



# Global metabolic profiling to model biological processes of aging in twins

Bryan J. Bunning<sup>1</sup> | K vin Contre    <sup>2</sup> | Brittany Lee-McMullen<sup>2</sup> | Gopal Krishna R. Dhondalay<sup>1</sup> | Wenming Zhang<sup>1</sup> | Dana Tupa<sup>1</sup> | Olivia Raeber<sup>1</sup> | Manisha Desai<sup>3</sup> | Kari C. Nadeau<sup>1</sup> | Michael P. Snyder<sup>2</sup> | Sandra Andorf<sup>1</sup>

<sup>1</sup>Department of Medicine, Sean N. Parker Center for Allergy and Asthma Research, Stanford University School of Medicine, Stanford, California

<sup>2</sup>Department of Genetics, Stanford University School of Medicine, Stanford, California

<sup>3</sup>Quantitative Science Unit, Department of Medicine, Stanford University School of Medicine, Stanford, California

## Correspondence

Sandra Andorf, Department of Medicine, Sean N. Parker Center for Allergy and Asthma Research, Stanford University School of Medicine, 300 Pasteur Drive, Grant Building, Stanford, CA 94305. Email: andorf@stanford.edu

## Funding information

Funding was provided by the National Institutes of Health (NIH) R01 (HL118612 PI, Nadeau), National Institute of Environmental Health Sciences (NIEHS) R21 (ES022153-02 PI, Nadeau), NIH Common Fund Human Microbiome Project (HMP) (1U54DE02378901, PI, Snyder), the Stanford Metabolic Health Center, and the Sean N. Parker Center for Allergy and Asthma Research at Stanford University.

## Abstract

Aging is intimately linked to system-wide metabolic changes that can be captured in blood. Understanding biological processes of aging in humans could help maintain a healthy aging trajectory and promote longevity. We performed untargeted plasma metabolomics quantifying 770 metabolites on a cross-sectional cohort of 268 healthy individuals including 125 twin pairs covering human lifespan (from 6 months to 82 years). Unsupervised clustering of metabolic profiles revealed 6 main aging trajectories throughout life that were associated with key metabolic pathways such as progestin steroids, xanthine metabolism, and long-chain fatty acids. A random forest (RF) model was successful to predict age in adult subjects ( $\geq 16$  years) using 52 metabolites ( $R^2 = .97$ ). Another RF model selected 54 metabolites to classify pediatric and adult participants (out-of-bag error = 8.58%). These RF models in combination with correlation network analysis were used to explore biological processes of healthy aging. The models highlighted established metabolites, like steroids, amino acids, and free fatty acids as well as novel metabolites and pathways. Finally, we show that metabolic profiles of twins become more dissimilar with age which provides insights into nongenetic age-related variability in metabolic profiles in response to environmental exposure.

## KEYWORDS

aging, LC-MS, machine learning, metabolomics, random forest, twins

## 1 | INTRODUCTION

Aging is a complex and multi-factorial process that involves system-wide dysregulation across many biological pathways and molecule types (Hoffman, Lyu, Pletcher, & Promislow, 2017). Understanding

mechanisms of healthy aging in humans is important because early detection of deviations from a healthy aging trajectory could be used to promote longevity by delaying, avoiding, or preventing the development of age-related diseases (Alpert et al., 2019). However, this task has been challenging in part because of the complex interplay

Bunning and Contre     contributed equally to the work.

Nadeau, Snyder and Andorf considered as co-senior authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

  2019 The Authors. *Aging Cell* published by Anatomical Society and John Wiley & Sons Ltd.

of genetic, environmental, and lifestyle factors involved in aging. Metabolomics—the unbiased profiling of a large panel of metabolites—has become the approach of choice to study aging with the realization that metabolic alterations are central to the aging process (Barzilai, Huffman, Muzumdar, & Bartke, 2012; Collino et al., 2013; López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013; Montoliu et al., 2014). In addition, metabolomics captures genetic and nongenetic factors since metabolites are influenced by the biology of the host, gut microbial activity, and environmental exposure (Rinschen, Ivanisevic, Giera, & Siuzdak, 2019). Recent studies combining genetic and metabolic profiles in twins were instrumental in quantifying the relative influence of genetic and environmental factors on blood metabolite levels and showed a wide range of heritability largely dependent on chemical classes (Darst, Kosciak, Hogan, Johnson, & Engelman, 2019; Kettunen et al., 2012; Long et al., 2017; Shin et al., 2014).

In the recent years, metabolomics has been successfully applied to the study of human aging highlighting many biomarkers and biological pathways associated with age (Chaleckis, Murakami, Takada, Kondoh, & Yanagida, 2016; Menni et al., 2013; Rist et al., 2017; Yu et al., 2012). Through various modeling and experimental methodologies, steroids, amino and fatty acids, and biomarkers of kidney function have been found to have significant associations with aging. However, these studies were often limited by the age range of participants focusing on adults (>18 years old) and by the range of metabolites covered (~100–500).

In this context, we profiled a wide array of 770 metabolites in blood from a cross-sectional cohort of healthy individuals enriched for twin pairs and aged 6 months to 82 years old. The aim of the study was threefold: (a) describe longitudinal trends of changes with age, (b) identify metabolites and pathways associated with age using random forest (RF) machine learning algorithm (regression and classification models), and (c) investigate biological variability of metabolic profiles in twin pairs with age.

## 2 | RESULTS

### 2.1 | Cohort characteristics and generation of untargeted metabolomics data

A cohort consisting of 268 healthy individuals including 125 twin pairs was recruited with the objective of having an even distribution of ages across a human lifespan (6 months to 82 years old) (Figure 1a). Demographic characteristics are shown in Table 1. The cohort was divided in two groups including pediatric (<16 years old) and adult participants (16–82 years) with the assumption that puberty is over at age 16. The pediatric population contained fewer female participants (45.3% vs. 69.8%) and a lower percentage of monozygotic twins among the twin pairs (64% vs. 91%) than the adult population.

Untargeted metabolomics was performed on randomized plasma samples using a broad range liquid chromatography-mass spectrometry (LC-MS) platform (Contrepois, Jiang, & Snyder, 2015). This platform uses complementary hydrophilic liquid chromatography (HILIC)

and reverse-phase liquid chromatography (RPLC) separations coupled with high-resolution mass spectrometry (HRMS). Over 17,000 MS peaks were robustly detected. After filtering and curation, 770 metabolites could be identified (Figure 1a). The named metabolites were categorized in eight main chemical classes and 59 pathways to help biological interpretation and enrichment analysis (Table S1). Annotation confidence levels for each metabolite were provided following the Metabolomics Standards Initiative (MSI) confidence scheme (Table S1) (Schymanski et al., 2014; Sumner et al., 2007).

### 2.2 | Metabolic profiles are influenced by age

Variance decomposition was performed on recorded confounders including age, body mass index (BMI), sex, asthma status (GINA score) (Bousquet, 2000), smoking and second-hand smoking status (Figure S1). Among the tested variables, age showed the most variance in our dataset (1.01% median, 44.50% maximum). In addition, a global aging trajectory in the cohort could be visualized in two dimensions by plotting the principal components (PCs) PC4 ( $p = 2.2\text{E-}12$ ) and PC5 ( $p = 1.4\text{E-}15$ ) that associated the most strongly with age (see methods, Figure 1b).

