Global metabolomics revealed deviations from the metabolic aging clock in colorectal cancer patients

Authors: Long Zhang[1,2,3,11], Shaobo Mo[1,2,11], Xiurui Zhu[4,11], C. James Chou[5,11], Bo Jin[4], Zhi Han[5], James Schilling[6,7,8], Weili Liao[4], Sheeno Thyparambil[4], Ruben Y. Luo[5], John C. Whitin[5], Lu Tian[5], Seema Nagpal[5], Scott Ceresnak[5], Harvey J. Cohen[5], Doff B. McElhinney[5], Karl G. Sylvester[5], Yangming Gong[9], Chen Fu[9,10]*, Xuefeng B. Ling[5]*, Junjie Peng[1,2]*

Affiliations:

[1] Department of Colorectal Surgery, Fudan University Shanghai Cancer Center; Shanghai, China.

[2] Department of Oncology, Shanghai Medical College, Fudan University; Shanghai, China.

[3] Cancer Research Institute, Fudan University Shanghai Cancer Center; Shanghai, China.

[4] mProbe Inc.; Rockville, MD, USA.

[5] School of Medicine, Stanford University; Stanford, CA, USA.

[6] Shanghai Yunxiang Medical Technology Co., Ltd.; Shanghai, China.

[7] Tianjin Yunjian Medical Technology Co. Ltd.; Tianjin, China.

[8] Binhai Industrial Technology Research Institute, Zhejiang University; Tianjin, China.

[9] Shanghai Municipal Center for Disease Control and Prevention; Shanghai, China.

[10] Shanghai Clinical Research Center for Aging and Medicine; Shanghai, China.

[11]These authors contributed equally to this work.

*Corresponding authors

*Junjie Peng Email: pengjj@shca.org.cn

*Xuefeng B. Ling Email: bxling@stanford.edu

*Chen Fu Email: fuchen@scdc.sh.cn

E-mails of all the authors:

Long Zhang: longzhangcrc@yeah.net

Shaobo Mo: 18111230026@fudan.edu.cn

Xiurui Zhu: xiuruizhu.001@gmail.com

C. James Chou: cjchou@stanford.edu

Bo Jin: bo.jin@mprobe.com

Zhi Han: Zhihan01@stanford.edu

James Schilling: james.s@mprobe.com

Weili Liao: weililiao@mprobe.com

Sheeno Thyparambil: sheeno@mprobe.com

Ruben Y. Luo: rubenluo@stanford.edu

John C. Whitin: cuke@stanford.edu

Lu Tian: lutian@stanford.edu

Seema Nagpal: snagpal@stanford.edu

Scott Ceresnak: ceresnak@stanford.edu

Harvey Cohen: punko@stanford.edu

Doff McElhinney: doff@stanford.edu

Karl G. Sylvester: karls@stanford.edu

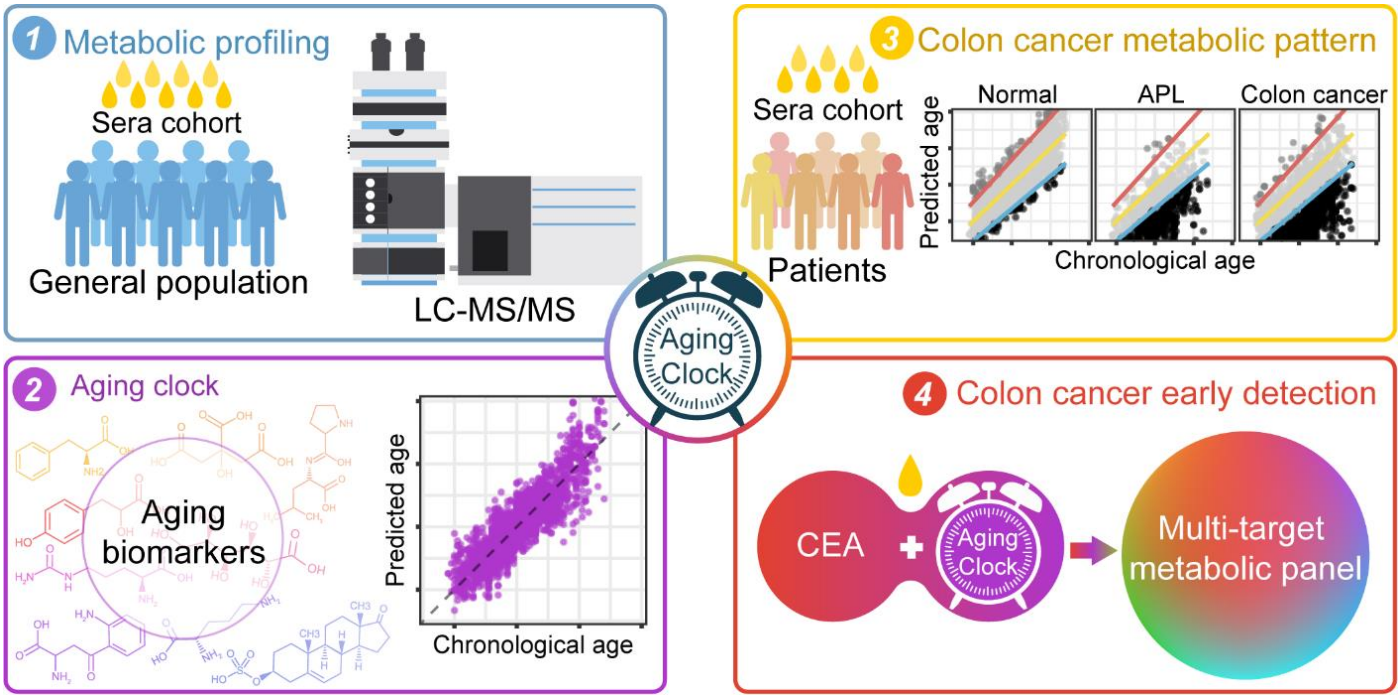
Yangming Gong: gongyangming@scdc.sh.cn

Chen Fu: fuchen@scdc.sh.cn

Xuefeng B. Ling: bxling@stanford.edu

Junjie Peng: pengjj@shca.org.cn

Graphical Abstract



Abstract:

Background: Markers of aging hold promise in the context of colorectal cancer (CRC) care. Utilizing high-resolution metabolomic profiling, we can unveil distinctive age-related patterns that have the potential to predict early CRC development. Our study aims to unearth a panel of aging markers and delve into the metabolomic alterations associated with aging and CRC.

Methods: We assembled a serum cohort comprising 5,649 individuals, consisting of 3,002 healthy volunteers, 715 patients diagnosed with colorectal advanced precancerous lesions (APL), and 1,932 CRC patients, to perform a comprehensive metabolomic analysis.

Results: We successfully identified unique age-associated patterns across 42 metabolic pathways. Moreover, we established a metabolic aging clock, comprising 9 key metabolites, using an elastic net regularized regression model that accurately estimates chronological age. Notably, we observed significant chronological disparities among the healthy population, APL patients, and CRC patients. By combining the analysis of circulative carcinoembryonic antigen levels with the categorization of individuals into the "hypo" metabolic aging subgroup, our blood test demonstrates the ability to detect APL and CRC with positive predictive values of 68.4% (64.3%, 72.2%) and 21.4% (17.8%, 25.9%), respectively.

Conclusions: This innovative approach utilizing our metabolic aging clock holds significant promise for accurately assessing biological age and enhancing our capacity to detect APL and CRC.

Keywords: Aging; Global metabolomics; Metabolic aging clock; Colorectal advanced precancerous lesions; Colorectal cancer

Running title: Metabolic deviation in colorectal cancer.

Introduction

Aging is an inevitable life-long decline in physiological functions and is the major risk factor for high impact chronic diseases such as cancers and cardiovascular diseases [1]. Aging involves extensive physiological changes and metabolic adaptations over decades[2]. Modern “omics” platforms, including genomic, transcriptomic,

proteomic, and metabolomic profiling assays, have provided new opportunities for the systematic and agnostic characterization of biological aging.

DNA methylation-based profiling, also known as “DNAm age”, is a powerful tool for predicting chronological age and assessing biological aging. It can be used across most tissues and cell types, and it incorporates composite clinical measures to capture risks for a wide range of age-related outcomes. Two of the most recent and promising DNAm age algorithms are “DNAm PhenoAge” [3] and “DNAm GrimAge” [4]. DNAm PhenoAge was developed to predict multifactorial phenotypic age, while DNAm GrimAge was developed to study aging and age-related traits. Both algorithms have been shown to be strongly associated with mortality and other age-related health outcomes. DNAm age is a valuable tool for researchers and clinicians alike. It can be used to study the aging process, identify individuals at risk for age-related diseases, and develop personalized interventions to promote longevity.

Metabolic syndrome [5], a cluster of metabolic abnormalities, is age-related and regulated by key metabolic proteins such as mechanistic target of rapamycin (mTOR), AMP-activated protein kinase (AMPK), and insulin/insulin growth factor (IGF) [6, 7]. Dysregulated metabolic control is a long-term cause of aging and increases the risk of chronic diseases. Metabolomic age models [8, 9], developed with unprecedented high-resolution metabolome coverage, assess biological age. Metabolomic [9] and epigenetic [10] aging clocks use different biomarkers, but both correlate with chronological age.

Age is the strongest risk factor for cancers, including advanced precancerous lesions (APL) and colorectal cancer (CRC). As people age, their risk of developing advanced polyps and CRC increases [11, 12]. In this study, we used a deep

metabolomic analysis of over 3,000 healthy individuals to investigate how the metabolome changes with age. We hypothesized that high-resolution metabolomic profiling could reveal unique age-associated patterns that could precisely predict chronological age. We also hypothesized that a metabolomic aging blood test could have clinical applications, such as assessing aging and detecting CRC early.

Results

Shanghai General Population and Cancer Center CRC Cohort Study

The study design and methods are outlined in Figure 1. We collected pretreatment serum samples (Supplementary Material 1) from 3,002 healthy individuals from the Shanghai Centers for Disease Control and Prevention (CDC), 715 patients with advanced precancerous lesions (APL), and 1,932 CRC patients without known contribution from germline causes or significant family history of cancer or inflammatory bowel disease (i.e., patients whose CRC is not thought to be caused by a genetic mutation or a strong family history of cancer). Demographic data are summarized in Table 1.

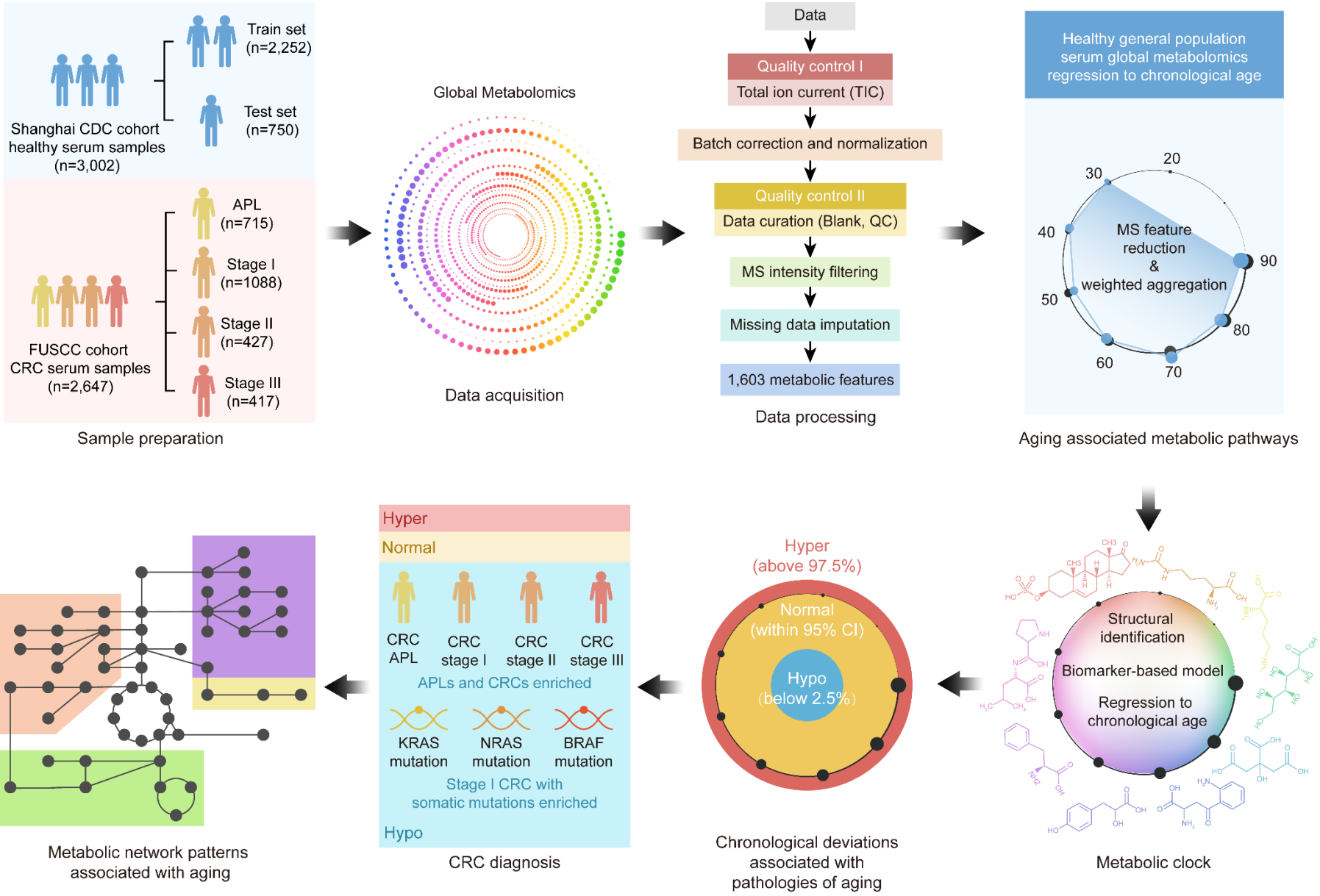

Train set
(n=2,252)
Test set
(n=750)
Shanghai CDC cohort
healthy serum samples
(n=3,002)
APL
(n=715)
Stage I
(n=1088)
Stage II
(n=427)
Stage III
(n=417)
FUSCC cohort
CRC serum samples
(n=2,647)
Sample preparation
Global Metabolomics
Data acquisition
Data
Quality control I
Total ion current (TIC)
Batch correction and normalization
Quality control II
Data curation (Blank, QC)
MS intensity filtering
Missing data imputation
1,603 metabolic features
Data processing
Healthy general population
serum global metabolomics
regression to chronological age
20
30
40
50
60
70
80
90
MS feature
reduction
&
weighted aggregation
Aging associated metabolic pathways
Structural
identification
Biomarker-based model
Regression to
chronological age
Metabolic clock
Hyper
(above 97.5%)
Normal
(within 95% CI)
Hypo
(below 2.5%)
Chronological deviations
associated with
pathologies of aging
Hyper
Normal
CRC
APL
CRC
stage I
CRC
stage II
CRC
stage III
APLs and CRCs enriched
KRAS
mutation
NRAS
mutation
BRAF
mutation
Stage I CRC with
somatic mutations enriched
Hypo
CRC diagnosis
Metabolic network patterns
associated with aging

Figure 1: Metabolomics data workflow. In this study, 3,002 healthy volunteers and 2,647 patients with colorectal advanced precancerous lesions (APL) or colorectal cancer (CRC) were enrolled, and their serum samples were collected. High-resolution mass spectrometry was used for serum global metabolomics data acquisition, and 1,603 metabolic features were identified after data processing. After metabolic feature analysis, aging associated metabolic pathways were found to regress to chronological age. Further exploration of the metabolic features identified nine metabolites as the metabolic aging clock. Comparing actual chronological age and metabolic aging clock predicted age of healthy people using a 95% confidence interval (CI, 2.5%~97.5%) defined hyper, normal and hypo subgroups. Compared with healthy people, APL and CRC patients usually bearing somatic genetic mutations significantly fell into the hypo subgroup. At last, the underlying metabolic patterns associated with aging were revealed.

120

Unique Metabolomic Patterns Predict Chronological Age

Using high-resolution mass spectrometry (Supplementary Material 2) to profile blood metabolomes, we identified 1,603 metabolic features. Of these, 157 were associated with aging (Pearson correlation, $|r| \geq 0.3$). We aggregated the aging associated features into KEGG pathways and calculated the value of each pathway as the weighted sum of the normalized measurement values of aging associated metabolites on the pathway divided by the number of mapped metabolites (Supplementary Material 3). Using an elastic net approach, we regressed these 59 pathways on chronological age and found that 42 of them contributed to the multivariate regression with positive importance scores.

The 42 pathways collectively achieved improved regression on chronological age (Figure 2A, Supplementary Figure 1): training, $r = 0.88$, 95%CI 0.88-0.89, $slope = 0.98$, 95%CI 0.95-1.00; testing, $r = 0.81$, 95%CI 0.80-0.83, $slope = 0.96$, 95%CI 0.90-1.01. The top 10 metabolic pathways, ranked by Pearson correlation to chronological age, were steroid hormone biosynthesis, bile secretion, ABC transporters, histidine metabolism, metabolism of xenobiotics by cytochrome P450, riboflavin metabolism, chemical carcinogenesis, phenylalanine metabolism, citrate cycle, and pyrimidine metabolism.

We performed the pathway-based multivariate regression analysis with men, and women separately. Using the same statistical pipeline for the general population (Figure 2A, 42 pathways), we identified 70 and 48 metabolic pathways (Figure 2B/2C) for the multivariate analyses of men and women populations, respectively. The Pearson

correlation coefficients were $0.90/0.91$ (men/women) in training and $0.78/$ 0.87 (men/women) in testing respectively (Figure 2B/2C). The Pearson correlation coefficients among these three populations, all/men/wo were statistically significant (p-value = 0.0001). The rankings of the importance scores of the significant aging-correlating pathway features were similar among all, men, and women (Figure 2A/2B/2C). For example, the steroid hormone biosynthesis pathway ranked top 1 in both the all's and men's models and ranked top 3 in the women's model.

