This approach will need to be adapted to a wider range of ethnicities both within the UK and globally. Future work should include other genders, wider ethnic diversity, larger cohorts, longitudinal data, and multi-centre validation to accelerate research in healthy ageing.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11306-024-02199-8.

Acknowledgements R.G. is grateful to UK Medical Research Council for funding the fRaill Project (190 MRC G1001375/1) and MAP/UK (MR/S010483/1). DKT is thankful to Community of Analytical and Measurement Sciences (CAMS) for supporting DKT's research position. We are grateful to Prof. Fred Wu and his team for EMAS samples and metadata. We thank Prof. Perdita Barran for facilitating access to the analytical platforms available in Michael Barber Centre for Collaborative Mass Spectrometry within Manchester Institute of Biotechnology.

Author contributions R.G. and D.T. conceived research and designed study. D.T, K.A.H. and D.K.T. acquired data on LC–MS and GC–MS platforms. D.T. processed and analysed data with advice from D.K.T. Y.X. and R.G. provided in-house tools for multivariate analysis. D.T. wrote the manuscript. R.G. provided the HUSERMET (SA and AC) samples with metadata from (Manchester) repository. FW provided the EMAS (WE) samples with metadata. All authors read and approved the final version of the manuscript.

Data availability All data files are available in *.mzXML* format at MetaboLights repository (Yurekten et al., 2024) at https://www.ebi.ac.uk/metabolights/editor/www.ebi.ac.uk/metabolights/MTBLS10303. The codes used for spectral pre-processing, auto-correction, normalisation, PCA, and PLS-DA are available at https://github.com/BioSpec/cluster-toolbox-v2.0.

Declarations

Conflict of interest All authors declare no conflict of interest.

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

References

Abdelmagid, S. A., Clarke, S. E., Roke, K., & Ma, D. W. L. (2015). Ethnicity, sex, FADS genetic variation, and hormonal contraceptive use influence delta-5- and delta-6-desaturase indices and plasma docosahexaenoic acid concentration in young Canadian adults: A cross-sectional study. *Nutrition and Metabolism, 12*(1), 14.

Akarachantachote, N., Chadcham, S., & Saithanu, K. (2014). Cutoff threshold of variable importance in projection for variable selection. *International Journal of Pure and Applied Mathematics, 94*(3), 307–322.

Arama, C., Maiga, B., Dolo, A., Kouriba, B., Traoré, B., Crompton, P. D., et al. (2015). Ethnic differences in susceptibility to malaria: What have we learned from immuno-epidemiological studies in West Africa? *Acta Tropica, 146*, 152–156. https://doi.org/10.1016/j.actatropica.2015.03.023

Bastianini, M., Faffa, C., Sisani, M., & Petracci, A. (2018). Caffeic acid-layered double hydroxide hybrid: A new raw material for cosmetic applications. *Cosmetics, 5*(3), 1–11.

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics: Official Journal of the Metabolomic Society, 14*(6), 72. https://doi.org/10.1007/s11306-018-1367-3

Caiazzo, E., Rezig, A. O. M., Bruzzese, D., Ialenti, A., Cicala, C., Cleland, J. G. F., et al. (2022). Systemic administration of glucocorticoids, cardiovascular complications and mortality in patients hospitalised with COVID-19, SARS, MERS or influenza: A systematic review and meta-analysis of randomised trials. *Pharmacological Research, 176*, 106053. https://doi.org/10.1016/j.phrs.2021.106053

Carroll, J. F., Fulda, K. G., Chiapa, A. L., & Cardarelli, R. (2009). Impact of race/ethnicity on the relationship between visceral fat and inflammatory biomarkers. *Obesity, 17*(7), 1420–1427.

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology, 30*(10), 918–920. https://doi.org/10.1038/nbt.2377

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*(1), 321–357.

Chen, H., Wang, Y., Dai, H., Tian, X., Cui, Z.-K., Chen, Z., et al. (2018). Bone and plasma citrate is reduced in osteoporosis. *Bone, 114*, 189–197. https://doi.org/10.1016/j.bone.2018.06.014

Chin, R. M., Fu, X., Pai, M. Y., & Huang, J. (2014). The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. *Nature, 510*(7505), 397–401.