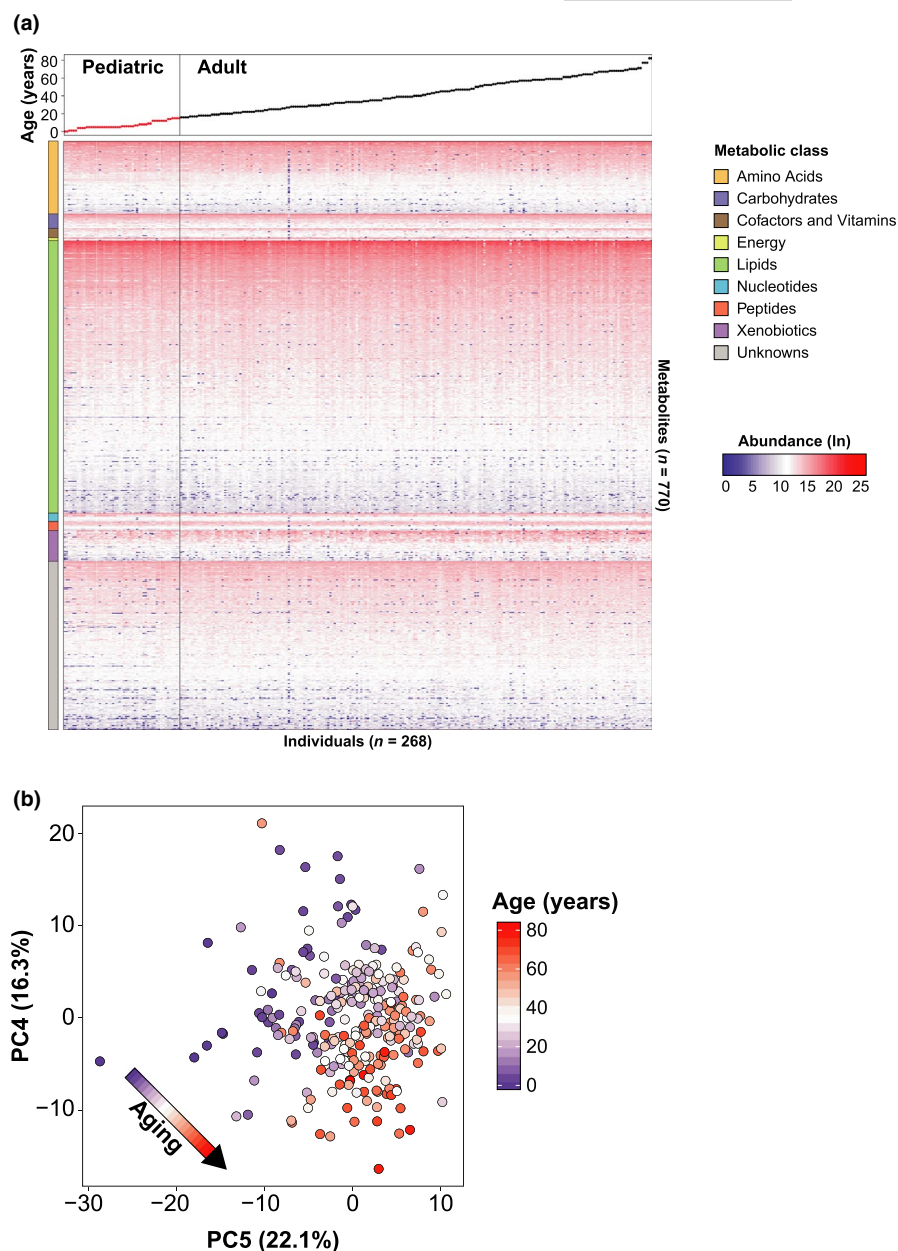
### 2.3 | Metabolic aging trajectories

We took advantage of the even age distribution of our cohort across a pediatric and an adult population and defined six clusters based on metabolite aging trajectories (Figure 2, Table S1). Pathway enrichment analysis via hypergeometric testing was used to highlight biological functions enriched in each cluster ( $\text{FDR} < 0.1$ ). Clusters 1, 4, 5, and 6 contained 84% of the metabolites. Metabolites in cluster 1 increased strongly until adult age and then tended to slowly increase with age. The metabolites belonging to this cluster were enriched for xanthine and histidine metabolism. Cluster 4 was enriched for acylcarnitines, as well as long and polyunsaturated fatty acids. Molecules in this cluster tended to remain constant at young age and linearly increased with age after the onset of adulthood. Clusters 5 and 6 presented opposing parabolic shapes with metabolites decreasing at early age and then increasing at adult age for cluster 5 and vice versa in cluster 6. The latter cluster was enriched for monoacylglycerols and progestin steroids. On average, metabolites in cluster 5 reached a minimum at age  $29.6 \pm 9.2$  years (mean, standard deviation) and metabolites in cluster 6 were the most concentrated at age  $31.0 \pm 11.2$  years. These cluster classifications were used to guide biological interpretations in the sections below.

### 2.4 | Machine learning reveals metabolites associated with age in a healthy adult population

A regression random forest (RF) model analysis to predict age was performed on adults aged 16 years and older to focus on healthy

**FIGURE 1** Untargeted metabolomics of aging plasma. (a) Natural log metabolite abundances of all 770 detected metabolites (rows) in 268 individuals (columns) ranging from low (blue) to high (red). Metabolites are ordered by metabolic class and median intensity across all the samples in the study, and individuals are ordered chronologically by age. (b) Principal component analysis using the top two principal components associated with age (*i.e.*, PC4 and PC5). Association with age is calculated via linear modeling of each PC



aging in adults ( $n = 215$ ). The dataset was modeled to correct for BMI and sex as these demographics are known to cause variability in metabolic profiles (Darst et al., 2019; Piening et al., 2018) (Figure S1). A model containing 52 metabolites (Figure 3a) successfully predicted healthy aging in adults with a mean squared error of 4.45, R-square of 0.97, and 54.78% of variance explained (Figure 3b). Interestingly, metabolites in clusters 4 (increase with age) and 6 (decrease with age) dominated representing 50.0% and 19.2% of the selected metabolites, respectively. As expected, androsterone and progesterone derivatives contributed the most to the model (Darst et al., 2019; Feldman et al., 2002) including dehydroisoandrosterone sulfate (DHEA-S), pregnanolone sulfate, and 16- $\alpha$ -hydroxy-DHEA-3-sulfate, of which all presided in cluster 6. In addition to steroid hormones, metabolites involved in amino acid (*i.e.*, phenylacetylglutamine, cystine, phenylalanine, citrulline, tryptophan) and lipid metabolism (*i.e.*, sn-glycero-3-phosphoethanolamine and polyunsaturated fatty acids

eicosapentaenoic acid and arachidonic acid) contributed to the RF model. Three unknown metabolites with elemental compositions C<sub>9</sub>H<sub>17</sub>NO<sub>3</sub>, C<sub>19</sub>H<sub>34</sub>O<sub>15</sub>, and C<sub>21</sub>H<sub>34</sub>O<sub>6</sub>S were among the most important metabolites in the model but could not be confidently annotated. Further work would be necessary to formally identify these molecules. Interestingly, C<sub>19</sub>H<sub>34</sub>O<sub>15</sub> and C<sub>21</sub>H<sub>34</sub>O<sub>6</sub>S resided in cluster 6 and had the same longitudinal pattern of change with age as androsterone and progesterone derivatives suggesting similar function and/or structure (Figure S2A). The unknown metabolite with the elemental composition C<sub>9</sub>H<sub>17</sub>NO<sub>3</sub> displayed a different aging trajectory with a positive linear association in adulthood as seen in cluster 4.

Correlation network analysis of metabolites included in the RF model revealed distinct functional clusters of co-regulated metabolites including androgenic and progestin hormones, fatty acids, and amino acids (Figure 3c). Betweenness centrality that represents

the importance of the node in the network was calculated for each metabolite. Interestingly, the unknown metabolite with elemental composition C<sub>12</sub>H<sub>25</sub>N<sub>3</sub>O<sub>3</sub>, residing in cluster 4, was central in the network and was ranked 9th for its importance in the RF regression model, suggesting it may have a critical function in the aging process.

2.5 | Machine learning reveals metabolites that distinguish pediatric and adult participants

Next, we were interested in identifying the metabolites that would classify samples from pediatric (*n* = 53) and adult (*n* = 215)

TABLE 1 Cohort demographics

	<16 years	≥16 years	Total
Individuals (twin sets)	53 (25)	215 <sup>a</sup> (100)	268 (125)
Age in years (mean, SD)	7.0 ± 4.1	42.5 ± 17.3	35.5 ± 21.1
Sex % female ( <i>n</i> )	45.3% (24)	69.8% (150)	64.9% (174)
BMI (mean, SD)	16.0 ± 2.7	27.0 ± 6.3	24.9 ± 7.2
Monozygotic individuals % of twins ( <i>n</i> )	64.0% (32) <sup>b</sup>	91.0% (182)	85.6% (214)

<sup>a</sup>12 Singletons, 1 set of triplets.