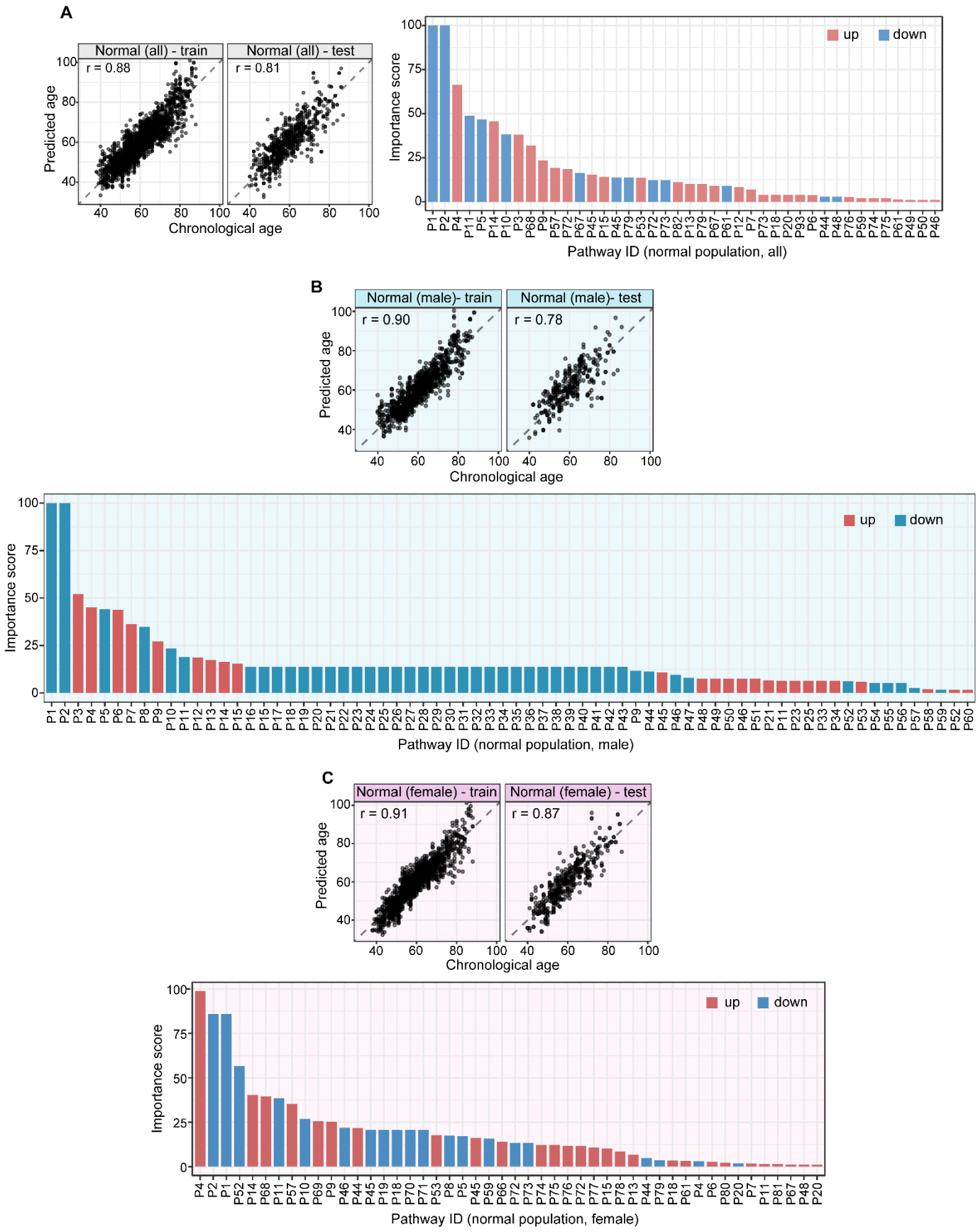

A
Normal (all) - train
r = 0.88
Normal (all) - test
r = 0.81
Predicted age
Chronological age
Importance score
up
down
Pathway ID (normal population, all)
B
Normal (male)- train
r = 0.90
Normal (male)- test
r = 0.78
Predicted age
Chronological age
Importance score
up
down
Pathway ID (normal population, male)
C
Normal (female) - train
r = 0.91
Normal (female) - test
r = 0.87
Predicted age
Chronological age
Importance score
up
down
Pathway ID (normal population, female)

Figure 2. A pathway based metabolic aging clock from the general population.

Importance scores of underlying pathways from untargeted metabolic profiling (training set:75%; testing set:25%). **(A).** Prediction of metabolic ages from all normal population (both sexes). **(B).** Prediction of metabolic ages from male normal population. **(C).** Prediction of metabolic ages from female normal population. The pathway IDs in the figure are as follows: P1, Steroid hormone biosynthesis; P2, Bile secretion; P3, Phenylalanine metabolism; P4, ABC transporters; P5, Metabolism of xenobiotics by cytochrome P450; P6, 2-Oxocarboxylic acid metabolism; P7, Lysine degradation; P8, Biosynthesis of amino acids; P9, Pyrimidine metabolism; P10, Chemical carcinogenesis; P11, Histidine metabolism; P12, Amino sugar and nucleotide sugar metabolism; P13, Protein digestion and absorption; P14, Riboflavin metabolism; P15, Pentose and glucuronate interconversions; P16, Arginine biosynthesis; P17, Ascorbate and aldarate metabolism; P18, Glyoxylate and dicarboxylate metabolism; P19, Carbon metabolism; P20, Taste transduction; P21, Ferroptosis; P22, Proximal tubule bicarbonate reclamation; P23, D-Glutamine and D-glutamate metabolism; P24, Neomycin, kanamycin and gentamicin biosynthesis; P25, Nitrogen metabolism; P26, FoxO signaling pathway; P27, Phospholipase D signaling pathway; P28, Gap junction; P29, Circadian entrainment; P30, Long-term potentiation; P31, Synaptic vesicle cycle; P32, Retrograde endocannabinoid signaling; P33, Glutamatergic synapse; P34, GABAergic synapse; P35, Long-term depression; P36, Amyotrophic lateral sclerosis; P37, Huntington disease; P38, Spinocerebellar ataxia; P39, Cocaine addiction; P40, Amphetamine addiction; P41, Nicotine addiction; P42, Alcoholism; P43, Sulfur relay system; P44, Valine, leucine and isoleucine degradation; P45, Glutathione metabolism; P46, Butanoate metabolism; P47, Antifolate resistance; P48, Valine, leucine and isoleucine biosynthesis; P49, Fatty acid degradation; P50, Vitamin B6 metabolism; P51, Pantothenate and CoA biosynthesis;

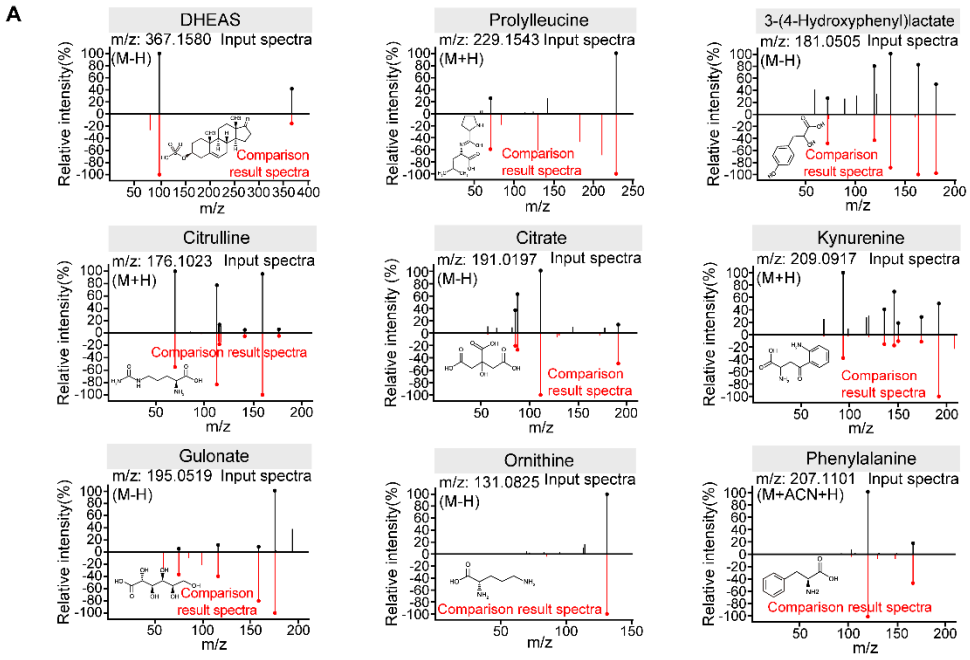
P52, Porphyrin and chlorophyll metabolism; P53, Adrenergic signaling in cardiomyocytes; P54, Taurine and hypotaurine metabolism; P55, Sulfur metabolism; P56, Neuroactive ligand-receptor interaction; P57, Ubiquinone and other terpenoid-quinone biosynthesis; P58, Pentose phosphate pathway; P59, Tryptophan metabolism; P60, beta-Alanine metabolism; P61, Arginine and proline metabolism; P62, Central carbon metabolism in cancer; P63, Aminoacyl-tRNA biosynthesis; P64, Mineral absorption; P65, Glycolysis / Gluconeogenesis; P66, Pyruvate metabolism; P67, Glycine, serine and threonine metabolism; P68, Citrate cycle (TCA cycle); P69, Insulin resistance; P70, Sphingolipid metabolism; P71, Sphingolipid signaling pathway; P72, Nicotinate and nicotinamide metabolism; P73, Alanine, aspartate and glutamate metabolism; P74, Serotonergic synapse; P75, Drug metabolism – cytochrome P450; P76, D-Arginine and D-ornithine metabolism; P77, Purine metabolism; P78, Biotin metabolism; P79, Cysteine and methionine metabolism; P80, Phenylalanine, tyrosine and tryptophan biosynthesis; P81, Steroid biosynthesis; P82, Dopaminergic synapse; P83, Galactose metabolism; P84, Fructose and mannose metabolism; P85, Caffeine metabolism; P86, Drug metabolism - other enzymes; P87, Phosphonate and phosphinate metabolism; P88, Glycosaminoglycan biosynthesis - heparan sulfate / heparin; P89, Cholinergic synapse; P90, Folate biosynthesis; P91, Propanoate metabolism; P92, Primary bile acid biosynthesis; P93, Glucagon signaling pathway; P94, Tyrosine metabolism.

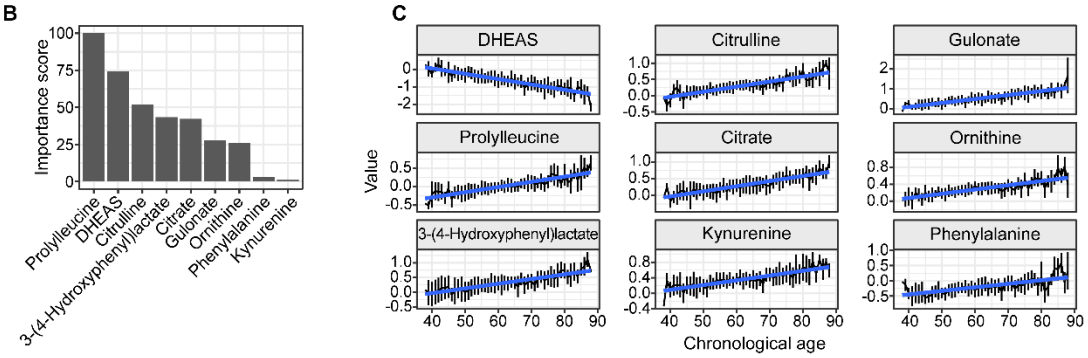
151

152 **Nine-Metabolite Metabolic Aging Clock**

153 Linear modeling links metabolic pathways to aging. We selected from the
154 significant aging-associated metabolic features to identify a panel of metabolite
155 biomarkers, called the "metabolic aging clock," to assess aging. The nine metabolites in

the metabolic aging clock were identified (Supplementary Material 4) using a combination of level 1 compound identification and multivariate regression (Supplementary Figure 3). The nine metabolites are shown in Figure 3A. The results of both the multivariate analysis (importance scores, IS, Figure 3B) and univariate analysis (Pearson correlation coefficient, r, Figure 3C) are summarized in Figure 3D. The metabolic aging clock achieved improved regression on chronological age, with a Pearson correlation coefficient r 0.95 (95%CI 0.946-0.954) and $slope$ 1.00 (95%CI 0.97-1.03) in the training, and r 0.87, 95%CI 0.85-0.89, $slope$ 0.96, 95%CI 0.90-1.03 in the testing set (Figure 3E). The results were similar for all subjects, men, and women (Figure 3E-G): the Pearson correlation coefficient (r) $0.95/0.97/0.97$ (all/men/women, *p*-value, 0.24) in training and $0.87/0.81/0.89$ (all/men/women, *p*-value < 0.0001) in testing respectively.





D

Metabolite	KEGG/HMDB entry	R	IS
Prolylleucine	HMDB0253028	0.534	100
Dehydroepiandrosterone sulfate (DHEAS)	C04555	-0.599	74.30
Citrulline	C00327	0.586	52.06
3-(4-hydroxyphenyl) lactate	C00642	0.443	43.37
Citrate	C00157	0.518	42.27
Gulonate	C00800	0.571	27.85
Ornithine	C00077	0.499	25.86
Phenylalanine	C00079	0.414	2.92
Kynurenine	C00328	0.442	1.04

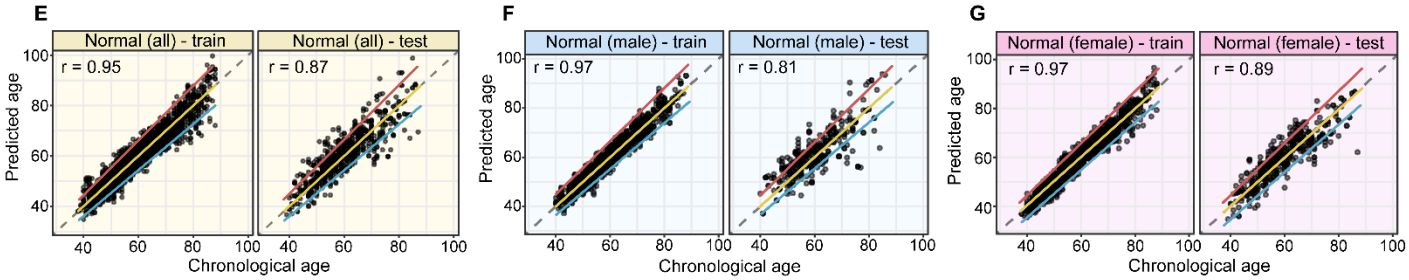


Figure 3. Structural identification of aging metabolite biomarkers and nine-metabolite-based metabolic aging clock. (A). Confirmation of the metabolites predicting chronological age by standard compounds. **(B).** Importance scores of the identified biomarkers in a biomarker-based metabolic aging clock. **(C).** Univariate trajectories of aging biomarkers as a function of the chronological age. **(D).** Nine compound biomarkers ranked by the importance scores in the aging clock model. KEGG: Kyoto Encyclopedia of Genes and Genomes; HMDB; Human Metabolome Database; R: Pearson Correlation Coefficient for regressing on the chronological age; IS: importance score in the multivariate model. **(E-G).** Prediction of metabolic ages in the training (75%) and testing (25%) sets from all the healthy general population (both sexes), from the male healthy population, and from the female healthy population.

Hypo-aging phenotypes in CRC patients

We used a metabolic clock to assess the metabolic ages of healthy people. We defined the "Δ metabolic aging" as the difference between the predicted and actual chronological ages. We used a 95% confidence interval (CI, 2.5%~97.5%) to define the "normal" (within the 95% CI), "hyper" (above the 97.5%), and "hypo" (below the 2.5%) Δ metabolic aging subgroup membership.

Compared to healthy people, individuals with "hypo" membership (Figure 4A) were more likely to have APL or CRC, regardless of the specific CI thresholds used to define the "normal", "hyper", and "hypo" Δ metabolic aging subgroups (Figure 4B, 4D-E, 4G-H). This suggests that a "hypo" metabolic aging phenotype is associated with an increased risk of APL and CRC (Supplementary Material 5, Supplementary Table 1, 2, and 3).