Chong, J., Wishart, D. S., & Xia, J. (2019). Using metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current Protocols in Bioinformatics, 68*(1), 1.

Clarke, E. D., Ferguson, J. J. A., Stanford, J., & Collins, C. E. (2023). Dietary assessment and metabolomic methodologies in human feeding studies: A scoping review. *Advances in Nutrition, 14*(6), 1453–1465. https://doi.org/10.1016/j.advnut.2023.08.010

Delon, C., Brown, K. F., Payne, N. W. S., Kotrotsios, Y., Vernon, S., & Shelton, J. (2022). Differences in cancer incidence by broad ethnic group in England, 2013–2017. *British Journal of Cancer, 126*(12), 1765–1773. https://doi.org/10.1038/s41416-022-01718-5

Doss, M. O., Kühnel, A., & Gross, U. (2000). Alcohol and porphyrin metabolism. *Alcohol and Alcoholism, 35*(2), 109–125. https://doi.org/10.1093/alcalc/35.2.109

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*(7), 1060–1083. https://doi.org/10.1038/nprot.2011.335

Dunn, W. B., Lin, W., Broadhurst, D., Begley, P., Brown, M., Zelena, E., et al. (2015). Molecular phenotyping of a UK population: Defining the human serum metabolome. *Metabolomics: Official Journal of the Metabolomic Society, 11*, 9–26. https://doi.org/10.1007/s11306-014-0707-1

Eendebak, R. J. A. H., Swiecicka, A., Gromski, P. S., Pye, S. R., O'Neill, T. W., Marshall, A., et al. (2017). Ethnic differences in male reproductive hormones and relationships with adiposity and insulin resistance in older men. *Clinical Endocrinology, 86*(5), 660–668. https://doi.org/10.1111/cen.13305

Gava, A. L., Freitas, F. P. S., Meyrelles, S. S., Silva, I. V., & Graceli, J. B. (2011). Gender-dependent effects of aging on the kidney. *Brazilian Journal of Medical and Biological Research, 44*(9), 905–913.

Gebrehiwot, A. G., Melka, D. S., Kassaye, Y. M., Rehan, I. F., Rangappa, S., Hinou, H., et al. (2018). Healthy human serum N-glycan profiling reveals the influence of ethnic variation on the identified cancer-relevant glycan biomarkers. *PloS One, 13*(12), e0209515. https://doi.org/10.1371/journal.pone.0209515

Goodacre, R. (2007). Metabolomics of a superorganism. *The Journal of Nutrition, 137*(1 Suppl), 259S-266S. https://doi.org/10.1093/jn/137.1.259S

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Analytica Chimica Acta, 879*, 10–23. https://doi.org/10.1016/j.aca.2015.02.012

Hajjar, G., Barros Santos, M. C., Bertrand-Michel, J., Canlet, C., Castelli, F., Creusot, N., et al. (2023). Scaling-up metabolomics: Current state and perspectives. *TrAC Trends in Analytical Chemistry, 167*, 117225. https://doi.org/10.1016/j.trac.2023.117225

Hu, J., Yao, J., Deng, S., Balasubramanian, R., Jiménez, M. C., Li, J., et al. (2022). Differences in metabolomic profiles between black and white women and risk of coronary heart disease: An observational study of women from four US cohorts. *Circulation Research, 131*(7), 601–615. https://doi.org/10.1161/CIRCRESAHA.121.320134

Huddleston, H. G., Rosen, M. P., Gibson, M., Cedars, M. I., & Fujimoto, V. Y. (2011). Ethnic variation in estradiol metabolism in reproductive age Asian and white women treated with transdermal estradiol. *Fertility and Sterility, 96*(3), 797–799.

Kahkeshani, N., Farzaei, F., Fotouhi, M., Alavi, S. S., Bahramsoltani, R., Naseri, R., et al. (2019). Pharmacological effects of gallic acid in health and diseases: A mechanistic review. *Iranian Journal of Basic Medical Sciences, 22*(3), 225–237. https://doi.org/10.22038/ijbms.2019.32806.7897

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research, 28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kim, C. X., Bailey, K. R., Klee, G. G., Ellington, A. A., Liu, G., Mosley, T. H., et al. (2010). Sex and ethnic differences in 47 candidate proteomic markers of cardiovascular disease: The Mayo Clinic proteomic markers of arteriosclerosis study. *PLoS ONE, 5*(2), e9065. https://doi.org/10.1371/journal.pone.0009065