<sup>b</sup>Zygosity unknown for 1 twin pair.

participants. An RF classification model determined 54 metabolites (Figure 4a) important in differentiating pediatric participants from adults with an out-of-bag estimate of error of 8.58%. In contrast to the regression model, the classification model selected metabolites enriched in cluster 1 (more abundant in adults) and 5 (less abundant in adults) representing 40.7% and 24.1%, respectively. The classification error was very small for adults (3%) and was greater for pediatric participants (28%) likely because of the smaller sample size of the latter (Figure 4b). As expected, metabolite biomarkers for the consumption of coffee (*i.e.*, caffeine, theophylline, trigonelline, quinic acid and 1-methyl uric acid), residing in cluster 1, as well as androsterone derivatives (*i.e.*, androsterone glucuronide and DHEA-S) in cluster 6, were important in the RF classification model. All these molecules were more abundant in adult participants. In addition, we observed two glycosylated amino acids that were important to the model, namely galactosyl-hydroxylysine and glucosyl-galactosyl-hydroxylysine. The level of these molecules was decreased in the adult group (Figure S2B). We also found creatinine, C-glycosyl tryptophan,  $\gamma$ -carboxyethyl hydroxymannan (CEHC) glucuronide,  $\gamma$ -glutamyl- $\epsilon$ -lysine,  $\alpha$ -glutamyl-lysine, proline-hydroxyproline and acetylcarnosine as main contributors in the model.

Correlation network analysis of metabolites selected in the model highlighted a cluster of steroid hormones containing androsterone derivatives as well as a cluster of xenobiotics coming from coffee consumption (Figure 4c). Interestingly, creatinine had a central position in the network and positively associated with 20 metabolites (Bonferroni adjusted *p*-value <.01) bridging the two main

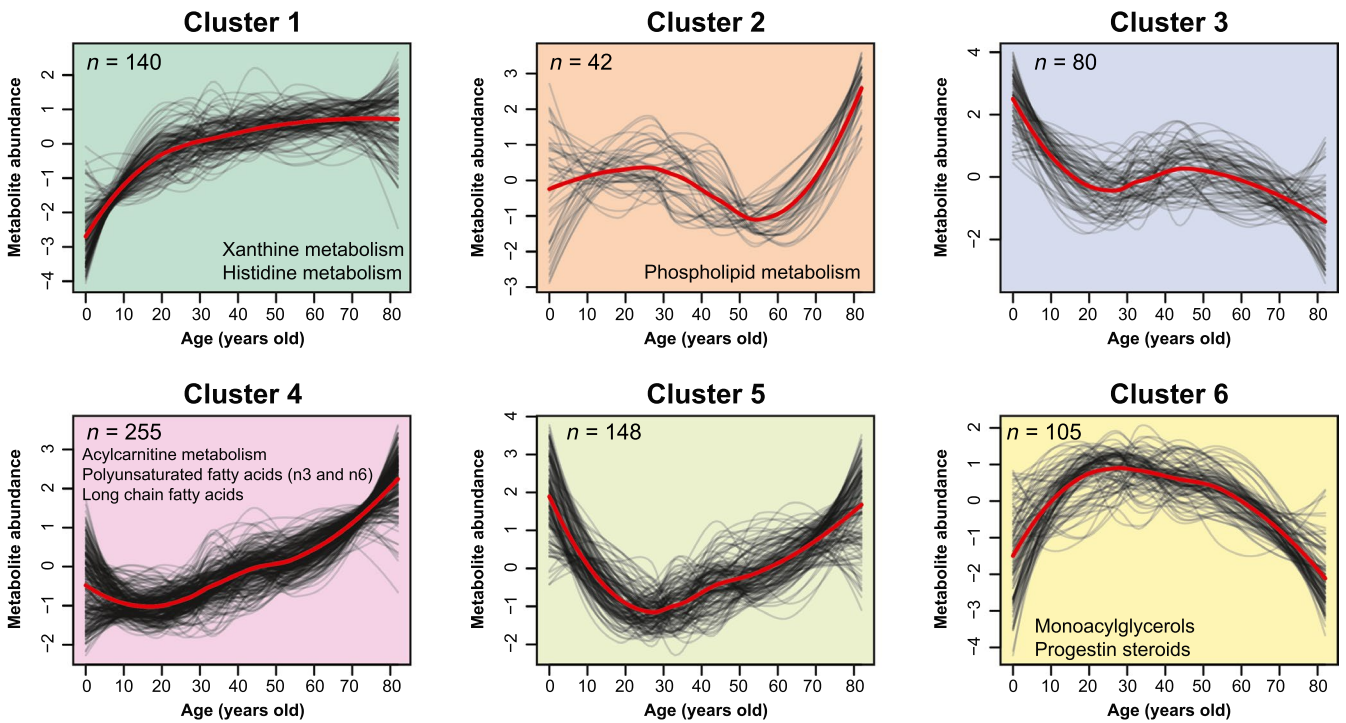
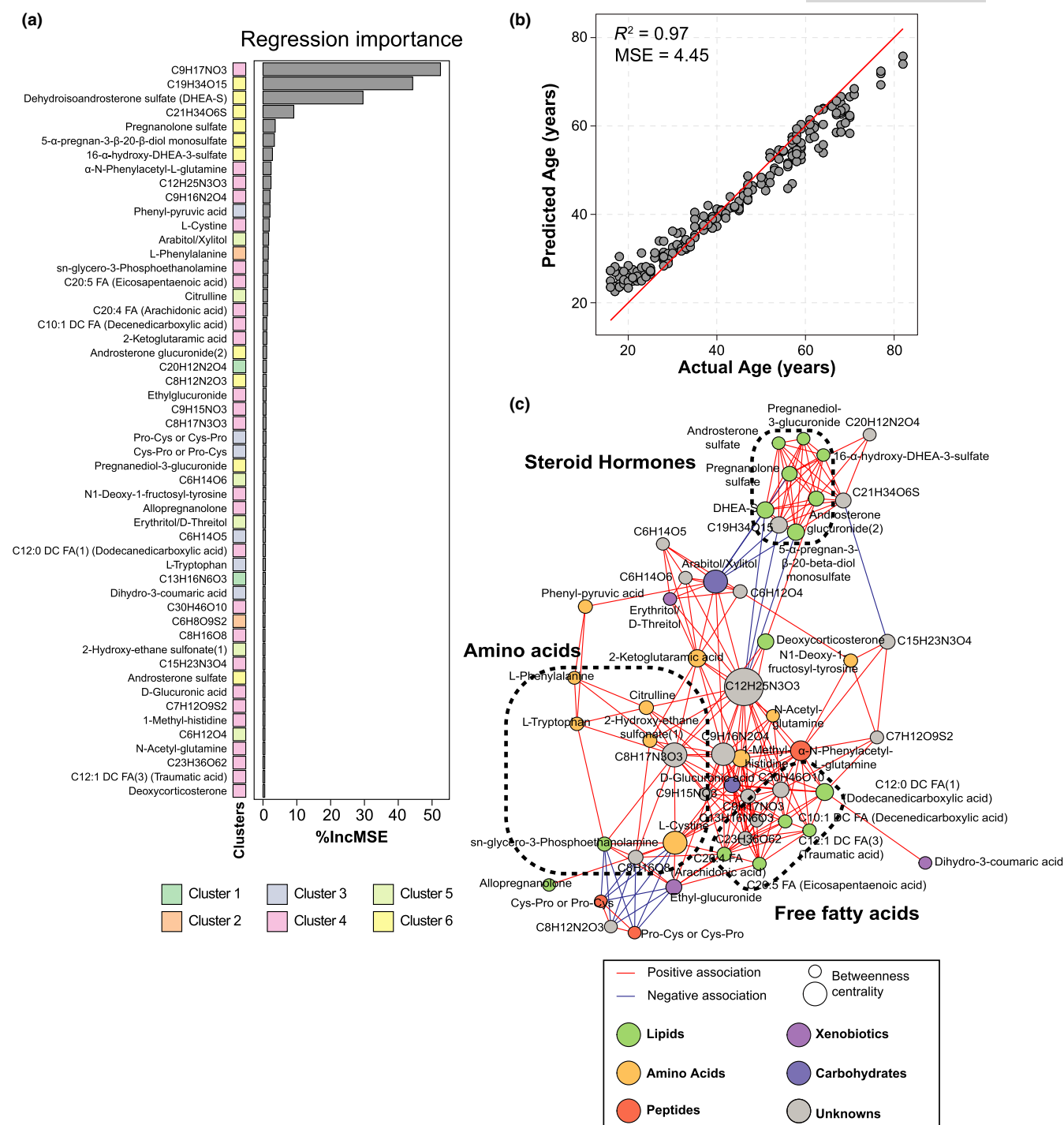