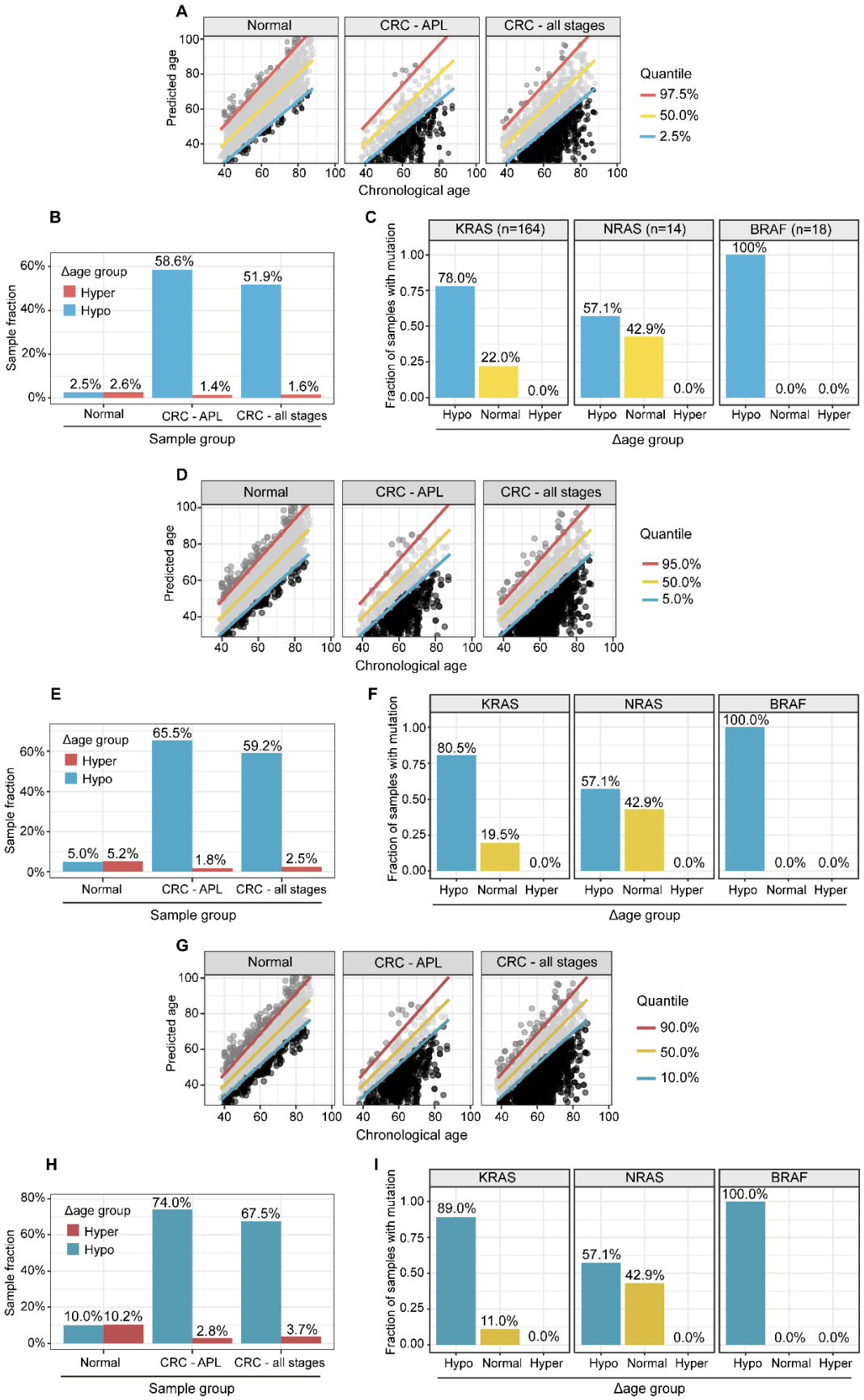

A
Normal
CRC - APL
CRC - all stages
Predicted age
Chronological age
Quantile
97.5%
50.0%
2.5%
B
Δage group
Hyper
Hypo
Sample fraction
58.6%
51.9%
2.5%
2.6%
1.4%
1.6%
Normal
CRC - APL
CRC - all stages
Sample group
C
KRAS (n=164)
NRAS (n=14)
BRAF (n=18)
Fraction of samples with mutation
78.0%
22.0%
0.0%
57.1%
42.9%
100%
Hypo
Normal
Hyper
Δage group
D
Quantile
95.0%
50.0%
5.0%
E
65.5%
59.2%
5.0%
5.2%
1.8%
2.5%
F
KRAS
NRAS
BRAF
80.5%
19.5%
100.0%
G
Quantile
90.0%
50.0%
10.0%
H
74.0%
67.5%
10.0%
10.2%
2.8%
3.7%
I
KRAS
NRAS
BRAF
89.0%
11.0%
100.0%

Figure 4. APL and CRC (all stages) samples have significantly higher fractions, and stage I CRC patients with gene mutations are more likely to have "hypo" metabolic age. (A, D, G). XY plot of the prediction of metabolic ages as a function of the chronological age in the normal general and CRC populations. (**B, E, H).** Sample fractions of the total in the hypo or hyper metabolic age subgroups. **(C, F, I).** KRAS, NRAS and BRAF mutations were significantly enriched in the hypo Δ age group ($p < 0.01$). The percentages were fractions of stage I CRC samples with corresponding mutations. **(A-C):** Quantile 2.5%, 50%, 97.5%. **(D-F):** Quantile 5%, 50%, 95%. **(G-I):** Quantile 10%, 50%, 90%.

182

183 **Hypo-Aging Phenotypes and Their Potential Clinical Utility for Improving**
184 **Colorectal Cancer Diagnosis**

185 The prevalence of APL and CRC in the general population is 7.6% and 0.7%
186 [13], respectively. Because there were many more APLs and CRCs in the hypo Δ
187 metabolic aging subgroup, we investigated whether the metabolic aging clock could be
188 used to detect CRC.

189 Specifically, we used hypo Δ metabolic aging subgroup membership to diagnose
190 CRC status. After adjusting for the true incidence rate of CRC in the general population,
191 the positive predictive values (PPVs, Table 2, Supplementary Table 2C) for APL and all
192 CRC stages were 65.5% (62.3-68.5%) and 12.7% (10.0-15.9%), respectively.

193

Although serum carcinoembryonic antigen (CEA) does not have sufficient sensitivity or specificity to diagnose CRC (Table 2 PPV: APL, 5.2% (4.6-5.7%); CRC, 0.4% (0.3-0.7%)), it is still considered the most important biomarker for detecting CRC.

By removing samples with normal CEA measurements from the positives predicted by the metabolic clock classifier, we created a multi-target panel (nine metabolites plus CEA) that improved the PPVs to 68.4% ((64.3%, 72.2%), $p=1.3x10^{-5}$) for APL, and 21.4% (17.8%-25.9%), $p= 1.2x10^{-10}$) for CRC. This suggests that CEA and the metabolic clock work together to improve APL/CRC diagnosis.

Metabolic Aging and CRC Mutation Profiles

To study the genomic mutation patterns of colorectal cancers (CRCs) in the metabolic Δ aging subgroups, we profiled 412 samples from patients with stage I CRC (Supplementary Table 3A). Among the 164/412 (39.8%) patients with KRAS, 14/412 (3.4%) with NRAS, and 18/412 (4.4%) with BRAF mutations/stage I CRCs (Supplementary Table 3B), 78.0% (71.3%, 84.2%) KRAS, 57.1% (28.8%, 85.7%) NRAS, and 100.0% (100.0%, 100.0%) BRAF mutants were found to be in the "hypo" metabolic age group (Figure 4C, 4F, and 4I). For reference, 66.1% of stage I CRC patients fell into the "hypo" subgroup (Figure 5).

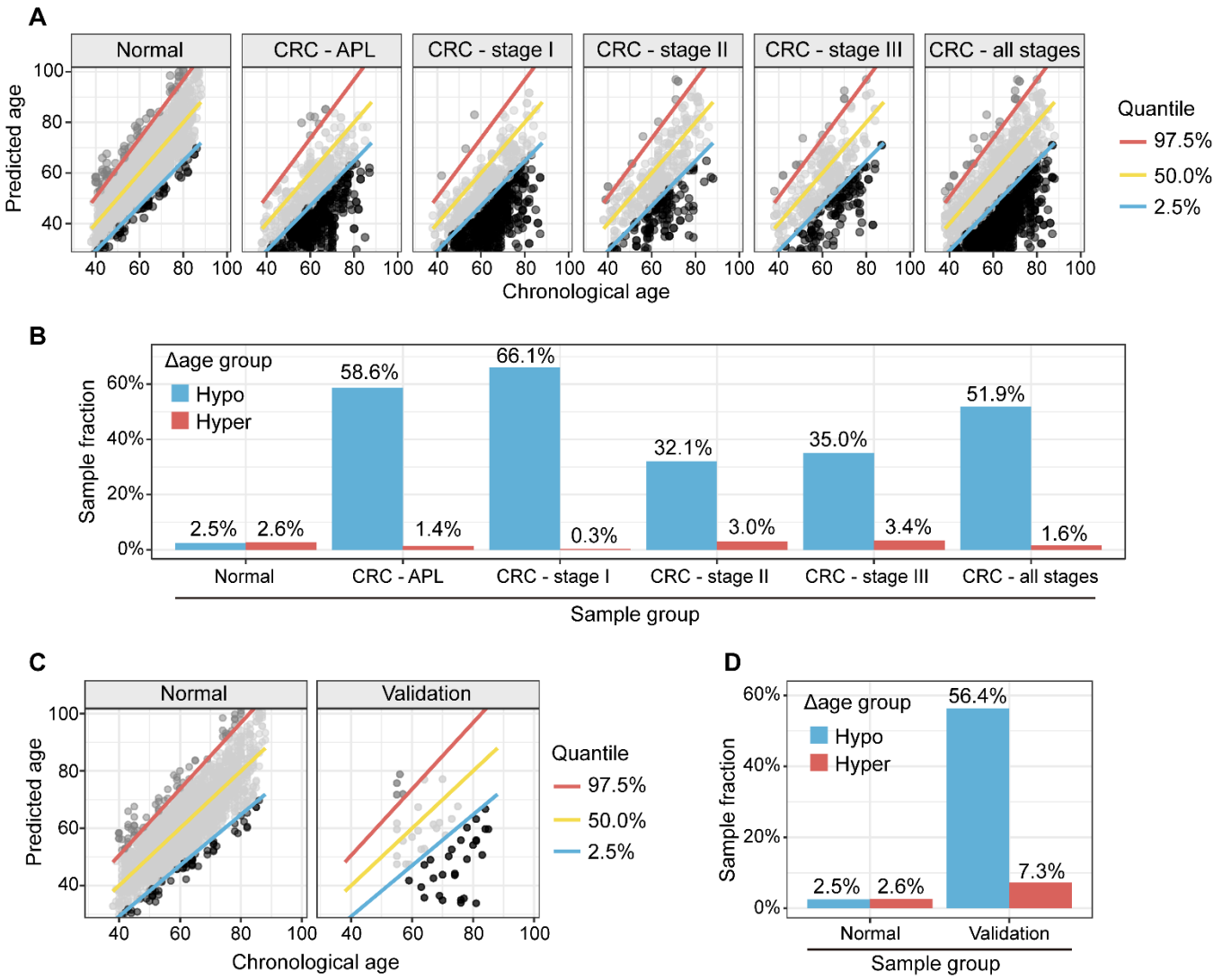


Figure 5. Metabolic clock analysis of APL and stage I CRC subjects. (A). XY plot of the predicted metabolic ages as a function of chronological ages. Samples below the 2.5% quantile line were defined as hypo Δ metabolic age subgroup and those above the 97.5% one as hyper Δ metabolic age subgroup. **(B).** Enrichment of APL and CRC stage I subjects in Δ metabolic age "hypo" subgroup. **(C).** XY plot of the prediction of metabolic ages as a function of the chronological age in the normal general and CRC populations. **(D).** Sample fractions of the total in the hypo or hyper Δ metabolic age subgroups.

Discussion

In this study, we used high-resolution mass spectrometry to identify key metabolic changes that correlate with chronological age in a healthy general population. We also developed a metabolic aging clock, a predictive model based on nine blood metabolites, to depict the age clock in the general population and in patients with colorectal cancer.

Previous studies have shown that certain biomarkers in our metabolic aging clock are associated with the aging process. Levels of kynurenine and phenylalanine increase with age, while levels of dehydroepiandrosterone (DHEAS) sulfate decrease. Kynurenine is produced by indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase from tryptophan. High levels of circulating kynurenine are thought to be a primary driver of aging [14-18], linked to increased frailty and mortality in humans. DHEAS peaks around age 20 and then gradually declines over time. By age 70, DHEA-S levels are about 20-30% lower than in younger adults [19-22]. Recently, researchers have reported that circulating phenylalanine also increases with age and is closely related to heart aging[23]. Older people also have a slower plasma clearance rate of phenylalanine, resulting in age-dependent increases [24]. Metabolites such as citrulline and ornithine are involved in the L-arginine/nitric oxide pathway and are thought to have anti-aging effects. The upregulation of circulating citrulline and ornithine could be a homeostatic response to the aging vesicular system in healthy individuals. Citrate, 3-(4-hdyroxyphenyl) lactate, and gulonate are closely related to cellular energy metabolism. An accumulation of these metabolites could indicate a shift in cellular metabolism between either glycolysis or mitochondrial respiration. Prolylleucine is associated with

muscle tone, type 2 diabetes, and insulin resistance, conditions that are highly associated with age [25, 26]. Prolylleucine levels are upregulated in males with insulin resistance and are significantly upregulated in people with type 2 diabetes, regardless of sex [25]. Prolylleucine has been proposed as a biomarker for type 2 diabetes [25].

Our findings that patients with CRC have a hypo metabolic age are consistent with a recent study of the PhenoAge clock (CpG markers: n=513), which showed a similar hypo-aging trend among high-risk CRC patients [27]. To develop a binary classifier for CRC assessment, we applied a random forest method to nine metabolite aging biomarkers. This improved the performance of the metabolic aging clock predictor of CRC status modestly: APL, 65.5% (62.3%, 68.5%); CRC, 12.7% (10.0%, 15.9%) (Table 2). The nine metabolic aging clock markers were originally discovered to regress to chronological age, but they can also be used directly with a more classical approach (a random forest method) to train a cancer binary classifier. This provides direct evidence to support the application of the metabolic aging clock in cancer assessment.

In this study, we validated our hypothesis that the metabolic aging clock and its hypo-aging membership could improve the early diagnosis of colorectal cancer (CRC). Our metabolic clock panel results significantly outperformed previous findings, with a much improved positive predictive value (PPV) for APL (65.5%) and CRC (12.7%) (Table 2) than Cologuard^{TM} [18] (20.0% for APL and 3.72% for CRC) and Septin 9 methylation tests (9.5% for APL and 2.3% for CRC) [28]. We further demonstrated that the CRC marker CEA could work together with our aging clock to improve the PPVs to identify APL (68.4%) and all stages of CRC cases (21.4%) (Table 2). Therefore, our models achieved higher PPV values to identify APL and stage I CRC subjects than

current clinically available diagnostic methods, using either the metabolic aging clock panel or the panel plus CEA. A good disease marker usually becomes more relevant with the severity of the disease, as it should accurately measure the presence and progression of the disease. However, our predictive performance with precancerous lesions and different CRC stages is counterintuitive. We hypothesize that this may be due to the different mechanisms of action between tumor genesis and later tumor progression. Tumor genesis is the process by which a normal cell transforms into a cancer cell, while tumor progression is the process by which cancer cells grow and spread throughout the body. Future research is needed to address this.

Clinically, KRAS and BRAF mutations are associated with a poor prognosis [29, 30]. Patients with CRC and BRAF mutations do not respond to cetuximab, and all but one patient with any of the three mutations did not respond [31]. Patients with any of the three mutations had a poor response rate (7.1%) and reduced survival (progression-free survival = 8 months) compared with wild-type counterparts (74.4% and 11.6 months). Our study showed that hypo-aging individuals were also highly enriched in the stage I CRC patient population with KRAS and BRAF mutations. A hypo-aging phenotype is typically associated with a less differentiated CRC phenotype and is usually less responsive to chemotherapeutic agents [32]. Epigenetic data suggests that decelerated epigenetic aging is associated with a poorer prognosis and lower overall survival rate in CRC [33]. Our study bridges the gap between clinical observations and epigenetic studies of CRC patients with KRAS and BRAF mutations in stage I CRC through a metabolic aging clock, illustrating a spectrum of malignancy with metabolic aging deviations in this stage.

Our study has several potential limitations:1. Enrolled patients were not required to take a germline mutation test, so it is possible that a small number of patients without a family history of CRC had germline mutations; 2. Our metabolic clock analytics for early CRC detection may have been confounded by pre-analytic variables and cohort differences in sex and age, which differed between the general population and CRC cohorts. A stronger single-site study design would help rule out the possibility of systematic confounding related to differences in blood collection and processing. In a subset of our cohort enrolled at Shanghai CDC, identical blood collection was performed in healthy controls and CRC patients (n=55) who were identified as part of the CDC screening. Similar hypo-aging membership patterns were observed in this subgroup of CRC patients (Figure 5C-5D), which supports the validity of our CRC results (Figure 4A-4B) from Fudan University Shanghai Cancer Center. In addition, we built 100 random models (Supplementary Material 6) using the statistical pipeline already established and compare if the deviation of the CRC/APL cohort from the CI derived from the general population is immanent/systematic (=bias) or specific for the age-related signature. Supplementary Figure 4A/B showed that our results are biologically meaningful and statistically significant. Our findings are unlikely to be due to technical bias; 3. This study is not designed to test the hypothesis that KRAS, NRAS, and BRAF mutations in other tumors cause hypometabolic age, or vice versa. Future studies with multi-cancer detection cohorts that include both pretreatment liquid and tissue biopsy samples could test these hypotheses, but they are beyond the scope of this study and is a limitation of the current study. 4. This study cannot test whether any metabolic age biomarkers change specifically in colorectal cancer (CRC). Although we

plan to explore the metabolic aging clock's clinical utility in other high-impact diseases, including other cancers, the generalizability of this approach needs to be validated with additional independent cohorts that can demonstrate minimal false positives and localization, evaluate the implementation and real-world performance of the test in clinical practice, confirm the results in a population with no known diagnosis, and validate the clinical utility in a high-risk population.

Conclusions

Our global metabolomic analysis revealed high-resolution metabolomic pattern changes (Figure 6) associated with aging progression and colorectal cancer (CRC) status. Our findings could lead to new approaches to longevity medicine and early detection of CRC, but further validation is needed in large, blinded clinical trials.