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research, 44*(D1), D515–D522. https://doi.org/10.1093/nar/gkv1049

Knight, M., Kurinczuk, J. J., Spark, P., & Brocklehurst, P. (2009). Inequalities in maternal health: National cohort study of ethnic variation in severe maternal morbidities. *BMJ, 338*, b542. https://doi.org/10.1136/bmj.b542

Koeken, V. A. C. M., Qi, C., Mourits, V. P., de Bree, L. C. J., Moorlag, S. J. C. F. M., Sonawane, V., et al. (2022). Plasma metabolome predicts trained immunity responses after antituberculosis BCG vaccination. *PLoS Biology, 20*(9), e3001765. https://doi.org/10.1371/journal.pbio.3001765

Kondoh, H., Kameda, M., & Yanagida, M. (2020). Whole blood metabolomics in aging research. *International Journal of Molecular Sciences, 22*(1), 175. https://doi.org/10.3390/ijms22010175

Lee, D. M., O'Neill, T. W., Pye, S. R., Silman, A. J., Finn, J. D., Pendleton, N., et al. (2009). The European Male Ageing Study (EMAS): Design, methods and recruitment. *International Journal of Andrology, 32*(1), 11–24. https://doi.org/10.1111/j.1365-2605.2008.00879.x

Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., et al. (2013). Predicting network activity from high throughput metabolomics. *PLoS Computational Biology, 9*(7), e1003123. https://doi.org/10.1371/journal.pcbi.1003123

Lip, G. Y. H., Barnett, A. H., Bradbury, A., Cappuccio, F. P., Gill, P. S., Hughes, E., et al. (2007). Ethnicity and cardiovascular disease prevention in the United Kingdom: A practical approach to management. *Journal of Human Hypertension, 21*(3), 183–211. https://doi.org/10.1038/sj.jhh.1002126

Liu, J., Dolikun, M., Štambuk, J., & Wang, W. (2018). The association between subclass-specific IgG Fc *N*-glycosylation profiles and hypertension in the Uygur, Kazak, Kirgiz, and Tajik populations. *Journal of Human Hypertension, 32*(8–9), 555–563.

Lo, C.-H., Nguyen, L. H., Drew, D. A., Warner, E. T., Joshi, A. D., Graham, M. S., et al. (2021). Race, ethnicity, community-level socioeconomic factors, and risk of COVID-19 in the United States and the United Kingdom. *eClinicalMedicine, 38*, 101029. https://doi.org/10.1016/j.eclinm.2021.101029

Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., & Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. *Molecular Systems Biology, 3*, 135. https://doi.org/10.1038/msb4100177

Martínez-Reyes, I., & Chandel, N. S. (2020). Mitochondrial TCA cycle metabolites control physiology and disease. *Nature Communications, 11*(1), 1.

McGarrah, R. W., Crown, S. B., Zhang, G.-F., Shah, S. H., & Newgard, C. B. (2018). Cardiovascular metabolomics. *Circulation Research, 122*(9), 1238–1258. https://doi.org/10.1161/CIRCRESAHA.117.311002

Merino, D. M., Johnston, H., Clarke, S., Roke, K., Nielsen, D., Badawi, A., et al. (2011). Polymorphisms in FADS1 and FADS2 alter desaturase activity in young Caucasian and Asian adults. *Molecular Genetics and Metabolism, 103*(2), 171–178. https://doi.org/10.1016/j.ymgme.2011.02.012

Mikhail, N., Wali, S., & Brown, A. F. (2021). Ethnic disparities in diabetes. *Endocrinology and Metabolism Clinics of North America, 50*(3), 475–490. https://doi.org/10.1016/j.ecl.2021.05.006

Misra, B. B. (2021). Advances in high resolution GC–MS technology: A focus on the application of GC-Orbitrap-MS in metabolomics and exposomics for FAIR practices. *Analytical Methods, 13*(20), 2265–2282. https://doi.org/10.1039/d1ay00173f