FIGURE 2 Metabolic aging trajectories. Fuzzy c-mean clustering of all 770 metabolite abundances fitted to a loess curve and Z-score scaled (black lines), adjusted for sex and BMI, as a function of age in years. Average trend of clusters is shown as a red line. Pathways with FDR < 0.1 are considered significant and displayed



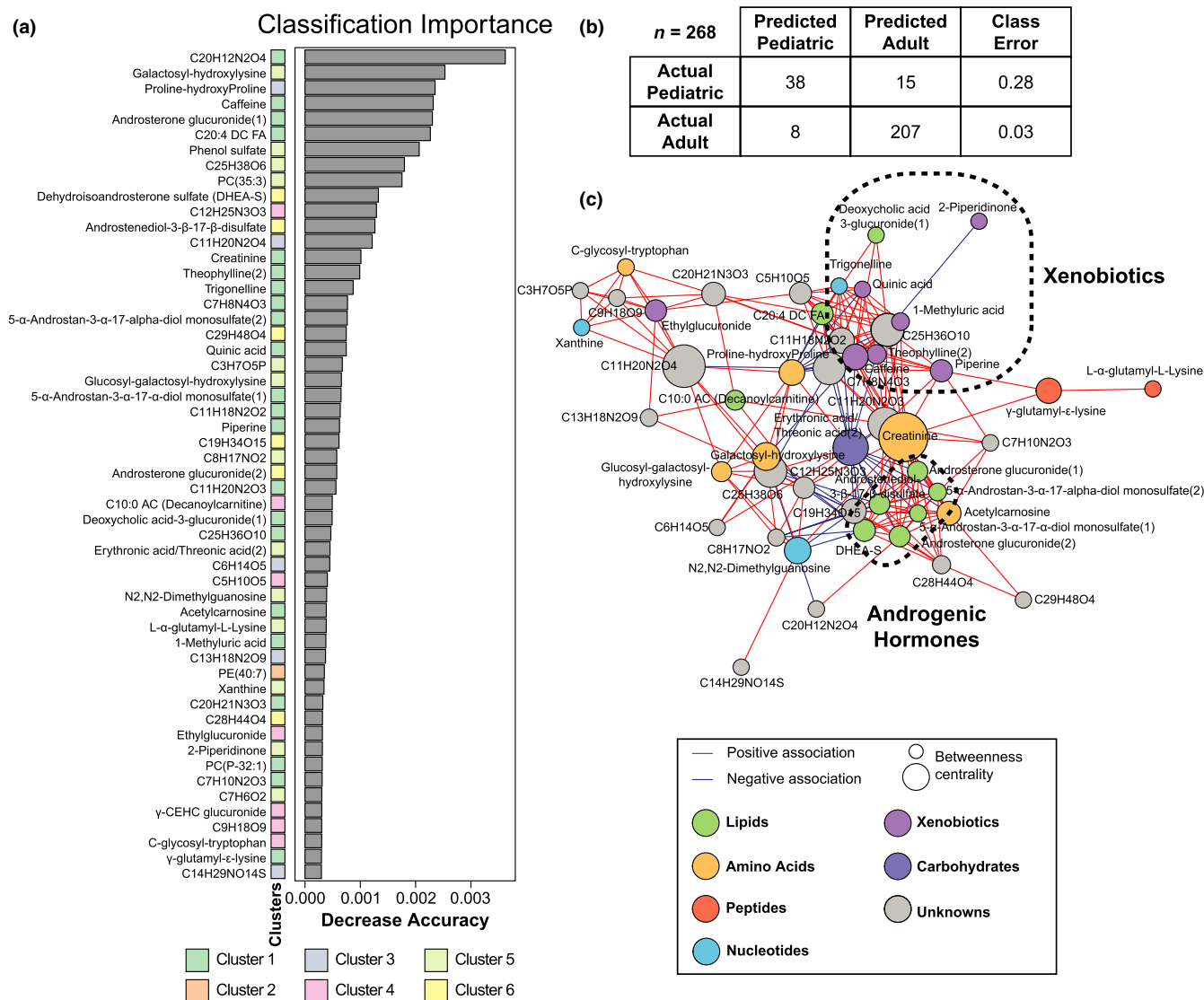


**FIGURE 3** Regression model Random Forest analysis. (a) Significant metabolites in the regression RF model ordered by importance and color-coded by cluster. (b) 2D scatter plot representing predicted age vs. actual age. MSE = mean squared error. (c) Correlation network analysis of significant metabolites in the model. Nodes are colored by metabolic class and sized by betweenness centrality. Edges are colored by association direction

clusters containing androgenic hormones and xenobiotics. Clinically, plasma creatinine is used to monitor renal function and was found to increase with age in our cohort (Shlipak et al., 2009; Tiao, Semmens, Masarei, & Lawrence-Brown, 2002).

Interestingly, the metabolites selected in the two RF models generated to predict age in a healthy adult population and to

classify pediatric and adult participants presented little overlap ( $n = 7$  metabolites). While the regression RF model successfully selected metabolites involved in energy metabolism (i.e., amino acid and lipid metabolism), the classification model selected many xenobiotics including metabolites derived from the consumption of coffee.



**FIGURE 4** Classification model Random Forest analysis. (a) Significant metabolites in the classification RF model ordered by importance and color-coded by cluster. (b) Two-way table of the classification model. (c) Correlation network analysis of significant metabolites in the model. Nodes are colored by metabolic class and sized by betweenness centrality. Edges are colored by association direction