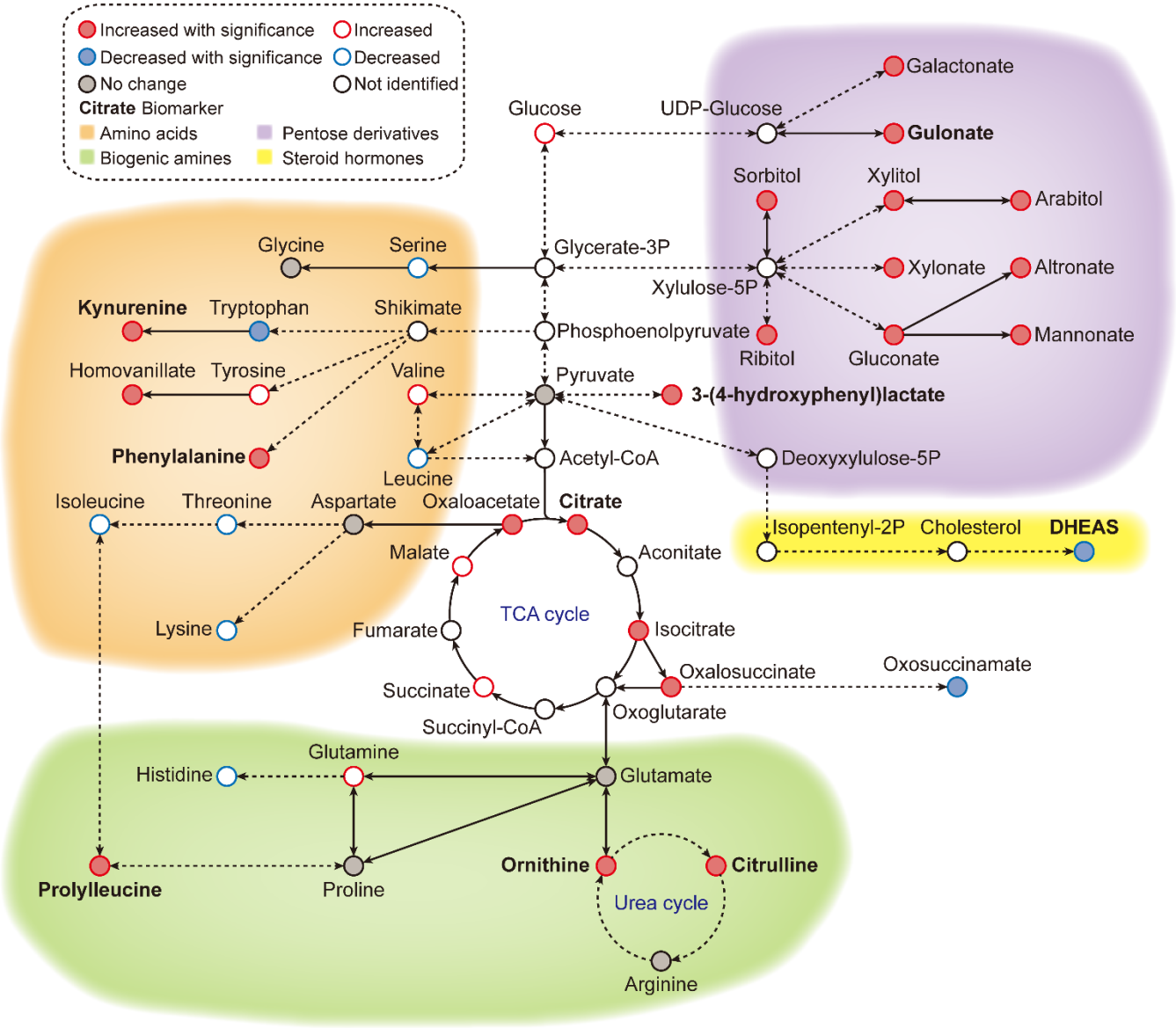


Figure 6. Metabolic overview of the reference ageotype. With global metabolomics, many metabolites were quantified and identified, forming a metabolic network of 4 clusters: amino acid metabolism (orange), biogenic amine metabolism (green), pentose/glucuronate conversions (purple) and steroid hormone biosynthesis (yellow). The names of the discovered biomarkers were highlighted in bold. Correlation between metabolites and aging was visualized with colored edges and fills in the nodes.

314

Methods

Study design and ethical approval

This study (Figure 1) was approved by the Shanghai Municipal Center for Disease Control and Prevention Ethical Review Committee (No. 2019-4) and the Ethical Committee and Institutional Review Board of Fudan University Shanghai Cancer Center (No. 1902197-15).

Healthy general population subjects in the Shanghai CDC cohort

To be eligible for the study, participants had to be at least 18 years old, not taking any medication for high-impact chronic diseases (such as cardiovascular diseases or diabetes mellitus) and have no history of tumors or cancers. Participants were followed for three years to identify any new cancer lesions or chronic diseases, and those who developed these conditions were excluded from the study.

Sporadic colorectal cancer (CRC) patients in Fudan University Shanghai Cancer Center cohort

We excluded patients with Lynch syndrome and FAP, which account for 5-7% of all colorectal cases and are mainly characterized by early-onset colorectal cancer and multiple polyps. These conditions may result in unique aging features. Our CRC subjects were defined as those with cancers that arise from the colorectum without known contribution from germline causes, a significant family history of cancer, or inflammatory bowel disease. We constructed the cohort by screening cases to exclude patients with common familial colorectal cancer according to their family history and clinical profile. For example, we excluded patients with Lynch syndrome according to the Amsterdam II criteria. Thus, our study focused on the aging characteristics of

sporadic colorectal cancer. We also excluded patients who were taking any medication for high-impact chronic diseases (such as cardiovascular diseases or diabetes mellitus). The collected samples were derived from the Department of Biobank, Fudan University Shanghai Cancer Center.

Sample preparation

Our cohort sera were collected from cancer patients before chemotherapy or radiotherapy administration. We used 3,002 serum samples (Supplementary Material 1) from healthy general population subjects in the Shanghai CDC cohort, 715 serum samples from patients with advanced precancerous lesions (APL), and 1,932 serum samples from CRC patients from Fudan University Shanghai Cancer Center in this study, after excluding ineligible participants. Demographic data are summarized in Table 1. We collected whole blood samples from patients and generated sera following the standard operating procedure (SOP) described in Supplementary Material 1 [34].

MS acquisition, QA/QC, annotation, structural identification

The MS analytic pipeline for data acquisition, QA/QC, annotation, and structure identification was described in detail in Supplementary Material 2 [35-61].

Identification of age associated metabolic pathways.

We were among the first groups to propose a pathway-based computational methodology for chronological event prediction with global metabolomics [62]. Detailed analyzing method was described in Supplementary Material 3 [47-65]. We provide a Supplementary Material 7 (Supplementary Table 5) describing aging associated KEGG metabolic pathways and their associated mapped metabolomic features.

Construction of a metabolic aging clock with nine compounds

Through an elastic net regularized regression ($\alpha = 0.125$, and $\lambda = 0.129$), a metabolic aging clock was trained with the metabolite biomarker candidates. Detailed analyzing method was described in Supplementary Material 4 [42, 43]. Evidence to support the appropriateness to use of ElasticNet is described in Supplementary Material 8 and Supplementary Figure 5 in this study.

Metabolic aging clock for CRC diagnosis

To leverage the clinical utility for the chronological deviations observed in CRC subjects, our metabolic panel classifies all samples in the hypo-aging group as CRC. In addition, a multi-target panel was assembled by removing samples with normal carcinoembryonic antigen (CEA) measurements (≤ 2.5 μg/L for non-smokers and ≤ 5.0 μg/L for smokers) from the positives assigned by the metabolic clock, since CEA is a known CRC biomarker [66]. To simulate CRC incidence in general population, predictions from the testing dataset and the CRC cohort were bootstrapped with replacement at 30x coverage to an incidence of 760/10,023 for CRC APL samples and 65/10,023 for all stages of CRC samples [13]. Positive predictive values (PPVs) were calculated with the ratio of true positive counts to total predictive positive counts. PPV CIs were calculated with logit transformation and central limit theorem assumption as previously published [67].

Metabolic clock in CRC mutation status

A subpopulation of 412 subjects in CRC stage I group were profiled with *KRAS*, *NRAS* and *BRAF* mutations in the tumor tissue samples. All mutants were assigned into hypo-, normal and hyper-aging groups by the metabolite-based metabolic aging clock. Then, sample fractions of mutant samples in each group were calculated for each mutation to

reveal enrichment trending. Finally, 95% CIs of sample fractions were calculated by bootstrapping the classification results of mutant samples with replacement at the same size for 10,000 times to derive the 2.5% and 97.5% quantiles.

Aging metabolic network construction

To visualize the metabolic network underlying the global metabolomic aging patterns, age correlating metabolites were annotated to various metabolic modules and pathways.

List of Supplementary Materials

Supplementary Material 1 for the standard operating procedure (SOP) of sera sample collection.

Supplementary Material 2 for the detailed MS analytic pipeline for data acquisition, QA/QC, annotation, and structure identification.

Supplementary Figure. 1. Performances of the pathway based multivariate modeling as a function of the univariate |r| thresholds.

Supplementary Material 3 for the identification of age associated metabolic pathways.

Supplementary Figure. 2. Age-related metabolic alterations in colorectal cancer patients, relative to healthy individuals.

Supplementary Material 4 for the identification the metabolic aging clock with nine compounds.

Supplementary Figure 3 (A). A histogram of chronological age correlation (r) values obtained for the 157 KEGG/HMDB annotated features. (B). The match scores derived from the MS/MS spectra comparison of the metabolic clock nine metabolite markers.

Supplementary Material 5.

Supplementary Table 1. A. Demographics showing metabolic aging clock normal Δ aging subgroup distributions; B. Demographics showing metabolic aging clock hyper Δ aging subgroup distributions; C. Demographics showing metabolic aging clock hypo Δ aging subgroup distributions.

Supplementary Table 2. A. Metabolic Δ aging subgroup membership, hyper/normal/hypo, determined by the nine-compound based metabolic aging clock; B. Metabolic Δ aging subgroup memberships, hyper/normal/hypo, determined by the pathway based multivariate analysis; C. Classification performance after bootstrapping cohort samples to the true incidence rate in the general population,

Supplementary Table 3. A. Demographics showing sample distributions across age and sex in CRC stage I population; B. Demographics showing sample distributions across CRC mutations in CRC stage I population (N=412). Supplementary Table 4. %CV of the biomarker metabolite abundance in QC matrix samples which were spaced every 10 testing sera.

Supplementary Material 6.

Supplementary Figure 4. A. The performance of 100 randomly built models to detect APL. All model PPVs are < 0.35, which is statistically lower (p-value < 0.05) than the original model (Table 2); B. The performance of 100 randomly built models to detect CRC. All model PPVs are < 0.35, which is statistically lower (p-value < 0.05) than the original model (Table 2).

Supplementary Material 7.

Supplementary Table 5. Aging associated KEGG metabolic pathways and their associated mapped metabolomic features.

Supplementary Material 8.

Supplementary Figure 5. Evidence to support the appropriateness to use of ElasticNet in this study.

References

1. Harman D. The aging process: major risk factor for disease and death. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88: 5360–3.
2. Valdes AM, Glass D, Spector TD. Omics technologies and the study of human ageing. Nature Reviews: Genetics. 2013; 14: 601–7.
3. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018; 10: 573–91.
4. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019; 11: 303–27.
5. Bonomini F, Rodella LF, Rezzani R. Metabolic syndrome, aging and involvement of oxidative stress. Aging Dis. 2015; 6: 109-20.
6. Johnson SC, Rabinovitch PS, Kaeberlein M. mTOR is a key modulator of ageing and age-related disease. Nature. 2013; 493: 338-45.

7. Kennedy BK, Lamming DW. The Mechanistic Target of Rapamycin: The Grand ConducTOR of Metabolism and Aging. Cell Metab. 2016; 23: 990-1003.
8. Hwangbo N, Zhang X, Raftery D, Gu H, Hu SC, Montine TJ, et al. A Metabolomic Aging Clock Using Human Cerebrospinal Fluid. J Gerontol A Biol Sci Med Sci. 2022; 77: 744-54.
9. Robinson O, Chadeau Hyam M, Karaman I, Climaco Pinto R, Ala-Korpela M, Handakas E, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. Aging Cell. 2020; 19: e13149.
10. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14: R115.
11. Ferlitsch M, Reinhart K, Pramhas S, Wiener C, Gal O, Bannert C, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. JAMA. 2011; 306: 1352-8.
12. Brenner DE, Normolle DP. Biomarkers for cancer risk, early detection, and prognosis: the validation conundrum. Cancer Epidemiol Biomarkers Prev. 2007; 16: 1918-20.
13. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. New England Journal of Medicine. 2014; 370: 1287–97.
14. De Bie J, Guest J, Guillemin GJ, Grant R. Central kynurenine pathway shift with age in women. Journal of Neurochemistry. 2016; 136: 995–1003.
15. Refaey ME, McGee‐Lawrence ME, Fulzele S, Kennedy EJ, Bollag WB, Elsalanty M, et al. Kynurenine, a tryptophan metabolite that accumulates with age, induces bone loss. Journal of Bone and Mineral Research. 2017; 32: 2182–93.
16. Kim B-J, Hamrick MW, Yoo HJ, Lee SH, Kim SJ, Koh J-M, et al. The detrimental effects of kynurenine, a tryptophan metabolite, on human bone metabolism. The Journal of Clinical Endocrinology & Metabolism. 2019; 104: 2334–42.
17. Pertovaara M, Raitala A, Lehtimäki T, Karhunen P, Oja S, Jylhä M, et al. Indoleamine 2, 3-dioxygenase activity in nonagenarians is markedly increased and predicts mortality. Mechanisms of ageing and development. 2006; 127: 497–9.
18. Valdiglesias V, Marcos-Pérez D, Lorenzi M, Onder G, Gostner JM, Strasser B, et al. Immunological alterations in frail older adults: a cross sectional study. Experimental Gerontology. 2018; 112: 119–26.
19. Barrou Z, Charru P, Lidy C. Dehydroepiandrosterone (DHEA) and aging. Archives of Gerontology and Geriatrics. 1997; 24: 233–41.
20. Birkenhäger‐Gillesse EG, Derksen J, Lagaay AM. Dehydroepiandrosterone sulphate (DHEAS) in the oldest old, aged 85 and over. Annals of the New York Academy of Sciences. 1994; 719: 543–52.
21. Orentreich N, Brind JL, Rizer RL, Vogelman JH. Age changes and sex differences in serum dehydroepiandrosterone sulfate concentrations throughout adulthood. Journal of Clinical Endocrinology & Metabolism. 1984; 59: 551–5.
22. Ravaglia G, Forti P, Maioli F, Boschi F, Bernardi M, Pratelli L, et al. The relationship of dehydroepiandrosterone sulfate (DHEAS) to endocrine-metabolic parameters and functional status in the oldest-old. Results from an Italian study on healthy free-living over-ninety-year-olds. Journal of Clinical Endocrinology & Metabolism. 1996; 81: 1173–8.
23. Czibik G, Mezdari Z, Murat Altintas D, Bréhat J, Pini M, d'Humières T, et al. Dysregulated phenylalanine catabolism plays a key role in the trajectory of cardiac aging. Circulation. 2021; 144: 559–74.
24. Rudman D, Abbasi AA, Chaudry F, Mattson DE. Delayed plasma clearance of phenylalanine and tyrosine in elderly men. Journal of the American Geriatrics Society. 1991; 39: 33–8.
25. Gu X, Al Dubayee M, Alshahrani A, Masood A, Benabdelkamel H, Zahra M, et al. Distinctive Metabolomics Patterns Associated With Insulin Resistance and Type 2 Diabetes Mellitus. Front Mol Biosci. 2020; 7: 609806.
26. Garvey SM, Russ DW, Skelding MB, Dugle JE, Edens NK. Molecular and metabolomic effects of voluntary running wheel activity on skeletal muscle in late middle-aged rats. Physiol Rep. 2015; 3.
27. Wang T, Maden SK, Luebeck GE, Li CI, Newcomb PA, Ulrich CM, et al. Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. Clin Epigenetics. 2020; 12: 5.
28. Lamb YN, Dhillon S. Epi proColon® 2.0 CE: a blood-based screening test for colorectal cancer. Molecular Diagnosis & Therapy. 2017; 21: 225–32.
29. Foltran L, Maglio GD, Pella N, Ermacora P, Aprile G, Masiero E, et al. Prognostic role of *KRAS*, *NRAS*, *BRAF* and *PIK3CA* mutations in advanced colorectal cancer. Future Oncology (London, England). 2015; 11: 629–40.

30. Arrington AK, Heinrich EL, Lee W, Duldulao M, Patel S, Sanchez J, et al. Prognostic and predictive roles of *KRAS* mutation in colorectal cancer. International Journal of Molecular Sciences. 2012; 13: 12153–68.
31. Hsu HC, Thiam TK, Lu YJ, Yeh CY, Tsai WS, You JF, et al. Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. Oncotarget. 2016; 7: 22257-70.
32. Yonemura Y, Canbay E, Ishibashi H. Prognostic factors of peritoneal metastases from colorectal cancer following cytoreductive surgery and perioperative chemotherapy. The Scientific World Journal. 2013; 2013: 798394.
33. Zheng C, Li L, Xu R. Association of epigenetic clock with consensus molecular subtypes and overall survival of colorectal cancer. Cancer Epidemiology, Biomarkers and Prevention. 2019; 28: 1720–4.
34. Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, et al. Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J Proteome Res. 2009; 8: 113-7.
35. Liang L, Rasmussen MH, Piening B, Shen X, Chen S, Rost H, et al. Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. Cell. 2020; 181: 1680-92 e15.
36. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015; 25: 141-51.
37. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008; 24: 2534-6.
38. Delabriere A, Hohenester UM, Colsch B, Junot C, Fenaille F, Thevenot EA. proFIA: a data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry. Bioinformatics. 2017; 33: 3767-75.
39. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006; 78: 779-87.
40. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011; 6: 1060-83.
41. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007; 3: 211-21.
42. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. Anal Chem. 2017; 89: 1063-7.
43. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27-30.
44. Viant MR, Kurland IJ, Jones MR, Dunn WB. How close are we to complete annotation of metabolomes? Curr Opin Chem Biol. 2017; 36: 64-9.
45. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007; 35: D521-6.
46. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. Ther Drug Monit. 2005; 27: 747-51.
47. Anderson NM, Mucka P, Kern JG, Feng H. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein Cell. 2018; 9: 216-37.
48. Chen JQ, Russo J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochim Biophys Acta. 2012; 1826: 370-84.
49. Eng C, Kiuru M, Fernandez MJ, Aaltonen LA. A role for mitochondrial enzymes in inherited neoplasia and beyond. Nat Rev Cancer. 2003; 3: 193-202.
50. Juang HH. Modulation of mitochondrial aconitase on the bioenergy of human prostate carcinoma cells. Mol Genet Metab. 2004; 81: 244-52.
51. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med. 2009; 360: 765-73.
52. Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics. 2009; 25: 3151-7.
53. Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. J Proteome Res. 2012; 11: 1354-63.
54. Zarei I, Baxter BA, Oppel RC, Borresen EC, Brown RJ, Ryan EP. Plasma and Urine Metabolite Profiles Impacted by Increased Dietary Navy Bean Intake in Colorectal Cancer Survivors: A Randomized-Controlled Trial. Cancer Prev Res (Phila). 2021; 14: 497-508.