Mojsak, P., Miniewska, K., Godlewski, A., Citko, A., Adamska-Patruno, E., Samczuk, P., et al. (2021). A preliminary study showing the impact of genetic and dietary factors on GC–MS-based plasma metabolome of patients with and without PROX1-genetic predisposition to T2DM up to 5 years prior to prediabetes appearance. *Current Issues in Molecular Biology, 43*(2), 513–528. https://doi.org/10.3390/cimb43020039

Mukadam, N., Marston, L., Lewis, G., Mathur, R., Rait, G., & Livingston, G. (2023). Incidence, age at diagnosis and survival with dementia across ethnic groups in England: A longitudinal study using electronic health records. *Alzheimer's and Dementia, 19*(4), 1300–1307. https://doi.org/10.1002/alz.12774

Mycielska, M. E., James, E. N., & Parkinson, E. K. (2022). Metabolic alterations in cellular senescence: The role of citrate in ageing and age-related disease. *International Journal of Molecular Sciences, 23*(7), 3652. https://doi.org/10.3390/ijms23073652

Navarro, S. L., Nagana Gowda, G. A., Bettcher, L. F., Pepin, R., Nguyen, N., Ellenberger, M., et al. (2023). Demographic, health and lifestyle factors associated with the metabolome in older women. *Metabolites, 13*(4), 514. https://doi.org/10.3390/metabo13040514

Orešič, M., McGlinchey, A., Wheelock, C. E., & Hyötyläinen, T. (2020). Metabolic signatures of the exposome—Quantifying the impact of exposure to environmental chemicals on human health. *Metabolites, 10*(11), 454. https://doi.org/10.3390/metabo10110454

Pageon, H., Azouaoui, A., Zucchi, H., Ricois, S., Tran, C., & Asselineau, D. (2019). Potentially beneficial effects of rhamnose on skin ageing: An in vitro and in vivo study. *International Journal of Cosmetic Science, 41*(3), 213–220. https://doi.org/10.1111/ics.12523

Pan, Y., Liu, P., Li, Y., & Ma, L. (2022). Abnormalities in the citrate cycle metabolism are associated with declined intrinsic capacity in older adults. *Innovation in Aging, 6*(Suppl 1), 325–326. https://doi.org/10.1093/geroni/igac059.1285

Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., & Xia, J. (2022). Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nature Protocols, 17*(8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w

Patel, J. V., Tracey, I., Hughes, E. A., & Lip, G. Y. (2010). Omega-3 polyunsaturated acids and cardiovascular disease: Notable ethnic differences or unfulfilled promise? *Journal of Thrombosis and Haemostasis, 8*(10), 2095–2104. https://doi.org/10.1111/j.1538-7836.2010.03956.x

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., et al. (2011). The human serum metabolome. *PLoS ONE, 6*(2), e16957. https://doi.org/10.1371/journal.pone.0016957

Ramos-Chávez, L. A., Roldán-Roldán, G., García-Juárez, B., González-Esquivel, D., Pérez de la Cruz, G., Pineda, B., et al. (2018). Low serum tryptophan levels as an indicator of global cognitive performance in nondemented women over 50 years of age. *Oxidative Medicine and Cellular Longevity, 2018*, 8604718. https://doi.org/10.1155/2018/8604718

Ren, J., Lu, Y., Qian, Y., Chen, B., Wu, T., & Ji, G. (2019). Recent progress regarding kaempferol for the treatment of various diseases (review). *Experimental and Therapeutic Medicine, 18*(4), 2759–2776. https://doi.org/10.3892/etm.2019.7886

Roberts, I., Wright Muelas, M., Taylor, J. M., Davison, A. S., Xu, Y., Grixti, J. M., et al. (2022). Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome. *Metabolomics, 18*(1), 6. https://doi.org/10.1007/s11306-021-01859-3

Saito, K., Maekawa, K., Kinchen, J. M., Tanaka, R., Kumagai, Y., & Saito, Y. (2016). Gender- and age-associated differences in serum metabolite profiles among Japanese populations. *Biological & Pharmaceutical Bulletin, 39*(7), 1179–1186. https://doi.org/10.1248/bpb.b16-00226

Salisbury, C., Johnson, L., Purdy, S., Valderas, J. M., & Montgomery, A. A. (2011). Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *The British Journal of General Practice, 61*(582), e12–e21. https://doi.org/10.3399/bjgp11X548929