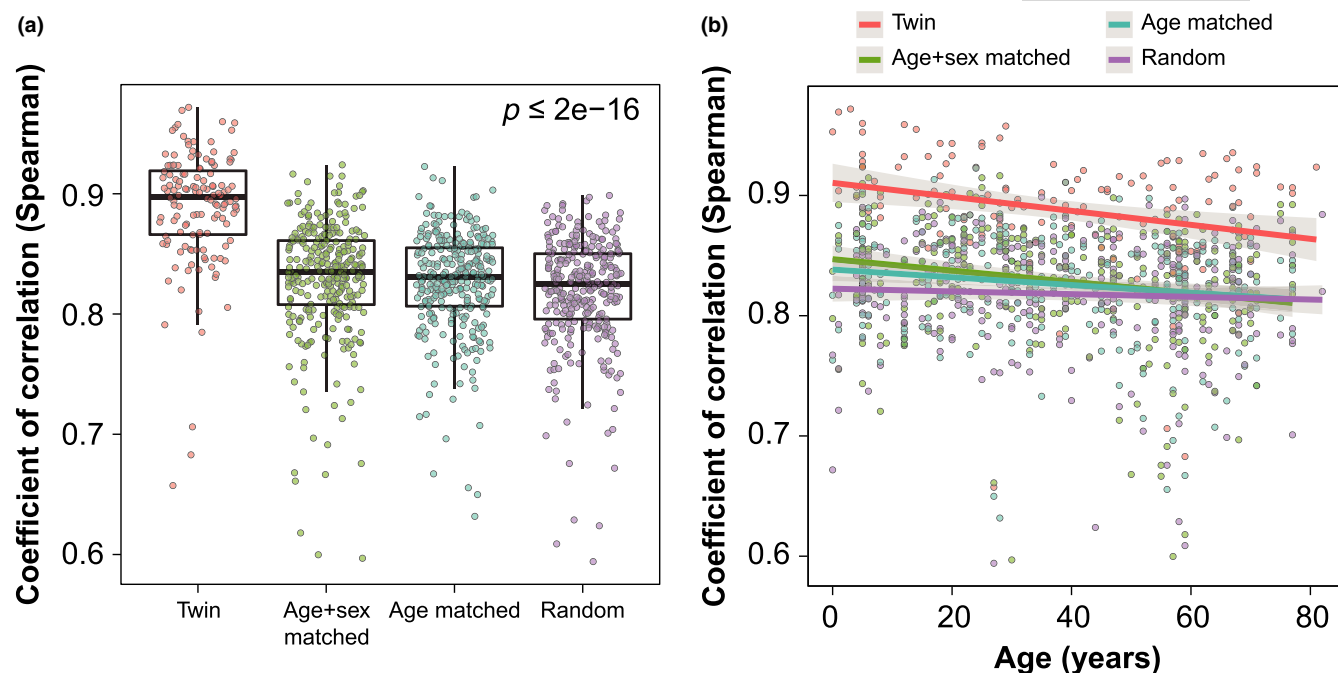
## 2.6 | Variability of metabolic profiles in twin pairs as a function of age

Since our cohort was enriched for twin pairs, and specifically monozygotic pairs, we assessed the variability of metabolic profiles between twins with age. As expected, we found that the Spearman rank correlation coefficient calculated across all detected metabolites was significantly higher between twin samples (median  $\rho = 0.894$ ) than between age- and sex-matched participants (median  $\rho = 0.835$ ,  $p = 2.7\text{E-}31$ ), age-matched participants (median  $\rho = 0.831$ ,  $p = 6.4\text{E-}36$ ), and randomly selected pairs (median  $\rho = 0.825$ ,  $p = 8.5\text{E-}40$ ) (Figure 5a). Linear modeling of pairwise Spearman's correlation as a function of age resulted in a significant negative term for age in twins ( $p = 1.1\text{E-}3$ ) as well as age- and sex-matched ( $p = 4.1\text{E-}4$ ) and age-matched pairs ( $p = 5.2\text{E-}3$ ). Random pairs without age or sex matching showed no significant association

of Spearman's rank correlations and age ( $p = .36$ ) (Figure 5b). The results in twins were not driven by twin status since the cohort had more dizygotic twins in pediatric than in adult individuals (Table 1). Via linear model, we calculated that the pairwise correlation between metabolic profiles of a given twin pair decreases by 0.0005906 each year. These observations may reflect the impact of differential environmental exposure on twin pairs that accumulate with age.

## 3 | DISCUSSION

Metabolomics has become a popular approach to study biological processes associated with age and has been applied in various human studies (Barzilai et al., 2012; Collino et al., 2013; López-Otín et al., 2013; Montoliu et al., 2014; Rinschen et al., 2019). Our study expands on previous work by profiling 770 chemically diverse



**FIGURE 5** Variability of metabolic profiles in twins. (a) Spearman's correlation of the whole collection of metabolites between various pairs of individuals.  $p$ -value shown was calculated via Kruskal–Wallis test. (b) Spearman's correlation plotted as a function of age shows a significant decreasing trend over time in twins

plasma metabolites from twin pairs of pediatric and adult participants ( $n = 268$ ) with an even distribution of age (6 months to 82 years old). We used a sensitive and robust LC-MS platform involving complementary hydrophilic liquid chromatography (HILIC) and reverse-phase liquid chromatography (RPLC) separations to capture a wide array of metabolites covering 59 biochemical pathways (Contrepois et al., 2015). Our novel bioinformatic approach combining unsupervised clustering, machine learning and correlation networks recapitulated known biology but also led to discoveries.

Regression and classification RF selected a small number of metabolites that successfully predicted age in an adult population and classified pediatric and adult participants. Xenobiotics were among the strongest predictors that differentiated pediatric and adult participants and included many metabolites associated with coffee consumption (*i.e.*, caffeine). These molecules were much more abundant in the blood of adult subjects (cluster 1) which was expected given that coffee is frequently ingested by adults but rarely by children (Azam, Hadi, Khan, & Hadi, 2003; Fulgoni, Keast, & Lieberman, 2015). These results validate our study demonstrating that we generated good quality untargeted metabolomics data and we employed an efficient bioinformatic approach to extract meaningful biology from the data.

The classification model revealed two glycosylated forms of hydroxylysine found in collagen. Galactosyl-hydroxylysine is a marker of bone resorption (Al-Dehaimi, Blumsohn, & Eastell, 1999). Previous studies found galactosyl-hydroxylysine to be (a) higher in the blood of girls compared with premenopausal women reflecting high bone turnover and (b) higher in postmenopausal compared with premenopausal women reflecting higher bone resorption rates in

postmenopausal older women due to estrogen deficiency related with menopause (Feng & McDonald, 2011). Our data validate these findings with galactosyl-hydroxylysine describing a positive parabolic trajectory (cluster 5), and we found that glucosyl-galactosyl-hydroxylysine follows the same aging trajectory.

As expected, androsterone and progesterone derivatives were selected by machine learning algorithms because they vary greatly with age (Darst et al., 2019; Feldman et al., 2002). Dehydroepiandrosterone sulfate (DHEA-S), a sulfate ester of adrenal steroid dehydroepiandrosterone (DHEA), was selected with both models. Its levels are known to correlate strongly with age, and it has been shown to be an important regulator of some age-related biological processes (Feldman et al., 2002). However, neither DHEA nor low-dose testosterone replacement in elderly people presented physiologically relevant beneficial effects (Nair et al., 2006). Steroid hormones showed a negative parabolic pattern, with levels spiking at around age 30 that decay throughout adulthood (cluster 6).

In addition to steroid hormones, many amino acids and lipids were selected in the models. We detected some expected changes including an accumulation with age of citrulline and cystine, a dimeric form of cysteine, presumably due to impaired muscle function in the elderly (Pitkänen, Oja, Kemppainen, Seppä, & Mero, 2003). Importantly, our study included pediatric samples which allows to discriminate metabolites based on their changes before puberty. Even though citrulline and cystine are both known to increase with age, we found that they belonged to different clusters. Citrulline, in cluster 5, presented a parabolic curve with decreased levels with age before puberty followed by an increase with age in adults while cystine in cluster 4 remained constant before puberty and increased

with age in the adult population. These findings demonstrate the power of our approach to discover new metabolic aging trajectories throughout the human lifespan.

We also observed some modified amino acids that accumulated with age (cluster 4). Phenylacetylglutamine was important in the regression model (8th position) and has been described as a marker of healthy aging (Collino et al., 2013). Interestingly, it is produced by the gut microbiome from amino acid fermentation which is in line with the growing body of literature related to gut microbial composition shift with age (Nagpal et al., 2018; Yatsunenkov et al., 2012). We also detected C-glycosyl tryptophan in cluster 4 that was shown to strongly correlate with chronological age (Menni et al., 2013).

Two polyunsaturated fatty acids, arachidonic acid (AA), and eicosapentaenoic acid (EPA) were highlighted in our analysis and positively associated with age (cluster 4). Even though docosahexaenoic acid (DHA) was not selected in the models, it presented the same longitudinal trajectory as AA and EPA (Otsuka, Kato, Imai, Ando, & Shimokata, 2013). These molecules are important modulators of cardiovascular health and inflammation (James, Gibson, & Cleland, 2000). In a recent study, Deelen et al. (2019) identified a metabolic signature containing 14 biomarkers predictive of all-cause mortality. Interestingly, polyunsaturated fatty acids were important molecules both in their model to predict all-cause mortality independent of age and sex and in our model to predict age.