55. Sheng L, Jena PK, Hu Y, Wan YY. Age-specific microbiota in altering host inflammatory and metabolic signaling as well as metabolome based on the sex. Hepatobiliary Surg Nutr. 2021; 10: 31-48.
56. Zhou H, Yang Z, Yue J, Chen Y, Chen T, Mu T, et al. Identification of potential hub genes via bioinformatics analysis combined with experimental verification in colorectal cancer. Mol Carcinog. 2020; 59: 425-38.
57. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. Bioinformatics. 2011; 27: 175-81.
58. Airley RE, Mobasheri A. Hypoxic regulation of glucose transport, anaerobic metabolism and angiogenesis in cancer: novel pathways and targets for anticancer therapeutics. Chemotherapy. 2007; 53: 233-56.
59. Tian J, Xue W, Yin H, Zhang N, Zhou J, Long Z, et al. Differential Metabolic Alterations and Biomarkers Between Gastric Cancer and Colorectal Cancer: A Systematic Review and Meta-Analysis. Onco Targets Ther. 2020; 13: 6093-108.
60. Malila N, Virtanen M, Pietinen P, Virtamo J, Albanes D, Hartman AM, et al. A comparison of prospective and retrospective assessments of diet in a study of colorectal cancer. Nutr Cancer. 1998; 32: 146-53.
61. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia (Auckl). 2015; 3: 83-92.
62. Sylvester KG, Hao S, You J, Zheng L, Tian L, Yao X, et al. Maternal metabolic profiling to assess fetal gestational age and predict preterm delivery: a two-centre retrospective cohort study in the US. BMJ Open. 2020; 10: e040647.
63. Max K. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008; 28.
64. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013; 29: 1830-1.
65. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498-504.
66. Reiter W, Stieber P, Reuter C, Nagel D, Lau-Werner U, Lamerz R. Multivariate analysis of the prognostic value of CEA and CA 19-9 serum levels in colorectal cancer. Anticancer Research. 2000; 20: 5195–8.
67. Ying G-S, Maguire MG, Glynn RJ, Rosner B. Calculating sensitivity, specificity, and predictive values for correlated eye data. Investigative Ophthalmology and Visual Science. 2020; 61: 29.
68. Brenner H, Werner S, Chen H. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 2014; 371: 184-5.

Declarations

Ethics approval and consent to participate:

This study was approved by the Shanghai Municipal Center for Disease Control and Prevention Ethical Review Committee (No. 2019-4) and the Ethical Committee and Institutional Review Board of Fudan University Shanghai Cancer Center (No. 1902197-15). Written informed consent was obtained from the general population and all the patients at enrolment.

Acknowledgments:

We would like to thank the Department of Biobank, Fudan University Shanghai Cancer Center, and mProbe Inc. R&D team for their support and collaboration. We are also grateful to Dr. Richard Mortensen of mProbe Inc. for his constructive feedback on the manuscript.

Funding:

Science and Technology Commission of Shanghai Municipality 18401933401 (CJC)

Science and Technology Commission of Shanghai Municipality 18401933402 (JP)

Science and Technology Commission of Shanghai Municipality 18401933403 (CF)

National Natural Science Foundation of China U1932145 (JP)

Shanghai Clinical Research Center for Aging and Medicine 19MC1910500 (CF)

Author contributions:

Conceptualization: CJC, YG, CF, JP, XBL

Sample preparation: LZ, SM, YG

Experiments design: CJC, XZ, BJ, XBL

LCMS metabolomic profiling guidance: WL, ST, JS, RYL, XBL

Data analysis: XZ, BJ, ZH, LT, XBL

Molecular finding and clinical interpretation: LZ, SM, CJC, YG, CF, JP, XBL

Visualization: LZ, BJ, XBL

Writing – original draft: XZ, CJC, XBL

Writing – review & editing: LZ, SM, CJC, XBL

Response to the internal and external peer reviewers: XBL, BJ, CJC, JP, DM, HC, JS, SN, SC, KGS, JCW, LT

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

Researchers interested in utilizing de-identified participant data may obtain access by submitting a written request to the respective corresponding authors affiliated with the Shanghai CDC (for the general population) and the Fudan University Cancer Center (for the APL and CRC populations). Access to the data is granted starting from the publication date, subject to the completion and approval of a Data Access Agreement. This agreement undergoes a review process by the ethics committees responsible for authorizing this research.

Table 1. Demographics showing sample distributions across age and sex in normal and CRC populations.

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	3002	715	1088	427	417
Age group, N (%)					
(30,40]	21 (0.7)	14 (2)	12 (1.1)	9 (2.1)	17 (4.1)
(40,50]	503 (16.8)	105 (14.7)	148 (13.6)	55 (12.9)	60 (14.4)
(50,60]	1124 (37.4)	219 (30.6)	295 (27.1)	104 (24.4)	123 (29.5)
(60,70]	851 (28.3)	268 (37.5)	426 (39.2)	161 (37.7)	138 (33.1)
(70,80]	387 (12.9)	90 (12.6)	175 (16.1)	80 (18.7)	66 (15.8)
(80,90]	116 (3.9)	19 (2.7)	32 (2.9)	18 (4.2)	13 (3.1)
Sex, N (%)					
Male	1312 (43.7)	410 (57.3)	623 (57.3)	272 (63.7)	234 (56.1)
Female	1690 (56.3)	305 (42.7)	465 (42.7)	155 (36.3)	183 (43.9)

Table 2. Comparison of positive predictive values (PPVs) of tests for CRC diagnosis.

	APL	CRC - all stages
Metabolic clock panel [a]	65.5% (62.3%, 68.5%)	12.7% (10.0%, 15.9%)
Multi-target panel [b]	68.4% (64.3%, 72.2%)	21.4% (17.8%, 25.9%)
CEA	5.2% (4.6%, 5.7%)	0.4% (0.3%, 0.7%)
Cologuard [c]	20.0% (18.0%, 22.0%)	3.72% (2.85%, 4.76%)
Septin 9 methylation [d]	9.5% (9.1%, 9.9%)	2.3% (1.8%. 2.9%)

a. Contains the 9 metabolic biomarkers

b. Contains the 9 metabolic biomarkers and carcinoembryonic antigen (CEA)

c. Data cited from reference [68].

d. Data cited from external source reference [28].

Supplementary Material 1

Standard Operating Procedure:

Serum sample collection and processing for molecular analysis

Serum is the liquid fraction of whole blood that is collected after the blood is allowed to clot. The clot is removed by centrifugation and the resulting supernatant, designated serum, is carefully removed using a Pasteur pipette.

INSTRUCTIONS for Sample Collection:

1. **BLOOD**: Collect **blood** via venipuncture into vacutainers (see below):

BLOOD SAMPLE COLLECTION AND PROCESSING[1]:

Serum:

1. Using sterile technique, perform venipuncture and collect 2-mL blood into a vacutainer containing no pro- or anticoagulant (red top or tiger-top tube). Document date and time of collection.

2. Whole blood can be stored at 4-8°C within 12 hours before the serum is separated, but it must not be frozen.

3. Allow blood to clot by leaving tube undisturbed at room temperature for 30 min*.

4. Centrifuge tubes at 1000 × g for 10 min in a refrigerated centrifuge. Document date and time. The resulting supernatant is designated serum.

5. Following centrifugation, it is important to immediately transfer the liquid component (serum) into a clean freezer compatible polypropylene tube using a Pasteur pipette. The samples should be maintained at 2-8°C while handling.

6. The serum should be apportioned into 2 × 250 µL aliquots, and labeled as: specimen type: blood (serum).

7. Freeze immediately at -20°C within 6 hours from the time of collection, and transfer to -80°C within 7 days for long-term storage. Document date and time。

8. Record the following information in the CASE REPORT FORM:

Specimen matrix

Collection date and time

Centrifugation times

Date and time placed in storage

Freezer ID# and temperature

Initials

It is important to avoid freeze-thaw cycles because this is detrimental to many serum and plasma components. Samples that are hemolyzed, icteric or lipemic can invalidate certain tests, and therefore should be documented.

[1] J Proteome Res. 2009 Jan; 8(1): 113–117. (PMID: 19072545)

*in order to obtain serum of high quality, blood samples should be allowed time to form a clot at room temperature for 30-60 min.

Serum tubes

For more information on the tubes, please visit Becton-Dickinson Resources:

https://www.bd.com/resource.aspx?IDX=7220

Red	No anticoagulant
Lavender	Treated with EDTA

STORAGE AND SHIPPING:

1. Sample Storage

All samples should be stored in a frozen environment (-80°C).

Avoid freeze/thaw cycles as this will negatively affect sample integrity.

Ensure that all samples are labeled with patient ID, type of sample and date and time of draw.

Each ID number should be associated with either a paper card orelectronic manifest including the following information:

2. Sample Shipping

All samples should be shipped in a frozen environment (-80°C).

Avoid freeze/thaw cycles, as this will damage the sample integrity.

Ensure that all samples are labeled with patient ID, type of sample and date and time of draw.

Shipments should include either card information or electronic manifestassociated with patient ID numbers.

First shipment should be sent once 100 serum and 100 plasma samples have been collected.

Second shipment should be sent at completion of study.

Shipping Address

Notify study collaborators of the shipment within 24 hours of sending the package. In this email, include the following:

Collection Site

Date of shipment

Tracking number

Number of samples sent

Electronic manifest summarizing the samples included in the shipment

Please name the file: YYYYMMDD Country_Shipment Manifest

Supplementary Material 2

Detailed description of the MS analytic pipeline for data acquisition, QA/QC, annotation, and structure identification

Deep metabolomic profiling pipeline and data analysis processes

Serum samples were retrieved from -80°C freezers and thawed on ice. A mixture of 240 μL extraction buffer (stored at -20°C) consisting of methanol, acetonitrile and ddH_2O (5:3:2 v/v) was added to 10 μL of each sample. Samples were vortexed at 4°C and then centrifuged at 12,000 g for 30 min at the same temperature. Supernatant (170 μL) from each sample was transferred to a vial and stored at -20°C before injection.

MS acquisition

Flow-injection analysis (FIA) was performed with QE-Plus mass spectrometer (Thermo Fisher, Shanghai, China). Samples were injected at 10 μL each for 3 min with mobile phase (flow rate at 50 μL/min) consisting of 0.1 % formic acid in 5% acetonitrile. The injected samples were sprayed at 4 kV and 325°C and scanned with a m/z range of 60-900 at a resolution of 140,000. Ionization modes were set to both positive and negative in turn to collect complete metabolic profiles for each sample. Meanwhile, these samples were also analyzed with liquid chromatography-tandem mass spectrometry (LC-MS/MS)

in data-dependent acquisition (DDA) mode according to a protocol previously published [1].

MS QA/QC

Several types of controls were analyzed in concert with the experimental samples: testing matrix sample, a pool generated by taking a small volume of each experimental sample; QC matrix sample, a pool of well-characterized human sera (Sigma-Aldrich, S1-M, human serum, normal) served as a technical replicate throughout the data sets; extracted water samples served as process blanks. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Bland–Altman plots [2] were used in analyzing the agreement between two different runs for the quality control. Overall process variability was determined by calculating the median RSD for all endogenous metabolites present in 100% of the pooled QC matrix samples. Experimental samples were randomized across the platform runs. To mitigate the potential blood processing and batch variance, QC matrix samples were spaced evenly (every 10 testing sera) among the injections to ensure intra- and inter-batch consistency (Supplementary Table 4) from metabolomics analytics. All the data analyses in this study were performed using R (version 4.0.3) programming language

and R software packages. Univariate statistical analysis was performed with R Stats (version 4.0.3, statistics analytics) package.

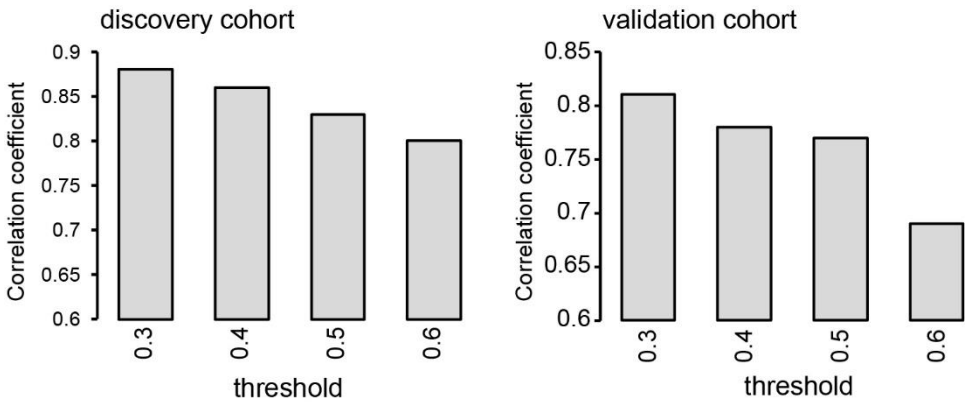
MS data processing

Raw MS files were converted and centroided with ProteoWizard and saved to mzXML files [3]. Injection zones and m/z bands were detected with proFIA (version 1.15.0, mass spectrometry data preprocessing analytics) package [4]. Detected bands were then aligned to derive data matrix across samples in the same batch with MzClust method from xcms (version 3.12.0, high resolution mass spectrometric spectra analytics) package [5]. Robust locally estimated scatterplot smoothing (LOESS) model was applied to each set of aligned bands to remove intra-batch drifting along injection process [6]. Inter-batch difference was removed by normalizing to the median of normal baseline per set of aligned bands (namely metabolic features). Metabolic features with more than 30% missing values were removed. The filtered data matrix was $\log_2$ transformed, centered, and scaled with the means and standard deviations (SDs) of the normal population in each metabolic feature. Finally, missing values were imputed with k-nearest neighbor (kNN) algorithm from impute (version 1.64.0, KNN imputation analytics) package.

Mass spectrometric feature annotation

Healthy general population subjects were split randomly into either training (75%) or testing (25%) sets. During the downstream elastic net analysis of the

age associated pathways, different univariate |r| thresholds (0.3, 0.4, 0.5, 0.6) were used for the selection of LCMS features correlating to the chronological age. Supplementary Fig. 1 summarized the resulting pathway based elastic net model performance as a function of different |r| thresholds.



Supplementary Figure 1. Performances of the pathway based multivariate modeling as a function of the univariate |r| thresholds. Prediction performance is evaluated as a correlation coefficient between the predicted metabolic age and the chronological age. Training (75%discovery cohort) and testing (25%, validation cohort) sets were constructed from the healthy general population.

LCMS metabolic features in the training set were first correlated with chronological age and those with absolute correlation coefficients $|r| \geq 0.3$ were preserved. As defined by the Metabolomics Standards Initiative[7], there

are four levels of metabolite identification confidence: confidently identified compounds (level 1), putatively annotated compounds (level 2), and putatively annotated compound classes (level 3) and unknown compounds (level 4). The preserved features were annotated (level 2 or 3) with 5 ppm tolerance with xMSannotator[8] via KEGG compound data base[9].

Structural identification and metabolite biomarker discovery

Definitive (level 1) identification requires comparing the two or more orthogonal properties such as retention time, m/z, and fragmentation mass spectrum for the metabolite of interest to the same properties of an authentic chemical standard observed under identical analytical conditions. Metabolite biomarker identification was performed as a level 1 identification with chemical standards according to MSI [10]. With tandem mass spectrometry (MS/MS, Thermo Q Exactive plus, Thermo Fisher) data of serum samples and manual review confirmation, the generated MS1/MS2 pairs were searched in the public databases: HMDB[11] (http://www.hmdb.ca/), MoNA (http://mona.fiehnlab.ucdavis.edu/), MassBank (http://www.massbank.jp/), METLIN[12] (https://metlin.scripps.edu), and NIST (https://www.nist.gov/). The metabolites of interest were procured and subjected to a level 1 identification comparing the retention time, MS1 and MS2 patterns with the biomarker candidates, using the same LCMS/MS protocol with the sample analysis.

Metabolite biomarker candidates were selected from the metabolites that met all the following criteria: (a) the metabolite appeared in the contributing pathways with importance score > 0; (b) the metabolite's corresponding metabolic feature met $|r| \geq 0.3$; (c) the metabolite was validated with structural identification.

Among the changing metabolites between the healthy general population and CRC patient samples, the major metabolic modules were "citrate cycle", "tyrosine metabolism" and "valine, leucine degradation", where similar changes across all stages were observed (Supplementary Fig. 2 in supplementary material 3). The citrate cycle serves as the central hub of energy metabolism and is well known for its dysregulation in oncogenesis.[13] Two metabolic modules were found to be stage-specific: "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism". Furthermore, we also identified a wide variety of metabolites scattering in many other metabolic modules, including "steroid hormone biosynthesis" and "oxidative phosphorylation". Many oncogenes and tumor suppressors regulate in a bi-directional manner the expression of fuel transporters and activity of cycle-related enzymes in cancer cells.[14] These enzymes include aconitase (also known as aconitate hydratase, AH), isocitrate dehydrogenase (IDH), fumarase (FH), succinate dehydrogenase (SDH) and α-ketoglutarate

dehydrogenase complex (KGDHC).[15-17] Our study validated the dysregulation of IDH activity by changes that occurred in isocitrate and oxalosuccinate levels (compounds at both ends of reaction) at the same time, while proposing other possibly changed reactions among oxaloacetate, citrate, and isocitrate, such as citrate synthase (oxaloacetate to citrate) and citrate hydro-lyase (citrate to isocitrate). There were two metabolites with significant changes in tyrosine metabolism. Homovanillate has been reported as a biomarker for breast cancer and colorectal cancer in urine [18, 19]. Regulated by tyrosine through levodopa (L-DOPA), homovanillate is the major product of monoamine oxidase and catechol-O-methyltransferase on dopamine. 3-(4-hydroxyphenyl)lactate has also been reported as a biomarker for colorectal cancer in both plasma and urine,[20] regulated by 3-(4-hydroxyphenyl) pyruvate. Although these two metabolites have been studied in CRC detection, our study unveiled their correlation with the aging process, suggesting them as possible evidence of oncogenic alterations on aging. On the other hand, sugar alcohols including xylitol, lyxitol, and ribitol are age-specific in microbiota of mice, by changing host inflammatory and metabolic signaling such as insulin sensitivity.[21] Our study shed light on the dysregulation of these metabolites as possible age-related biomarker candidates in human colorectal cancer in its early stages. In general, our study constructed a high-resolution panoramic view of metabolic changes in

CRC samples against the metabolic aging clock trained from a healthy general population with nine age-related metabolites, complementing and revealing the underlying biology of both aging and oncogenic pathophysiology.