Smith, G. D. (2011). Epidemiology, epigenetics and the 'Gloomy Prospect': Embracing randomness in population health research and practice. *International Journal of Epidemiology, 40*(3), 537–562. https://doi.org/10.1093/ije/dyr117

Stowe, R. P., Peek, M. K., Cutchin, M. P., & Goodwin, J. S. (2010). Plasma cytokine levels in a population-based study: Relation to age and ethnicity. *The Journals of Gerontology Series a, Biological Sciences and Medical Sciences, 65*(4), 429–433. https://doi.org/10.1093/gerona/glp198

Stratakis, N., Siskos, A. P., Papadopoulou, E., Nguyen, A. N., Zhao, Y., Margetaki, K., et al. (2022). Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health. *eLife, 11*, e71332. https://doi.org/10.7554/eLife.71332

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics : Official Journal of the Metabolomic Society, 3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

Tai, D. B. G., Sia, I. G., Doubeni, C. A., & Wieland, M. L. (2022). Disproportionate impact of COVID-19 on racial and ethnic minority groups in the United States: A 2021 update. *Journal of Racial and Ethnic Health Disparities, 9*(6), 2334–2339. https://doi.org/10.1007/s40615-021-01170-w

Theodoridis, G., Gika, H., Raftery, D., Goodacre, R., Plumb, R. S., & Wilson, I. D. (2023). Ensuring fact-based metabolite identification in liquid chromatography–mass spectrometry-based metabolomics. *Analytical Chemistry, 95*(8), 3909–3916. https://doi.org/10.1021/acs.analchem.2c05192

Tishkoff, S. A., & Kidd, K. K. (2004). Implications of biogeography of human populations for "race" and medicine. *Nature Genetics, 36*(11), S21–S27. https://doi.org/10.1038/ng1438

Trivedi, D. K. (2012). The application of SIMCA P+ in shotgun metabolomics analysis of ZICRHILIC-MS spectra of human urine—Experience with the Shimadzu IT-T of and profiling solutions data

extraction software. *Journal of Chromatography and Separation Techniques, 03*(06), 1.

Trivedi, D. K., Hollywood, K. A., & Goodacre, R. (2017). Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons in Translational Medicine, 3*(6), 294–305.

van der Goot, A. T., & Nollen, E. A. A. (2013). Tryptophan metabolism: Entering the field of aging and age-related pathologies. *Trends in Molecular Medicine, 19*(6), 336–344. https://doi.org/10.1016/j.molmed.2013.02.007

Vasishta, S., Ganesh, K., Umakanth, S., & Joshi, M. B. (2022). Ethnic disparities attributed to the manifestation in and response to type 2 diabetes: Insights from metabolomics. *Metabolomics, 18*(7), 45. https://doi.org/10.1007/s11306-022-01905-8

Wang, Y., Klarić, L., Yu, X., & Wang, W. (2016). The association between glycosylation of immunoglobulin g and hypertension. *Medicine (United States), 95*(17), 1–11.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics*. https://doi.org/10.1007/s11306-007-0099-6

WHO. (2022, October 1). Ageing and health. www.who.int. Organisation. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed 15 September 2023

Winkler, C., An, P., & O'Brien, S. J. (2004). Patterns of ethnic diversity among the genes that influence AIDS. *Human Molecular Genetics, 13*(suppl_1), R9–R19. https://doi.org/10.1093/hmg/ddh075

Yaya, S., Yeboah, H., Charles, C. H., Otu, A., & Labonte, R. (2020). Ethnic and racial disparities in COVID-19-related deaths: Counting the trees, hiding the forest. *BMJ Global Health, 5*(6), e002913. https://doi.org/10.1136/bmjgh-2020-002913

Yurekten, O., Payne, T., Tejera, N., Amaladoss, F. X., Martin, C., Williams, M., & O'Donovan, C. (2024). MetaboLights: Open data repository for metabolomics. *Nucleic Acids Research, 52*(D1), D640–D646. https://doi.org/10.1093/nar/gkad1045

Zeki, Ö. C., Eylem, C. C., Reçber, T., Kır, S., & Nemutlu, E. (2020). Integration of GC–MS and LC–MS for untargeted metabolomics profiling. *Journal of Pharmaceutical and Biomedical Analysis, 190*, 113509. https://doi.org/10.1016/j.jpba.2020.113509

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