Correlation network analysis revealed that creatinine presented the most connections in the classification model highlighting its central role in the aging process. Creatinine is an important molecule that is used to measure renal function in the clinic and was found to increase strongly with age in the pediatrics cohort (cluster 1). Unknown metabolite C12H25N3O3 was central in the correlation network for the regression model bridging steroid hormones, free fatty acids, and amino acids which highlights its potential functions in the biological process of aging in healthy adults. Interestingly, some unknown metabolites were among the most important molecules in our models with C9H17NO3 that was ranked 1st in the regression RF model.

Importantly, the longitudinal trajectories of metabolites with age described in our analysis are best generalized as parabolic. Hence, a model using a second-order term ( $y = \beta \cdot \text{age}^2 + \beta \cdot \text{age}$ ) should be more appropriate than a simple linear model of age ( $y = \beta \cdot \text{age}$ ) to identify metabolites associated with age. Linear models have been widely used potentially impacting accuracy and significance of the findings (Menni et al., 2013; Yu et al., 2012).

We are the first, to our knowledge, to show a decrease in Spearman's rank correlation between twins' untargeted metabolic profiles as age increases. This observation is likely explained by the influence of nongenetic factors including lifestyle and environmental exposure on the metabolome. This increased variability with age was also observed in other biological systems such as the immune system (Brodin et al., 2015; Cheung et al., 2018). The average change in correlation over our cohorts age range of 80 years between twins ( $-0.0005906/\text{year} \cdot 80 \text{ years} = -0.046$ ) is on the order of the difference between the average of the twin correlations

and the average of the random age-matched correlations over the same age range ( $-0.073$ ). Comparing these two values helps to contextualize changes in correlations due to genetic factors, or due to aging. The relatively high correlations of metabolic profiles among randomly age-matched participants reflect the small variability in abundance of most metabolites between healthy individuals. As an example, a modest increase of circulating branched-chain amino acids (10%–20%) most consistently differentiates individuals with insulin resistance and/or type 2 diabetes (Guasch-Ferré et al., 2016; Schüssler-Fiorenza Rose et al., 2019).

This study should be assessed in the context of its limitations. First, the participants were not required to fast before the blood draw and blood samples were collected at any time of the day. However, we demonstrated the biological relevance of the metabolomics data generated since we were able to reproduce main expected changes associated with age. Second, our study is cross-sectional by design which limits its statistical power due to inter-individual variability of metabolic profiles. To circumvent this issue, two recent studies profiled metabolites in a longitudinal fashion (Chak et al., 2019; Darst et al., 2019). Even though these studies improve the robustness of the findings, they do not give insights into mechanisms resulting in the observed metabolic changes. Longitudinal multi-omics studies will likely be key to unravel molecular mechanisms that results in age-dependent metabolism dysregulation in human plasma by integrating information from multiple regulatory levels (Piening et al., 2018; Zhou et al., 2019). Lastly, younger twins most likely still live together or have lived apart for a shorter period of time and, thus, have more recently shared environmental exposures than older twins, which has to be considered when interpreting the results of the twin analysis.

Overall, we show that our computational approach involving clustering, RF machine learning, and network analysis was successful to describe metabolic aging trajectories throughout life and discover metabolites and pathways dysregulated with age. Additionally, we revealed an increased variability of metabolic profiles in twin pairs with age suggesting accumulating effects of differential exposure on human metabolome.

## 4 | EXPERIMENTAL PROCEDURES

### 4.1 | Study design and participants

A cohort of 268 participants enriched for twins was recruited as part of an observational study (ClinicalTrials.gov Identifier: NCT01613885), which included twins of all ages from sites in Washington and California between 2010 and 2016. Individuals with an active infection, history of cancer, autoimmune disease, or that use oral steroids or immunomodulators were not included in this analysis. As such, participants were considered healthy. The cohort was composed of 125 twin pairs, 1 triplet set, and 15 singlets consisting of family members or instances of only one twin arriving at the appointment. In addition to a routine examination, participants



completed a detailed questionnaire regarding their medical history as well as dietary habits. Individual-level demographic information, self-reported disease status, and self-reported frequent medication usage are provided in Table S2. This study was conducted with the approval of Stanford University Institutional Review Board (IRB-19495), and all subjects provided signed informed consent before participation.

## 4.2 | Blood collection and plasma preparation

Intravenous whole-blood was collected nonfasted and at any time of the day in 10 ml EDTA-vacutainer tubes, and plasma was prepared within 24 hr following centrifugation at 2,200 rpm for 20 min at 24°C. Blood was kept at room temperature prior to processing. The top layer plasma was pipetted off, aliquoted, and immediately frozen at -80°C. Plasma aliquots were de-identified in agreement with CFR/GCP (code of federal regulation/good clinical practice) guidelines. Samples from twin pairs were collected the same day at the same time.

## 4.3 | Chemicals

LC-MS-grade solvents and mobile phase modifiers were obtained from Fisher Scientific (water, acetonitrile, methanol, formic acid) and Sigma-Aldrich (ammonium acetate).

## 4.4 | Metabolite extraction

Fresh plasma aliquots were used for the study, and metabolites were extracted in a randomized order as previously described (Contrepois et al., 2015). Briefly, metabolites were prepared from 100 µl of plasma using 1:1:1 acetone:acetonitrile:methanol, evaporated to dryness under nitrogen, and reconstituted in 1:1 methanol:water.

## 4.5 | Untargeted metabolomics profiling

Metabolic extracts were analyzed four times using hydrophilic liquid chromatography (HILIC) and reverse-phase liquid chromatography (RPLC) separation in both positive and negative ionization modes. Data were acquired on a Thermo Q Exactive plus mass spectrometer equipped with a HESI-II probe and operated in full MS scan mode. MS/MS data were acquired on pool samples consisting of an equimolar mixture of 100 randomized samples in the study. HILIC experiments were performed using a ZIC-HILIC column 2.1 × 100 mm, 3.5 µm, 200Å (Merck Millipore), and mobile phase solvents consisting of 10 mM ammonium acetate in 50/50 acetonitrile/water (A) and 10 mM ammonium acetate in 95/5 acetonitrile/water (B). RPLC experiments were performed using a Zorbax SBaq column 2.1 × 50 mm, 1.7 µm, 100Å (Agilent Technologies), and mobile phase

solvents consisting of 0.06% acetic acid in water (A) and 0.06% acetic acid in methanol (B).

Data quality was ensured by (a) sample randomization for metabolite extraction and data acquisition, (b) multiple injections of a pool sample to equilibrate the LC-MS system prior to run the sequence (12 and 6 injections for HILIC and RPLC methods, respectively), (c) spike-in 9-labeled internal standards (IS) during the sample preparation to control for extraction efficiency and evaluate LC-MS performance, (d) checking mass accuracy, retention time, and peak shape of IS in every samples, and (e) injection of a pool sample every 10 injections to control for signal deviation with time.

## 4.6 | Data processing and metabolite annotation

Data from each mode were independently analyzed using Progenesis Q1 software v2.3 (Nonlinear Dynamics). Metabolic features from blanks and that did not show sufficient linearity upon dilution were discarded ( $R^2 < .6$ ). Only metabolic features present in >33% of the samples were kept for further analysis, and missing values were imputed using minimum value imputation (Tyanova et al., 2016). MS signal drift with time was corrected for each metabolite using LOESS normalization method (locally estimated scatterplot smoothing Local Regression) on pool samples. Data from each mode were merged, and metabolites were formally identified by matching the fragmentation spectra and retention time to analytical-grade standards when possible or matching experimental MS/MS to fragmentation spectra in publicly available databases. We used the Metabolomics Standards Initiative (MSI) level of confidence to grade metabolite annotation confidence (level 1 - level 4) (Table S1). Level 1 represents formal identifications where the biological signal matches accurate mass, retention time, and fragmentation spectra of an authentic standard ran on the same platform. For level 2 identification, the biological signal matches accurate mass and fragmentation spectra available in METLIN database. Acylcarnitines, free fatty acids, and complex lipids do not necessarily all have MS/MS data in public databases but were annotated based on expected signature fragments. Level 3 represents putative identifications that are the most likely name based on previous knowledge of blood composition. Level 4 consists in unknown metabolites. After careful annotation of the metabolic features, a total of 770 metabolites were measured using our metabolite profiling platform. Metabolites were then categorized in classes and pathways based on the KEGG database where possible. Metabolite abundances (spectral counts) for all participants can be found in Table S3.