Our study revealed progressive alteration of metabolic profiles as a function of different CRC stages (Supplementary Fig. 2 in supplementary material 3). During APL and stage I, metabolites in "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism" modules were dysregulated in correlation with chronological age, resulting in a hypo-aging trend. Bioinformatics analysis combined with experimental verification in colorectal cancer found differentially expressed hub genes significantly enriched in the "pentose and glucuronate conversion" pathway[22]. Changes in metabolic driver genes were observed at the initial stages of oncogenesis,[23] and glucose transport and anaerobic metabolism have been studied in cancers (including CRC) for over a decade [24]. Both "pentose and glucuronate conversion" and "starch as sucrose metabolism" are downstream modules of glycolysis on metabolic pathways (hsa01100). This is in line with our observation that metabolites in the "pentose and glucuronate conversion" module are dysregulated accordingly during the early stages of CRC.

Another meta-analysis study of colorectal cancer discovered that differentially expressed metabolites enriched in the ascorbate and aldarate metabolism in and all the other four modules above[25, 26]. Our study showed the correlation of metabolites in "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism" as major contributors in the metabolic aging clock. Although our study revealed the dysregulated ascorbate and aldarate metabolism in the early stages of CRC, including APL and stage I CRC, the temporal correlations of metabolites in these two modules were to certain extent re-established during stage II and III CRC. Possibly, the cancer-associated hypoxia rewires the metabolism and re-activates glycolysis during CRC development [27].

Z statistics test analysis

Cohort sex-composition impact on the metabolomic clock performance was performed as follows. Let $\hat{\rho}_M$ and $\hat{\rho}_F$ be the observed correlation coefficients of the metabolomic clock in males and female testing cohorts, respectively. The test Z statistic in comparing them is given by

$$Z = \frac{\left[\frac{1}{2}\log\left(\frac{1+\hat{\rho}_F}{1-\hat{\rho}_F}\right) - \frac{1}{2}\log\left(\frac{1+\hat{\rho}_M}{1-\hat{\rho}_M}\right)\right]}{\sqrt{\frac{1}{n_F-3} + \frac{1}{n_M-3}}}$$

The *p*-value can be calculated as $P(|N(0,1)| > |Z|)$.

Similar Z statistics analysis was performed to evaluate the positive predictive value (PPV) differences when different approaches were used to detect APL and CRC.

REFERENCE

1. Liang L, Rasmussen M-LH, Piening B, Shen X, Chen S, Röst H, et al. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. Cell. 2020; 181: 1680–92.
2. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015; 25: 141-51.
3. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008; 24: 2534–6.
4. Delabrière A, Hohenester UM, Colsch B, Junot C, Fenaille F, Thévenot EA. *proFIA*: a data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry. Bioinformatics. 2017; 33: 3767–75.
5. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical Chemistry. 2006; 78: 779–87.
6. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature Protocols. 2011; 6: 1060–83.
7. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007; 3: 211-21.
8. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. Anal Chem. 2017; 89: 1063-7.
9. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27-30.
10. Viant MR, Kurland IJ, Jones MR, Dunn WB. How close are we to complete annotation of metabolomes? Curr Opin Chem Biol. 2017; 36: 64-9.
11. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the human metabolome database. Nucleic Acids Research. 2007; 35: D521–D6.
12. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. Therapeutic Drug Monitoring. 2005; 27: 747–51.
13. Anderson NM, Mucka P, Kern JG, Feng H. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein & Cell. 2018; 9: 216–37.
14. Chen J-Q, Russo J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2012; 1826: 370–84.
15. Eng C, Kiuru M, Fernandez MJ, Aaltonen LA. A role for mitochondrial enzymes in inherited neoplasia and beyond. Nature Reviews: Cancer. 2003; 3: 193–202.

16. Juang H-H. Modulation of mitochondrial aconitase on the bioenergy of human prostate carcinoma cells. Molecular Genetics and Metabolism. 2004; 81: 244–52.
17. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. New England Journal of Medicine. 2009; 360: 765–73.
18. Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics. 2009; 25: 3151–7.
19. Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. Journal of Proteome Research. 2012; 11: 1354–63.
20. Zarei I, Baxter BA, Oppel RC, Borresen EC, Brown RJ, Ryan EP. Plasma and urine metabolite profiles impacted by increased dietary navy bean intake in colorectal cancer survivors: a randomized-controlled trial. Cancer Prevention Research. 2021; 14: 497–508.
21. Sheng L, Jena PK, Hu Y, Wan Y-JY. Age-specific microbiota in altering host inflammatory and metabolic signaling as well as metabolome based on the sex. Hepatobiliary Surgery and Nutrition. 2021; 10: 31–48.
22. Zhou H, Yang Z, Yue J, Chen Y, Chen T, Mu T, et al. Identification of potential hub genes via bioinformatics analysis combined with experimental verification in colorectal cancer. Molecular Carcinogenesis. 2020; 59: 425–38.
23. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. Bioinformatics. 2011; 27: 175–81.
24. Airley RE, Mobasheri A. Hypoxic regulation of glucose transport, anaerobic metabolism and angiogenesis in cancer: novel pathways and targets for anticancer therapeutics. Chemotherapy. 2007; 53: 233–56.
25. Tian J, Xue W, Yin H, Zhang N, Zhou J, Long Z, et al. Differential metabolic alterations and biomarkers between gastric cancer and colorectal cancer: a systematic review and meta-analysis. OncoTargets and Therapy. 2020; 13: 6093–108.
26. Malila N, Virtanen M, Pietinen P, Virtamo J, Albanes D, Hartman AM, et al. A comparison of prospective and retrospective assessments of diet in a study of colorectal cancer. Nutrition and Cancer. 1998; 32: 146–53.
27. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia. 2015; 3: 83–92.

Supplementary Material 3

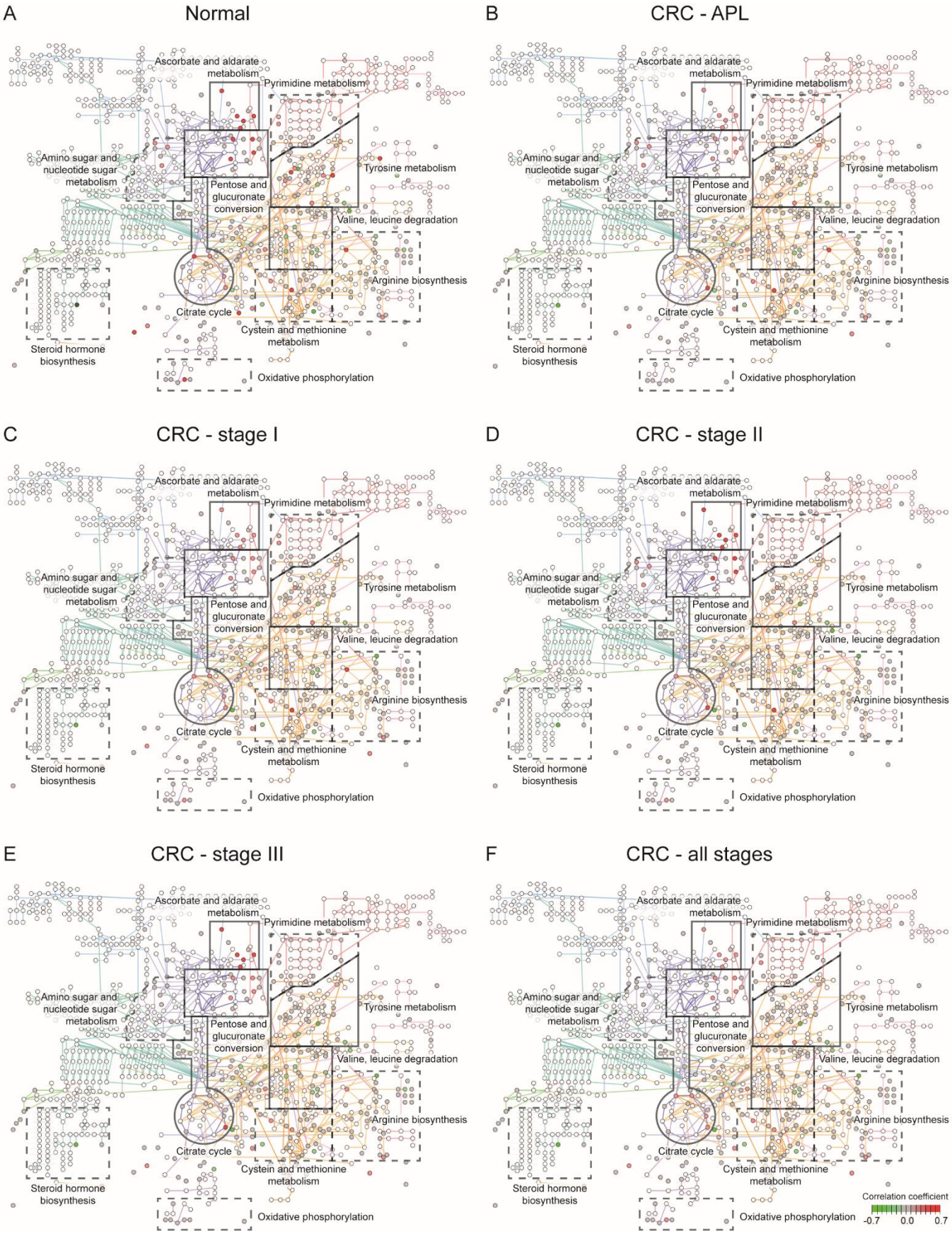
We were among the first groups to propose a pathway-based computational methodology for chronological event prediction with global metabolomics [1]. Representing the biological grouping of individual metabolomic features to associate to targeted clinical outcomes, the pathway-based modeling analysis was found to have less variability and higher sensitivity than direct global metabolomic feature-based modeling.

Annotated mass spectrometric features were aggregated into relevant KEGG pathways, and the value of each pathway was calculated as the weighted sum of the normalized measurement values of metabolites on the pathway divided by the number of mapped metabolites. Specifically, univariate correlations r_i of each metabolic feature i were first calculated with chronological age. Next, feature directionalities (up/down) were calculated according to the positive/negative signs of the correlation coefficients. Then, the annotated metabolites were mapped to KEGG pathways k in each directionality as $P_{k,\,\text{up/down}} = \{i | \forall \text{compound}_i \in \text{pathway}_k, r_i \gtrless 0\}$. Finally, pathway expression values p_{kj} for sample j were calculated with relevant metabolite serological abundance values f_{ij} as follows, where $\text{n}(A)$ stands for the number of elements in set A:

$$p_{kj,\,\text{up/down}} = \frac{\sum_{i \in P_{k,\,\text{up/down}}} (r_i f_{ij})}{\text{n}(P_{k,\,\text{up/down}})}$$

The derived pathway matrix was used to perform the multivariate analysis by fitting an elastic net model using all pathways against age with tuned hyperparameters α and λ from cross validation (CV) on the training set to minimize the mean squared prediction error. The importance score of each pathway biomarker was measured by the absolute value of the estimated β coefficient of the normalized pathway level. The higher the importance score suggests a bigger role in predicting age. Caret (version 6.0-86, elastic net machine learning analytics) package [2] was used. Pathways with an importance score > 0 were defined as age predictor contributing pathways. Age predictions were made on both training (75%) and testing (25%) sets of the healthy general population.

To further unveil an overview of metabolic changes from the reference metabolic ageotype of healthy general population to CRC, aging associating metabolites were annotated in KEGG hsa01100 metabolic pathways. Pathview (version 1.30.1, KEGG metabolic pathway visualization analytics) package [3] in Cytoscape (KEGG metabolic pathway network visualization analytics) [4] was used with correlation coefficients of each sample group as color key (Supplementary Figure 2). Clusters of metabolites with significant correlations in the healthy general population were identified with KEGG modules and pattern changes in CRC samples were identified from these metabolite clusters.


A
Normal
B
CRC - APL
C
CRC - stage I
D
CRC - stage II
E
CRC - stage III
F
CRC - all stages
Ascorbate and aldarate metabolism
Pyrimidine metabolism
Amino sugar and nucleotide sugar metabolism
Tyrosine metabolism
Pentose and glucuronate conversion
Valine, leucine degradation
Arginine biosynthesis
Citrate cycle
Cystein and methionine metabolism
Steroid hormone biosynthesis
Oxidative phosphorylation
Correlation coefficient
-0.7
0.0
0.7

Supplementary Figure. 2. Age-related metabolic alterations in colorectal cancer patients, relative to healthy individuals. With reference to **A.** healthy baseline, changes were observed in the "citrate cycle" and "tyrosine metabolism" for all CRC stages (**B-F**). There were also stage-specific alterations, including those in the "ascorbate and aldarate metabolism" for APL and stage I CRC (**B-C**), "pentose and glucuronate conversion" in all but stage I CRC (B, C, F) and "valine, leucine degradation" in all but stage Ill CRC **(B-D)**.

Among the changing metabolites between the healthy general population and CRC patient samples, the major metabolic modules were "citrate cycle", "tyrosine metabolism" and "valine, leucine degradation", where similar changes across all stages were observed (Supplementary Fig. 2). The citrate cycle serves as the central hub of energy metabolism and is well known for its dysregulation in oncogenesis [5]. Two metabolic modules were found to be stage-specific: "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism". Furthermore, we also identified a wide variety of metabolites scattering in many other metabolic modules, including "steroid hormone biosynthesis" and "oxidative phosphorylation". Many oncogenes and tumor suppressors regulate in a bi-directional manner the expression of fuel transporters and activity of cycle-related enzymes in cancer cells [6]. These enzymes include aconitase (also known as aconitate hydratase, AH), isocitrate dehydrogenase (IDH), fumarase (FH), succinate dehydrogenase

(SDH) and α-ketoglutarate dehydrogenase complex (KGDHC) [7-9]. Our study validated the dysregulation of IDH activity by changes that occurred in isocitrate and oxalosuccinate levels (compounds at both ends of reaction) at the same time, while proposing other possibly changed reactions among oxaloacetate, citrate, and isocitrate, such as citrate synthase (oxaloacetate to citrate) and citrate hydro-lyase (citrate to isocitrate). There were two metabolites with significant changes in tyrosine metabolism. Homovanillate has been reported as a biomarker for breast cancer and colorectal cancer in urine [10, 11]. Regulated by tyrosine through levodopa (L-DOPA), homovanillate is the major product of monoamine oxidase and catechol-O-methyltransferase on dopamine. 3-(4-hydroxyphenyl)lactate has also been reported as a biomarker for colorectal cancer in both plasma and urine [12], regulated by 3-(4-hydroxyphenyl) pyruvate. Although these two metabolites have been studied in CRC detection, our study unveiled their correlation with the aging process, suggesting them as possible evidence of oncogenic alterations on aging. On the other hand, sugar alcohols including xylitol, lyxitol, and ribitol are age-specific in microbiota of mice, by changing host inflammatory and metabolic signaling such as insulin sensitivity [13]. Our study shed light on the dysregulation of these metabolites as possible age-related biomarker candidates in human colorectal cancer in its early stages. In general, our study constructed a high-resolution panoramic view of metabolic

changes in CRC samples against the metabolic aging clock trained from a healthy general population with nine age-related metabolites, complementing and revealing the underlying biology of both aging and oncogenic pathophysiology.

Our study revealed progressive alteration of metabolic profiles as a function of different CRC stages (Supplementary Fig. 2). During APL and stage I, metabolites in "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism" modules were dysregulated in correlation with chronological age, resulting in a hypo-aging trend. Bioinformatics analysis combined with experimental verification in colorectal cancer found differentially expressed hub genes significantly enriched in the "pentose and glucuronate conversion" pathway [14]. Changes in metabolic driver genes were observed at the initial stages of oncogenesis [15], and glucose transport and anaerobic metabolism have been studied in cancers (including CRC) for over a decade [16]. Both "pentose and glucuronate conversion" and "starch as sucrose metabolism" are downstream modules of glycolysis on metabolic pathways (hsa01100). This is in line with our observation that metabolites in the "pentose and glucuronate conversion" module are dysregulated accordingly during the early stages of CRC.

Another meta-analysis study of colorectal cancer discovered that differentially expressed metabolites enriched in the ascorbate and aldarate metabolism in and all the other four modules above [17, 18]. Our study showed the correlation of metabolites in "pentose and glucuronate conversion" and "ascorbate and aldarate metabolism" as major contributors in the metabolic aging clock. Although our study revealed the dysregulated ascorbate and aldarate metabolism in the early stages of CRC, including APL and stage I CRC, the temporal correlations of metabolites in these two modules were to certain extent re-established during stage II and III CRC. Possibly, the cancer-associated hypoxia rewires the metabolism and re-activates glycolysis during CRC development [19].

REFERENCE

1. Sylvester KG, Hao S, You J, Zheng L, Tian L, Yao X, et al. Maternal metabolic profiling to assess fetal gestational age and predict preterm delivery: a two-centre retrospective cohort study in the US. BMJ Open. 2020; 10: e040647.
2. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008; 28: 1 - 26.
3. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013; 29: 1830–1.
4. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research. 2003; 13: 2498–504.
5. Anderson NM, Mucka P, Kern JG, Feng H. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein & Cell. 2018; 9: 216–37.
6. Chen J-Q, Russo J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2012; 1826: 370–84.
7. Eng C, Kiuru M, Fernandez MJ, Aaltonen LA. A role for mitochondrial enzymes in inherited neoplasia and beyond. Nature Reviews: Cancer. 2003; 3: 193–202.

8. Juang H-H. Modulation of mitochondrial aconitase on the bioenergy of human prostate carcinoma cells. Molecular Genetics and Metabolism. 2004; 81: 244–52.
9. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. New England Journal of Medicine. 2009; 360: 765–73.
10. Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics. 2009; 25: 3151–7.
11. Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. Journal of Proteome Research. 2012; 11: 1354–63.
12. Zarei I, Baxter BA, Oppel RC, Borresen EC, Brown RJ, Ryan EP. Plasma and urine metabolite profiles impacted by increased dietary navy bean intake in colorectal cancer survivors: a randomized-controlled trial. Cancer Prevention Research. 2021; 14: 497–508.
13. Sheng L, Jena PK, Hu Y, Wan Y-JY. Age-specific microbiota in altering host inflammatory and metabolic signaling as well as metabolome based on the sex. Hepatobiliary Surgery and Nutrition. 2021; 10: 31–48.
14. Zhou H, Yang Z, Yue J, Chen Y, Chen T, Mu T, et al. Identification of potential hub genes via bioinformatics analysis combined with experimental verification in colorectal cancer. Molecular Carcinogenesis. 2020; 59: 425–38.
15. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. Bioinformatics. 2011; 27: 175–81.
16. Airley RE, Mobasheri A. Hypoxic regulation of glucose transport, anaerobic metabolism and angiogenesis in cancer: novel pathways and targets for anticancer therapeutics. Chemotherapy. 2007; 53: 233–56.
17. Tian J, Xue W, Yin H, Zhang N, Zhou J, Long Z, et al. Differential metabolic alterations and biomarkers between gastric cancer and colorectal cancer: a systematic review and meta-analysis. OncoTargets and Therapy. 2020; 13: 6093–108.
18. Malila N, Virtanen M, Pietinen P, Virtamo J, Albanes D, Hartman AM, et al. A comparison of prospective and retrospective assessments of diet in a study of colorectal cancer. Nutrition and Cancer. 1998; 32: 146–53.
19. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia. 2015; 3: 83–92.