## 4.7 | Data analysis and visualization

### 4.7.1 | Heatmap and principal component analysis

Natural log metabolite abundances of all 770 detected metabolites (rows) in 268 individuals (columns) were plotted as a heatmap using

the R package “pheatmap” (v1.0.8). Principal component analysis was performed using the function `prcomp` in the base “stats” package in R (v3.3.3). Abundances were natural log-transformed and scaled by metabolite using the Z-score method (mean of zero and standard deviation of 1) prior to the analysis. In order to identify the 2 principal components (PCs) most strongly associated with age, we applied linear regression models between PC scores and age for each PC and selected the 2 PC with the smallest *p* values. The 4th and 5th PCs ordered by overall variation were selected.

#### 4.7.2 | Variance partition analysis

Variance partition was performed using the “variancePartition” package in R (v1.4.2). Age and BMI were considered continuous variables whereas sex, asthma status (GINA scoring) defined by a pulmonologist (Bousquet, 2000), smoking, and second-hand smoking exposure were considered categorical variables.

#### 4.7.3 | Clustering of metabolic aging trajectories

Fuzzy c-means clustering was performed using the R package “Mfuzz” (v2.20.0). Data inputted into the clustering algorithm were created by fitting a LOESS model of age in years (6 months to 82 years) vs. metabolite abundances using  $\text{span} = 0.75$  with R package “stats” (v3.3.3). Data were then scaled so that the mean equal 0 and the standard deviation equal 1 for each metabolite. We calculated the minimum centroid distance for a range of cluster numbers and chose six clusters as the optimal number of clusters.

#### 4.7.4 | Random forest prediction modeling

Two predictive models were generated using random forest (RF) algorithm (“RandomForest” package in R (v4.6)). The regression model used age as a continuous variable in adult population ( $\geq 16$  years old). Importance values were assessed (“importance = T”). The classification model used age as a discrete variable to distinguish adults ( $\geq 16$  years old) from pediatric subjects ( $< 16$  years old). Abundance profiles of all metabolite were natural log-transformed and adjusted for sex and BMI using a linear model prior to modeling. Significance was defined as the model importance metric greater than the absolute value of the most negative metabolite. Sub-sampling (“sampsiz” parameter) was utilized when performing the classification model on the adult group to account for the unbalanced size between pediatric and adult populations in the cohort.

#### 4.7.5 | Correlation network analysis

Pairwise Spearman's rank correlations were calculated using the R package “Hmisc” (v3.15-0), and weighted, undirected networks

were plotted with “igraph” (v0.7.1). Correlations with Bonferroni adjusted *p*-values below 0.01 were included and displayed via the Fruchterman–Reingold method. Only vertices with at least one connection were plotted. Nodes were color-coded by chemical classes, and their size represents the betweenness centrality.

#### 4.7.6 | Pathway enrichment analysis

Pathway enrichment analysis was performed using hypergeometric tests and the background of the detected metabolites in the study. *p* values were adjusted for multiple comparisons using Benjamini and Hochberg correction (Benjamini, 2010).

#### 4.7.7 | Twin metabolic variability

To evaluate metabolic variability between twins, we calculated Spearman's rank correlations using all the detected metabolites (R package “stats” (v 3.3.3)). We also calculated the Spearman's rank correlations between randomly selected, age-matched, and between randomly selected, sex- and age-matched pairs. Age matching was done by randomly selecting an individual within three years of age. We also included a set of randomly selected pairs without any matching restrictions. Singlets were not included in this analysis. For triplet participants, we randomly picked two sex-matched triplets.

#### ACKNOWLEDGMENTS

The authors would like to thank the participants for their willingness to contribute to scientific research, the SRI and Washington State Twin Registries, and the staff at the Sean N Parker Center for Allergy and Asthma Research at Stanford University, and the University of Washington and Asthma and Allergy Associates of Seattle who conducted the study.

#### CONFLICT OF INTEREST

MPS is a cofounder of Personalis, SensOmics, January, Filtricine, Qbio, and Akna. KCN has received funding from, is currently funded by, or cofounded the following: the NIH, Food Allergy Research & Education, End Allergies Together, Before Brands, Alladapt Immunotherapeutics, Adare Pharmaceuticals, AstraZeneca, Novartis, Genentech, Astellas, DBV Technologies, ForTra, Aimmune Therapeutics, Regeneron, Nestle, Sanofi, ImmuneWorks, Cour, Allergenics, Ukko, and the Environmental Protection Agency. The remaining authors declare no relevant competing interests.

#### AUTHORS' CONTRIBUTION

KN, MPS, and KC conceived and assisted with the design of the project. BB and KC wrote the manuscript with assistance from SA and KN. KC and BL-M performed the experiment with assistance from WZ. BB and KC performed data analysis with assistance from SA, GD, and MD. DT, BB, and OR assisted in the clinic.

## ORCID

Bryan J. Bunning  <https://orcid.org/0000-0003-3092-0038>

Sandra Andorf  <https://orcid.org/0000-0002-3093-2568>

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are shared as supplemental tables (Tables S2 and S3).