Supplementary Material 4

Analytical pipeline and workflow.

Step 1: LCMS feature reduction

Out of the 1,603 LCMS features, 157 features (Pearson correlation, $|r| \geq 0.3$) were selected for downstream modeling analysis. To prepare for the elastic net analysis of age associated pathways, different univariate |r| thresholds (0.3, 0.4, 0.5, 0.6) were used for the selection of LCMS features correlating to the chronological age. Supplementary Figure 1 summarized the resulting pathway based elastic net model performance as a function of the univariate |r| thresholds.

Step 2: 157 LCMS feature mapping to metabolic pathways

The preserved 157 features were annotated (level 2 or 3) with 5 ppm tolerance with xMSannotator[1] via KEGG and HMDB compound data base[2]. Annotated mass spectrometric features were aggregated into relevant KEGG pathways (n=59), and the value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of mapped metabolites.

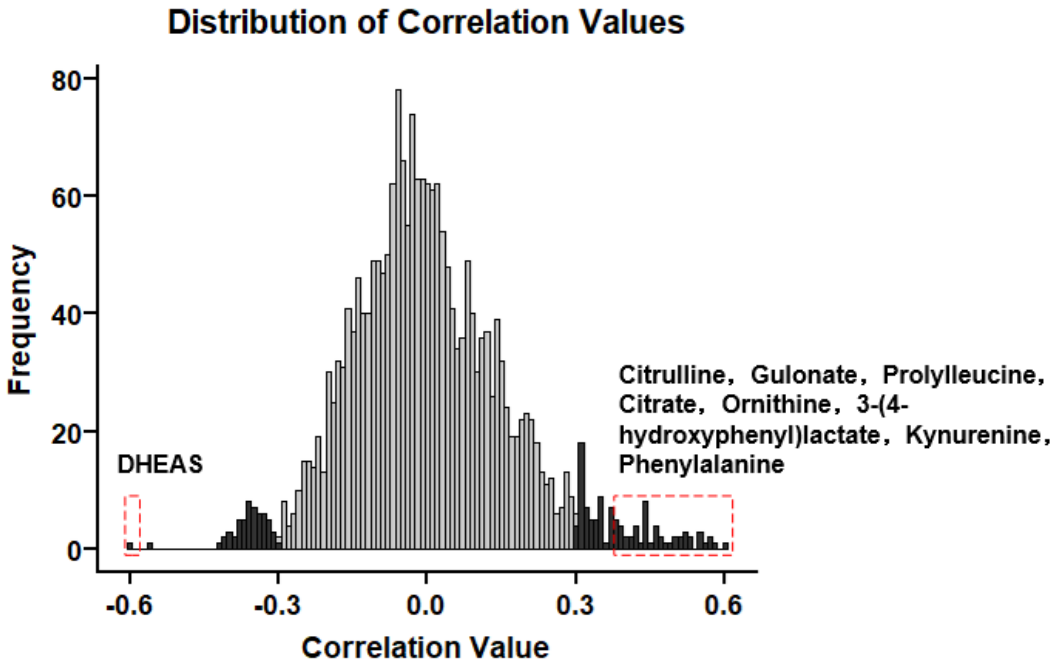
Step 3: Elastic net method was used to construct a linear model to predict chronological age

With the elastic net analysis, 42/59 pathways were identified (beta coefficient not equal to 0) to contribute to the age prediction.

Step 4: Metabolic pathway based aging clock deviation analysis. To search for chronological deviations associated with pathologies of aging, a quantile linear regression model was fit with quantreg (version 5.75, quantile linear regression modeling analytics) package on predictions against chronological age from the normal population at 2.5%, 50% and 97.5%. The derived regression lines were overlaid with the predictions in both normal and CRC populations. Samples with predictions below the 2.5% quantile regression line was classified as “hypo-aging”, those with predictions above the 97.5% line as “hyper-aging”, and the rest (within 95% confidence interval) as normal aging, thus deriving three Δ age groups. Sample fractions in these Δ age groups were calculated in targeted categories of the interest.

Step 5: Structure identification

Supplementary Figure 3 (A). A histogram of all correlation (r) values obtained for the 157 annotated features. Black bars represent absolute correlation values $\geq$ 0.3, gray bars represent absolute correlation values < 0.3. The red dashed box represents the range of the nine features selected by the model.



Supplementary Figure 3 (B). The match scores derived from the MS/MS spectra comparison of the metabolic clock nine metabolite markers. mzCloud: https://www.mzcloud.org/; MassBank: https://massbank.jp/Index.

Mode	Compund Name	mzCloud Score	MassBank Score	Adduct	m/z
Postitive	Phenylalanine	94.7		M+ACN+H	207.1101
				M+H	166.0855
	Prolylleucine	75.4		M+H	229.1543
	Citrulline		99.4	M+H	176.1023
	Kynurenine	86.3		M+H	209.0917
Negative	Ornithine	77.6		M-H	131.0825
	3-(4-Hydroxyphenyl)lactate	83.7		M-H	181.0505
	Citrate		89.3	M-H	191.0197
	Gulonate		62.1	M-H	195.0519
	DHEAS	89.4		M-H	367.1580

To identify the structures of LCMS features, we combined two pools of features:

1. Features mapped to HMDB but not KEGG: We selected features with absolute correlation values (r) greater than a threshold of 0.5.
2. Features mapped to 42 age-associated pathways: We included all mapped features from these pathways.

The combined pool of features contained 137 features, which we subjected to structure characterization and identification. We were able to determine the structures of 49 of these features.

Step 6: Construction of the nine-metabolite-based aging clock

Through an elastic net regularized regression ($\alpha = 0.125$, and $\lambda = 0.129$), a metabolic aging clock was trained with the metabolite biomarker candidates. The penalty parameters of the elastic net were chosen via cross validations to minimize the mean squared prediction error. The importance score of each metabolite biomarker was measured by the absolute value of the estimated b coefficient of the normalized metabolite level. The higher the importance score suggests a bigger role in predicting age. With positive importance scores, nine metabolites were identified to as the panel ("metabolic aging clock") to regress to the chronological age, deriving the metabolic age for each healthy or CRC subject. Regarding the metabolite-based metabolic clock, all the final 9 biomarkers' univariate |r| are > 0.4.

Step 7: The nine-metabolite-based aging clock deviation analysis.

Same algorithmic process will be used as Step 4.

With the elastic net analysis, 9/49 metabolites were identified (beta coefficient not equal to 0) to model the metabolic aging using a bootstrapped general population of Shanghai CDC balancing age and gender (Figure 3E/3F/3G). The resulted elastic net model was defined as the “metabolic aging clock”. Comparative analysis of the enrolled individuals at Shanghai CDC and Fudan University cancer center revealed unique metabolic aging patterns associated with APL/CRC subjects (Figure 4A/4D/4G; Figure 5A/5C). A lower limit in our age prediction model is set at 30 years old.

To summarize, we only had one threshold (Pearson correlation, $|r| \geq 0.3$) to select LCMS features for downstream modeling analysis. Then we used two elastic net modeling processes to select pathways and structure-determined metabolic compounds for final linear modeling of chronological aging.

REFERENCES

1. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. Anal Chem. 2017; 89: 1063-7.

2. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27-30.

Supplementary Material 5

Supplementary tables

Supplementary Table 1A. Demographics showing metabolic aging clock normal Δ aging subgroup distributions

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	2860	290	377	277	264
Age group, N (%)					
(30,40]	15 (0.5)	14 (4.8)	7 (1.9)	2 (0.7)	15 (5.7)
(40,50]	469 (16.4)	47 (16.2)	62 (16.4)	43 (15.5)	37 (14)
(50,60]	1093 (38.2)	82 (28.3)	113 (30)	63 (22.7)	87 (33)
(60,70]	824 (28.8)	106 (36.6)	125 (33.2)	103 (37.2)	93 (35.2)
(70,80]	361 (12.6)	37 (12.8)	56 (14.9)	52 (18.8)	28 (10.6)
(80,90]	98 (3.4)	4 (1.4)	14 (3.7)	14 (5.1)	4 (1.5)
Sex, N (%)					
Male	1233 (43.1)	181 (62.4)	230 (61)	182 (65.7)	159 (60.2)
Female	1627 (56.9)	109 (37.6)	147 (39)	95 (34.3)	105 (39.8)

Supplementary Table 1B. Demographics showing metabolic aging clock hyper Δ aging subgroup distributions

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	76	10	3	15	13
Age group, N (%)					
(30,40]	5 (6.6)	0 (0.0)	0 (0.0)	4 (26.7)	1 (7.7)
(40,50]	22 (28.9)	0 (0.0)	1 (33.3)	0 (0.0)	5 (38.5)
(50,60]	16 (21.1)	3 (30.0)	2 (66.7)	3 (20.0)	2 (15.4)
(60,70]	6 (7.9)	4 (40.0)	0 (0.0)	1 (6.7)	4 (30.8)
(70,80]	16 (21.1)	0 (0.0)	0 (0.0)	7 (46.7)	1 (7.7)
(80,90]	11 (14.5)	3 (30.0)	0 (0.0)	0 (0.0)	0 (0.0)
Sex, N (%)					
Male	35 (46.1)	8 (80)	2 (66.7)	12 (80)	11 (84.6)
Female	41 (53.9)	2 (20)	1 (33.3)	3 (20)	2 (15.4)

Supplementary Table 1C. Demographics showing metabolic aging clock hypo Δ aging subgroup distributions

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	66	415	708	135	140
Age group, N (%)					
(30,40]	1 (1.5)	0 (0)	5 (0.7)	3 (2.2)	1 (0.7)
(40,50]	12 (18.2)	58 (14)	85 (12)	12 (8.9)	18 (12.9)
(50,60]	15 (22.7)	134 (32.3)	180 (25.4)	38 (28.1)	34 (24.3)
(60,70]	21 (31.8)	158 (38.1)	301 (42.5)	57 (42.2)	41 (29.3)
(70,80]	10 (15.2)	53 (12.8)	119 (16.8)	21 (15.6)	37 (26.4)
(80,90]	7 (10.6)	12 (2.9)	18 (2.5)	4 (3)	9 (6.4)
Sex, N (%)					
Male	44 (66.7)	221 (53.3)	391 (55.2)	78 (57.8)	64 (45.7)
Female	22 (33.3)	194 (46.7)	317 (44.8)	57 (42.2)	76 (54.3)

Supplementary Table 2A. Metabolic Δ aging subgroup membership, hyper/normal/hypo, determined by the nine-compound based metabolic aging clock

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	3002	715	1088	427	417
Δ age group, N (%)					
Hyper	76 (2.5)	10 (1.4)	3 (0.3)	15 (3.5)	13 (3.1)
Normal	2860 (95.3)	290 (40.6)	377 (34.7)	277 (64.9)	264 (63.3)
Hypo	66 (2.2)	415 (58)	708 (65.1)	135 (31.6)	140 (33.6)

Supplementary Table 2B. Metabolic Δ aging subgroup memberships, hyper/normal/hypo, determined by the pathway based multivariate analysis.

	Normal	APL	CRC - stage I	CRC - stage II	CRC - stage III
Total	3002	715	1088	427	417
Δ age group, N (%)					
Hyper	102 (3.4)	36 (5)	46 (4.2)	63 (14.8)	56 (13.4)
Normal	2818 (93.9)	552 (77.2)	811 (74.5)	330 (77.3)	334 (80.1)
Hypo	82 (2.7)	127 (17.8)	231 (21.2)	34 (8)	27 (6.5)

Supplementary Table 2C. Classification performance after bootstrapping cohort samples to the true incidence rate in the general population.

Class	Model name	True positive	True negative	False positive	False negative	Sensitivity	Specificity
CRC – APL	**Metabolic clock panel**	13388	268873	7067	9419	58.70%	97.44%
CRC - stage I~III	**Metabolic clock panel**	1028	268873	7067	922	52.72%	97.44%
CRC – APL	**Multi-target panel**	9532	271539	4401	13268	41.81%	98.41%
CRC - stage I~III	**Multi-target panel**	1201	271539	4401	737	61.97%	98.41%

Supplementary Table 3A. Demographics showing sample distributions across age and sex in CRC stage I population

Total	412
Age group, N (%)	
(30,40]	6 (1.5)
(40,50]	56 (13.6)
(50,60]	102 (24.8)
(60,70]	154 (37.4)
(70,80]	76 (18.4)
(80,90]	18 (4.4)
Sex, N (%)	
Male	250 (60.7)
Female	162 (39.3)

Supplementary Table 3B. Demographics showing sample distributions across CRC mutations in CRC stage I population (N=412)

Mutant	KRAS	NRAS	BRAF
Mutation, N (%)	164 (39.8)	14 (3.4)	18 (4.4)

Supplementary Table 4. %CV of the biomarker metabolite abundance in QC matrix samples which were spaced every 10 testing sera

Biomarkers	Within-run %CV	Between-day %CV	Total %CV
Ornithine	6.4	4.7	7.9
3(4-Hydroxyphenyl)lactate	7.7	3.9	8.6
Citrate	7.7	3.9	8.6
Gulonate	7.8	3.6	8.6
DHEAS	8.5	4.5	9.6
Citrulline	6.5	4.8	8
Phenylalanine	4.5	6.4	7.8
Kynurenine	9	5.4	10.6
Prolylleucine	5.2	5.7	7.7

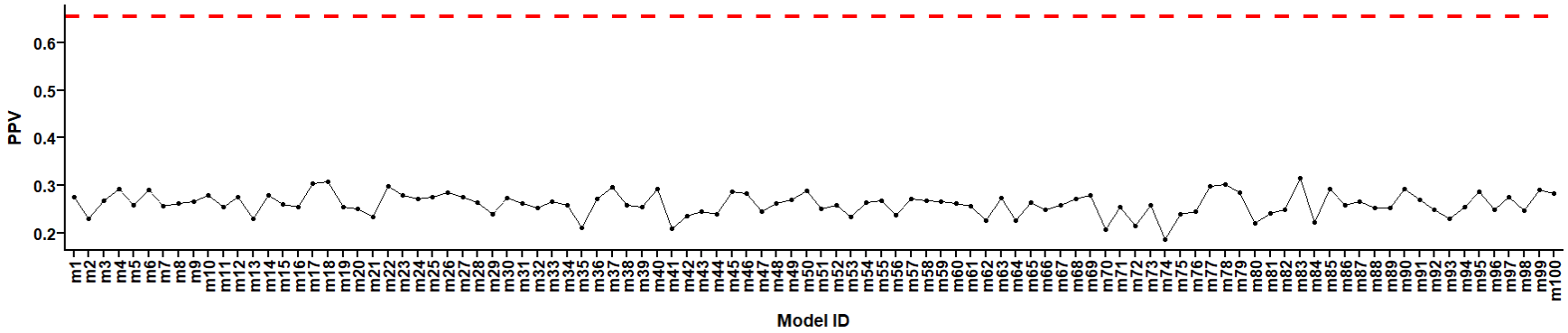
Supplementary Material 6

It is important to exclude potential technical bias in our study.

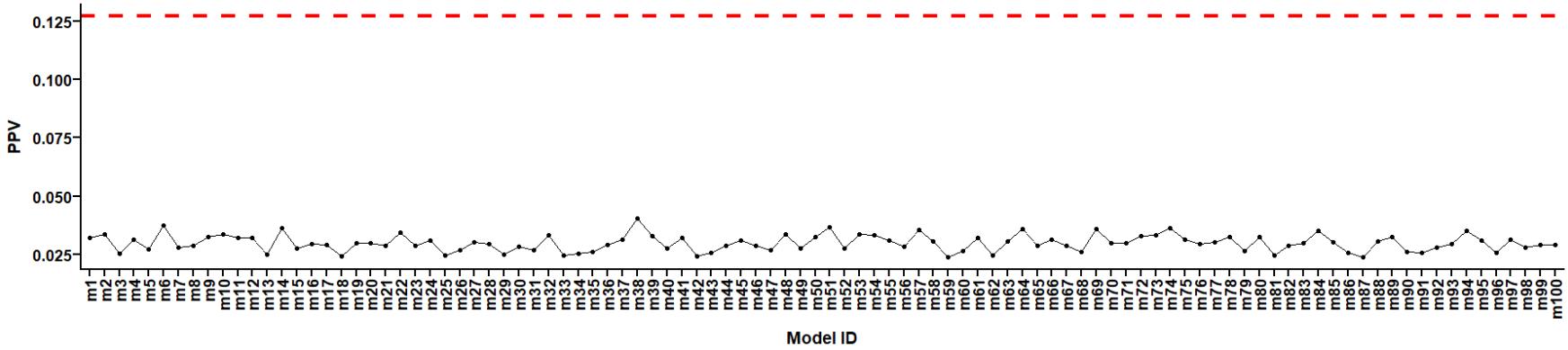
We have performed a follow up analysis in a subset of our cohort enrolled at Shanghai CDC, where identical blood collection was performed in CRC patients. We observed similar hypo-aging membership patterns in this subgroup of CRC patients, which supports the validity of our CRC results from Fudan University Shanghai Cancer Center. The following evidence was added in the discussion: “In a subset of our cohort enrolled at Shanghai CDC, identical blood collection was performed in healthy controls and CRC patients (n=55) who were identified as part of the CDC screening. Similar hypo-aging membership patterns were observed in this subgroup of CRC patients, (Figure. 5C-D), which supports the validity of our CRC results (Figure. 4A-B) from Fudan University Shanghai Cancer Center.