## REFERENCES

- Al-Dehaimi, A. W., Blumsohn, A., & Eastell, R. (1999). Serum galactosyl hydroxyllysine as a biochemical marker of bone resorption. *Clinical Chemistry*, 45(5), 676–681.
- Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., ... Shen-Orr, S. S. (2019). A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nature Medicine*, 25(3), 487–495. <https://doi.org/10.1038/s41591-019-0381-y>
- Azam, S., Hadi, N., Khan, N. U., & Hadi, S. M. (2003). Antioxidant and pro-oxidant properties of caffeine, theobromine and xanthine. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 9(9), BR325–R330.
- Barzilai, N., Huffman, D. M., Muzumdar, R. H., & Bartke, A. (2012). The critical role of metabolic pathways in aging. *Diabetes*, 61(6), 1315–1322. <https://doi.org/10.2337/db11-1300>
- Benjamini, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 72(4), 405–416. <https://doi.org/10.1111/j.1467-9868.2010.00746.x>
- Bousquet, J. (2000). Global Initiative for Asthma (GINA) and its objectives. *Clinical & Experimental Allergy*, 30(Suppl 1), 2–5. <https://doi.org/10.1046/j.1365-2222.2000.00088.x>
- Brodin, P., Jovic, V., Gao, T., Bhattacharya, S., Angel, C. J. L., Furman, D., ... Davis, M. M. (2015). Variation in the human immune system is largely driven by non-heritable influences. *Cell*, 160(1–2), 37–47. <https://doi.org/10.1016/j.cell.2014.12.020>
- Montoliu, I., Scherer, M., Beguelin, F., DaSilva, L., Mari, D., Salvioli, S., ... Collino, S. (2014). Serum profiling of healthy aging identifies phospho- and sphingolipid species as markers of human longevity. *Aging*, 6(1), 9–25. <https://doi.org/10.18632/aging.100630>
- Chak, C. M., Lacruz, M. E., Adam, J., Brandmaier, S., Covic, M., Huang, J., ... Wang-Sattler, R. (2019). Ageing investigation using two-time-point metabolomics data from KORA and CARLA studies. *Metabolites*, 9(3), 44. <https://doi.org/10.3390/metabo9030044>
- Chaleckis, R., Murakami, I., Takada, J., Kondoh, H., & Yanagida, M. (2016). Individual variability in human blood metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), 4252–4259. <https://doi.org/10.1073/pnas.1603023113>
- Cheung, P., Vallania, F., Warsinske, H. C., Donato, M., Schaffert, S., Chang, S. E., ... Kuo, A. J. (2018). Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. *Cell*, 173(6), 1385–1397.e14. <https://doi.org/10.1016/j.cell.2018.03.079>
- Collino, S., Montoliu, I., Martin, F.-P., Scherer, M., Mari, D., Salvioli, S., ... Rezzi, S. (2013). Metabolic signatures of extreme longevity in Northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. *PLoS ONE*, 8(3), e56564. <https://doi.org/10.1371/journal.pone.0056564>
- Contrepois, K., Jiang, L., & Snyder, M. (2015). Optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-mass spectrometry. *Molecular & Cellular Proteomics*, 14(6), 1684–1695. <https://doi.org/10.1074/mcp.M114.046508>
- Darst, B. F., Kosciak, R. L., Hogan, K. J., Johnson, S. C., & Engelman, C. D. (2019). Longitudinal plasma metabolomics of aging and sex. *Aging*, 11(4), 1262–1282. <https://doi.org/10.18632/aging.101837>
- Deelen, J., Kettunen, J., Fischer, K., van der Spek, A., Trompet, S., Kastenmüller, G., ... Slagboom, P. E. (2019). A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications*, 10(1), 3346. <https://doi.org/10.1038/s41467-019-11311-9>
- Feldman, H. A., Longcope, C., Derby, C. A., Johannes, C. B., Araujo, A. B., Coviello, A. D., ..., McKinlay, J. B. (2002). Age trends in the level of serum testosterone and other hormones in middle-aged men: Longitudinal results from the Massachusetts Male Aging Study. *Journal of Clinical Endocrinology and Metabolism*, 87(2), 589–598. <https://doi.org/10.1210/jcem.87.2.8201>
- Feng, X., & McDonald, J. M. (2011). Disorders of bone remodeling. *Annual Review of Pathology*, 6, 121–145. <https://doi.org/10.1146/annurev-pathol-011110-130203>
- Fulgoni, V. L., Keast, D. R., & Lieberman, H. R. (2015). Trends in intake and sources of caffeine in the diets of US adults: 2001–2010. *American Journal of Clinical Nutrition*, 101(5), 1081–1087. <https://doi.org/10.3945/ajcn.113.080077>
- Guasch-Ferré, M., Hruby, A., Toledo, E., Clish, C. B., Martínez-González, M. A., Salas-Salvadó, J., & Hu, F. B. (2016). Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. *Diabetes Care*, 39(5), 833–846. <https://doi.org/10.2337/dc15-2251>
- Hoffman, J. M., Lyu, Y., Pletcher, S. D., Promislow, D. E. L. (2017). Proteomics and metabolomics in ageing research: from biomarkers to systems biology. *Essays in Biochemistry*, 61(3), 379–388. <https://doi.org/10.1042/EBC20160083>
- James, M. J., Gibson, R. A., & Cleland, L. G. (2000). Dietary polyunsaturated fatty acids and inflammatory mediator production. *American Journal of Clinical Nutrition*, 71(Suppl. 1), 343S–348S. <https://doi.org/10.1093/ajcn/71.1.343s>
- Kettunen, J., Tukiainen, T., Sarin, A.-P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.-P., ..., Ripatti, S. (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genetics*, 44, 269–276. <https://doi.org/10.1038/ng.1073>
- Long, T., Hicks, M., Yu, H.-C., Biggs, W. H., Kirkness, E. F., Menni, C., ... Telenti, A. (2017). Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature Genetics*, 49(4), 568–578. <https://doi.org/10.1038/ng.3809>
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153(6), 1194–1217. <https://doi.org/10.1016/j.cell.2013.05.039>
- Menni, C., Kastenmüller, G., Petersen, A. K., Bell, J. T., Psatha, M., Tsai, P.-C., ... Valdes, A. M. (2013). Metabolomic markers reveal novel pathways of ageing and early development in human populations. *International Journal of Epidemiology*, 42(4), 1111–1119. <https://doi.org/10.1093/ije/dyt094>
- Nagpal, R., Mainali, R., Ahmadi, S., Wang, S., Singh, R., Kavanagh, K., ... Yadav, H. (2018). Gut microbiome and aging: Physiological and mechanistic insights. *Nutrition and Healthy Aging*, 4(4), 267–285. <https://doi.org/10.3233/NHA-170030>
- Nair, K. S., Rizza, R. A., O'Brien, P., Dhatariya, K., Short, K. R., Nehra, A., ... Jensen, M. D. (2006). DHEA in elderly women and DHEA or testosterone in elderly men. *New England Journal of Medicine*, 355(16), 1647–1659. <https://doi.org/10.1056/NEJMoa054629>
- Otsuka, R., Kato, Y., Imai, T., Ando, F., & Shimokata, H. (2013). Higher serum EPA or DHA, and lower ARA compositions with age independent fatty acid intake in Japanese aged 40 to 79. *Lipids*, 48(7), 719–727. <https://doi.org/10.1007/s11745-013-3763-9>
- Piening, B. D., Zhou, W., Contrepois, K., Röst, H., Gu Urban, G. J., Mishra, T., ... Snyder, M. P. (2018). Integrative personal omics profiles during

- periods of weight gain and loss. *Cell Systems*, 6(2), 157–170.e8. <https://doi.org/10.1016/j.cels.2017.12.013>
- Pitkänen, H. T., Oja, S. S., Kemppainen, K., Seppä, J. M., & Mero, A. A. (2003). Serum amino acid concentrations in aging men and women. *Amino Acids*, 24(4), 413–421. <https://doi.org/10.1007/s00726-002-0338-0>
- Rinschen, M. M., Ivanisevic, J., Giera, M., & Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology*, 20(6), 353–367. <https://doi.org/10.1038/s41580-019-0108-4>
- Rist, M. J., Roth, A., Frommherz, L., Weinert, C. H., Krüger, R., Merz, B., ... Watzl, B. (2017). Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS ONE*, 12(8), e0183228. <https://doi.org/10.1371/journal.pone.0183228>
- Schüssler-Fiorenza Rose, S. M., Contrepois, K., Moneghetti, K. J., Zhou, W., Mishra, T., Mataraso, S., ..., Snyder, M. P. (2019). A longitudinal big data approach for precision health. *Nature Medicine*, 25(5), 792–804. <https://doi.org/10.1038/s41591-019-0414-6>
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science & Technology*, 48(4), 2097–2098. <https://doi.org/10.1021/es5002105>
- Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., ... Soranzo, N. (2014). An atlas of genetic influences on human blood metabolites. *Nature Genetics*, 46(6), 543–550. <https://doi.org/10.1038/ng.2982>
- Shlipak, M. G., Katz, R., Kestenbaum, B., Fried, L. F., Newman, A. B., Siscovick, D. S., ... Sarnak, M. J. (2009). Rate of kidney function decline in older adults: A comparison using creatinine and cystatin C. *American Journal of Nephrology*, 30(3), 171–178. <https://doi.org/10.1159/000212381>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ..., Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tiao, J. Y.-H., Semmens, J. B., Masarei, J. R., & Lawrence-Brown, M. M. (2002). The effect of age on serum creatinine levels in an aging population: Relevance to vascular surgery. *Cardiovascular Surgery*, 10(5), 445–451. [https://doi.org/10.1016/S0967-2109\(02\)00056-X](https://doi.org/10.1016/S0967-2109(02)00056-X)
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., ... Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), 731–740. <https://doi.org/10.1038/nmeth.3901>
- Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., ... Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, 486(7402), 222–227. <https://doi.org/10.1038/nature11053>
- Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., ... Wang-Sattler, R. (2012). Human serum metabolic profiles are age dependent. *Aging Cell*, 11(6), 960–967. <https://doi.org/10.1111/j.1474-9726.2012.00865.x>
- Zhou W., Sailani M. R., Contrepois K., Zhou Y., Ahadi S., Leopold S. R., ... Snyder M.. Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature*. 2019; 569(7758), 663–671.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Bunning BJ, Contrepois K, Lee-McMullen B, et al. Global metabolic profiling to model biological processes of aging in twins. *Aging Cell*. 2020;19:e13073. <https://doi.org/10.1111/ace1.13073>