In addition, we developed 100 random models to assess whether the deviation of the CRC/APL cohort from the CI derived from CTRL is inherent to the age-related signature or due to other factors.

Supplementary Figure 4A shows the performance of 100 randomly built models to detect APL. All model PPVs are < 0.35, which is statistically lower (p-value < 0.05) than the original model (Table 2).



Supplementary Figure 4B shows the performance of 100 randomly built models to detect CRC. All model PPVs are < 0.35, which is statistically lower (p-value < 0.05) than the original model (Table 2).



In summary, our computational analysis results suggest that our results are biologically meaningful and statistically significant. Our findings are unlikely to be due to technical bias. This analysis supports our observations with the Shanghai CDC subset follow-up study,

Supplementary Material 7

Deep metabolomics profiling of healthy individuals revealed unique age-associated patterns in KEGG metabolic pathways. Each canonical pathway was divided into either a positive ("up") or negative ("down") age-correlated version based on the age correlation of its component compounds. An "up" pathway comprises significantly positive age-correlated pathway component compounds, while a "down" pathway comprises significantly negative age-correlated pathway component compounds. As detailed in Supplementary Material 3, the subject's value of either "up" or "down" pathway was calculated as the weighted sum of the normalized measurement values of the corresponding "up" or "down" metabolites on the pathway divided by the number of mapped metabolites.

Supplementary Table 5 provides a detailed breakdown of these pathways and their associated mapped metabolite features, including pathway name, KEGG ID, total number of metabolites in the pathway, number of annotated metabolites, number of significant metabolites, names of significant metabolites within the pathways, and pathway rank in the all-combined men and women cohort, as well as in the men and women cohorts separately.
Supplementary Table 5. Aging associated KEGG metabolic pathways and their associated mapped metabolomic features."

Pathway name	Direction	Total number of metabolites in pathway	Number of metabolites annotated	Number of metabolites significant	Names of significant metabolites within pathways	Pathway rank in all	Pathway rank in men	Pathway rank in women
map00140: Steroid hormone biosynthesis	down	99	5	1	Dehydroepiandrosterone sulfate;3beta-Hydroxyandrost-5-en-17-one 3-sulfate;DHEA sulfate; KEGG ID: C04555	1	1	2
map04976: Bile secretion	down	97	8	1	Dehydroepiandrosterone sulfate;3beta-Hydroxyandrost-5-en-17-one 3-sulfate;DHEA sulfate; KEGG ID: C04555	1	1	2

map02010: ABC transporters	up	137	45	12	D-Xylose;Wood sugar; KEGG ID: C00181 Hydroxyproline;L-Hydroxyproline;trans-4-Hydroxy-L-proline; KEGG ID: C01157 5-Aminolevulinate;5-Amino-4-oxopentanoate;5-Amino-4-oxovaleric acid; KEGG ID: C00430 Xylitol; KEGG ID: C00379 L-Arabinose;L-Arabinopyranose; KEGG ID: C00259 L-Ornithine;(S)-2,5-Diaminovaleric acid;(S)-2,5-Diaminopentanoic acid;(S)-2,5-Diaminopentanoate; KEGG ID: C00077 D-Methionine;D-2-Amino-4-(methylthio)butyric acid; KEGG ID: C00855 L-Cystine;L-Dicysteine;L-alpha-Diamino-beta-dithiolactic acid; KEGG ID: C00491 D-Sorbitol;D-Glucitol;L-Gulitol;Sorbitol; KEGG ID: C00794 Mannitol;D-Mannitol; KEGG ID: C00392 D-Ribose; KEGG ID: C00121 Uridine; KEGG ID: C00299	2	3	1
map00340: Histidine metabolism	down	47	17	1	3-(Imidazol-4-yl)-2-oxopropyl phosphate;Imidazole-acetol phosphate; KEGG ID: C01267	3	10	6
map00980: Metabolism of xenobiotics by cytochrome P450	down	121	11	1	N-Hydroxy-1-aminonaphthalene;1-Naphthylhydroxylamine; KEGG ID: C14789	4	4	14
map00740: Riboflavin metabolism	up	20	2	1	Ribitol;Adonitol; KEGG ID: C00474	5	12	4

map00360: Phenylalanine metabolism	up	60	26	8	Phenyllactate;(R)-Phenyllactate;(R)-3-(phenyl)lactate; KEGG ID: C05607 Benzoate;Benzoic acid;Benzenecarboxylic acid;Phenylformic acid;Dracylic acid; KEGG ID: C00180 cis-3-(3-Carboxyethenyl)-3,5-cyclohexadiene-1,2-diol;(2E)-3-(cis-5,6-Dihydroxycyclohexa-1,3-dien-1-yl)prop-2-enoate; KEGG ID: C12622 4-Hydroxy-2-oxopentanoate;4-Hydroxy-2-oxovalerate; KEGG ID: C03589 2-Hydroxy-2,4-pentadienoate;cis-2-Hydroxypenta-2,4-dienoate;Oxopent-4-enoate;2-Oxopent-4-enoate;2-Hydroxypenta-2,4-dienoate; KEGG ID: C00596 3-(2-Hydroxyphenyl)propanoate;2-Hydroxyphenylpropanoate;3-(2-Hydroxyphenyl)propionic acid;Melilotate;3-(2-Hydroxyphenyl)propionate; KEGG ID: C01198 3-(2,3-Dihydroxyphenyl)propanoate;2,3-Dihydroxyphenylpropanoate; KEGG ID: C04044 3-(3-Hydroxyphenyl)propanoic acid;Dihydro-3-coumaric acid;3-Hydroxyphenylpropanoate;3-(3-Hydroxyphenyl)propanoate; KEGG ID: C11457	6	2	NA
map05204: Chemical carcinogenesis	down	99	11	1	S-(1,2-Dichlorovinyl)-L-cysteine;DCVC; KEGG ID: C20317	6	9	8

map00020: Citrate cycle (TCA cycle)	up	20	7	4	Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158 Oxaloacetate;Oxalacetic acid;Oxaloacetic acid;2-Oxobutanedioic acid;2-Oxosuccinic acid;keto-Oxaloacetate; KEGG ID: C00036 Oxalosuccinate;Oxalosuccinic acid; KEGG ID: C05379 Isocitrate;Isocitric acid;1-Hydroxytricarballylic acid;1-Hydroxypropane-1,2,3-tricarboxylic acid; KEGG ID: C00311	7	NA	5
map00240: Pyrimidine metabolism	up	65	19	3	Uridine; KEGG ID: C00299 Pseudouridine; KEGG ID: C02067 (R)-5,6-Dihydrothymine;(5R)-Dihydrothymine; KEGG ID: C21028	8	8	9
map00130: Ubiquinone and other terpenoid-quinone biosynthesis	up	92	11	1	(R)-3-(4-Hydroxyphenyl)lactate; KEGG ID: C03964	9	NA	7
map00760: Nicotinate and nicotinamide metabolism	up	55	20	3	2-Methyleneglutarate;alpha-Methylene glutarate; KEGG ID: C02930 2,3-Dimethylmaleate;Dimethylmaleic acid; KEGG ID: C00922 Methylitaconate;2-Methylene-3-methylsuccinate; KEGG ID: C02295	10	NA	21
map00260: Glycine, serine and threonine metabolism	down	50	22	2	Creatine;alpha-Methylguanidino acetic acid;Methylglycocyamine; KEGG ID: C00300 L-Tryptophan;Tryptophan;(S)-alpha-Amino-beta-(3-indolyl)-propionic acid; KEGG ID: C00078	11	NA	NA
map00480: Glutathione metabolism	up	38	6	2	L-Ornithine;(S)-2,5-Diaminovaleric acid;(S)-2,5-Diaminopentanoic acid;(S)-2,5-Diaminopentanoate; KEGG ID: C00077	12	17	15

map00040: Pentose and glucuronate interconversions	up	55	26	20	Xylitol; KEGG ID: C00379 L-Lyxonate;L-Lyxonic acid; KEGG ID: C05412 Ribitol;Adonitol; KEGG ID: C00474 L-Gulonate;L-Gulonic acid;Gulonate;Gulonic acid; KEGG ID: C00800 L-Arabitol;L-Arabinol;L-Arabinitol;L-Lyxitol; KEGG ID: C00532 D-Arabitol;D-Arabinitol;D-Arabinol;D-Lyxitol; KEGG ID: C01904 (4R,5S)-4,5,6-Trihydroxy-2,3-dioxohexanoate;2,3-Diketo-L-gulonate; KEGG ID: C04575 D-Ribulose;D-erythro-2-Pentulose;D-Arabinoketose;D-Arabinulose;D-Riboketose; KEGG ID: C00309 D-Xylulose;D-threo-Pentulose;D-Lyxulose; KEGG ID: C00310 L-Ribulose;L-erythro-Pentulose;L-Arabinoketose;L-Arabinulose;L-Riboketose; KEGG ID: C00508 L-Lyxose; KEGG ID: C01508 D-Xylose;Wood sugar; KEGG ID: C00181 D-Mannonate; KEGG ID: C00514 L-Arabinose;L-Arabinopyranose; KEGG ID: C00259 L-Galactonate;L-Galactonic acid; KEGG ID: C15930 D-Lyxose; KEGG ID: C00476 D-Xylonate; KEGG ID: C00502 L-Xylulose;L-threo-Pentulose;L-Lyxulose; KEGG ID: C00312 L-Xylonate;L-Xylonic acid; KEGG ID: C05411 D-Altronate; KEGG ID: C00817	13	13	23

map00270: Cysteine and methionine metabolism	down	63	19	1	L-Homocystine;Homocystine; KEGG ID: C01817	14	NA	27
map00480: Glutathione metabolism	down	38	6	1	Bis-gamma-glutamylcystine;Oxidized gamma-glutamylcysteine;Bis-gamma-L-glutamyl-L-cystine;Oxidized gamma-L-glutamyl-L-cysteine; KEGG ID: C03646	14	NA	11
map04261: Adrenergic signaling in cardiomyocytes	up	10	1	1	Isoproterenol;Isoprenaline; KEGG ID: C07056	14	23	12
map00250: Alanine, aspartate and glutamate metabolism	down	28	14	1	2-Oxosuccinamate;2-Oxosuccinamic acid;gamma-Aminooxaloacetate;Oxaloacetamide; KEGG ID: C02362	15	NA	18
map00760: Nicotinate and nicotinamide metabolism	down	55	20	1	Iminoaspartate;Iminoaspartic acid;Iminosuccinate; KEGG ID: C05840	15	NA	18
map04728: Dopaminergic synapse	up	12	4	2	3-Methoxytyramine; KEGG ID: C05587 Homovanillate;Homovanillic acid;3-Methoxy-4-hydroxyphenylacetate; KEGG ID: C05582	16	NA	NA
map00270: Cysteine and methionine metabolism	up	63	19	2	L-Cystine;L-Dicysteine;L-alpha-Diamino-beta-dithiolactic acid; KEGG ID: C00491 L-Methionine;Methionine;L-2-Amino-4methylthiobutyric acid; KEGG ID: C00073	17	NA	NA
map04974: Protein digestion and absorption	up	47	24	2	L-Methionine;Methionine;L-2-Amino-4methylthiobutyric acid; KEGG ID: C00073 L-Cystine;L-Dicysteine;L-alpha-Diamino-beta-dithiolactic acid; KEGG ID: C00491	17	11	25
map00260: Glycine, serine and threonine metabolism	up	50	22	2	5-Hydroxyectoine; KEGG ID: C16432 5-Aminolevulinate;5-Amino-4-oxopentanoate;5-Amino-4-oxovaleric acid; KEGG ID: C00430	18	NA	36

map00330: Arginine and proline metabolism	down	78	34	2	N-Methylhydantoin;N-Methylimidazolidine-2,4-dione; KEGG ID: C02565 Creatine;alpha-Methylguanidino acetic acid;Methylglycocyamine; KEGG ID: C00300	19	NA	NA
map00520: Amino sugar and nucleotide sugar metabolism	up	108	10	2	L-Arabinose;L-Arabinopyranose; KEGG ID: C00259 D-Xylose;Wood sugar; KEGG ID: C00181	20	10	NA
map00310: Lysine degradation	up	54	20	5	6-Amino-2-oxohexanoate;2-Oxo-6-aminocaproate; KEGG ID: C03239 L-2-Aminoadipate 6-semialdehyde;2-Aminoadipate 6-semialdehyde;L-Allysine;Allysine;(S)-2-Amino-6-oxohexanoate; KEGG ID: C04076 2-Amino-5-oxohexanoate; KEGG ID: C05825 (S)-5-Amino-3-oxohexanoic acid;(S)-5-Amino-3-oxohexanoate; KEGG ID: C03656 Glutarate;Glutaric acid;Pentanedioic acid;1,3-Propanedicarboxylic acid; KEGG ID: C00489	21	6	34
map00250: Alanine, aspartate and glutamate metabolism	up	28	14	2	Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158 Oxaloacetate;Oxalacetic acid;Oxaloacetic acid;2-Oxobutanedioic acid;2-Oxosuccinic acid;keto-Oxaloacetate; KEGG ID: C00036	22	NA	NA
map00630: Glyoxylate and dicarboxylate metabolism	up	62	18	4	Isocitrate;Isocitric acid;1-Hydroxytricarballylic acid;1-Hydroxypropane-1,2,3-tricarboxylic acid; KEGG ID: C00311 trans-2,3-Epoxysuccinate; KEGG ID: C03548 Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158 Oxaloacetate;Oxalacetic acid;Oxaloacetic acid;2-	22	NA	28

					Oxobutanedioic acid;2-Oxosuccinic acid;keto-Oxaloacetate; KEGG ID: C00036			
map04742: Taste transduction	up	32	9	1	Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158	22	NA	38
map04922: Glucagon signaling pathway	up	26	8	3	Isocitrate;Isocitric acid;1-Hydroxytricarballylic acid;1-Hydroxypropane-1,2,3-tricarboxylic acid; KEGG ID: C00311 Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158 Oxaloacetate;Oxalacetic acid;Oxaloacetic acid;2-Oxobutanedioic acid;2-Oxosuccinic acid;keto-Oxaloacetate; KEGG ID: C00036	22	NA	NA

map01210: 2-Oxocarboxylic acid metabolism	up	134	39	8	Oxalosuccinate;Oxalosuccinic acid; KEGG ID: C05379 L-Ornithine;(S)-2,5-Diaminovaleric acid;(S)-2,5-Diaminopentanoic acid;(S)-2,5-Diaminopentanoate; KEGG ID: C00077 Isocitrate;Isocitric acid;1-Hydroxytricarballylic acid;1-Hydroxypropane-1,2,3-tricarboxylic acid; KEGG ID: C00311 L-Methionine;Methionine;L-2-Amino-4methylthiobutyric acid; KEGG ID: C00073 Citrate;Citric acid;2-Hydroxy-1,2,3-propanetricarboxylic acid;2-Hydroxytricarballylic acid; KEGG ID: C00158 (S)-2-Acetolactate;(S)-2-Hydroxy-2-methyl-3-oxobutanoate;(2S)-2-Hydroxy-2-methyl-3-oxobutanoate; KEGG ID: C06010 3-Hydroxy-3-methyl-2-oxobutanoic acid;3-Hydroxy-3-methyl-2-oxobutanoate;2-Oxo-3-hydroxyisovalerate; KEGG ID: C04181 Oxaloacetate;Oxalacetic acid;Oxaloacetic acid;2-Oxobutanedioic acid;2-Oxosuccinic acid;keto-Oxaloacetate; KEGG ID: C00036 (R)-2,3-Dihydroxy-3-methylpentanoate;(R)-2,3-Dihydroxy-3-methylvalerate;(2R,3R)-2,3-Dihydroxy-3-methylpentanoate; KEGG ID: C06007	23	5	31
map00280: Valine, leucine and isoleucine degradation	down	42	14	2	(S)-3-Methyl-2-oxopentanoic acid;(S)-3-Methyl-2-oxopentanoate;(3S)-3-Methyl-2-oxopentanoic acid;(3S)-3-Methyl-2-oxopentanoate; KEGG ID: C00671 4-Methyl-2-oxopentanoate;2-Oxoisocaproate; KEGG ID: C00233	24	16	26

map00290: Valine, leucine and isoleucine biosynthesis	down	23	14	2	(S)-3-Methyl-2-oxopentanoic acid;(S)-3-Methyl-2-oxopentanoate;(3S)-3-Methyl-2-oxopentanoic acid;(3S)-3-Methyl-2-oxopentanoate; KEGG ID: C00671 4-Methyl-2-oxopentanoate;2-Oxoisocaproate; KEGG ID: C00233	24	NA	NA
map00472: D-Arginine and D-ornithine metabolism	up	11	7	5	L-Ornithine;(S)-2,5-Diaminovaleric acid;(S)-2,5-Diaminopentanoic acid;(S)-2,5-Diaminopentanoate; KEGG ID: C00077 2-Amino-4-oxopentanoic acid;2-Amino-4-oxopentanoate;(2R)-2-Amino-4-oxopentanoate; KEGG ID: C03341 5-Amino-2-oxopentanoic acid;5-Amino-2-oxopentanoate;2-Oxo-5-amino-pentanoate;2-Oxo-5-aminopentanoate;alpha-Keto-delta-aminopentanoate;2-Oxo-5-aminovalerate; KEGG ID: C01110 D-Ornithine; KEGG ID: C00515 (2R,4S)-2,4-Diaminopentanoate;D-threo-2,4-Diaminopentanoate;2,4-Diaminopentanoate; KEGG ID: C03943	25	NA	20
map00380: Tryptophan metabolism	up	83	12	4	5-Hydroxyindoleacetate; KEGG ID: C05635 L-Kynurenine;3-Anthraniloyl-L-alanine; KEGG ID: C00328 2-Oxoindole-3-acetate;2-Oxoindole-3-acetic acid;2-Indolinone-3-acetate;2-(2-Oxo-1,3-dihydroindol-3-yl)acetate; KEGG ID: C22202 Formyl-5-hydroxykynurenamine; KEGG ID: C05647	26	NA	NA
map00982: Drug metabolism - cytochrome P450	up	87	12	1	5-Phenyl-1,3-oxazinane-2,4-dione; KEGG ID: C16596	26	NA	19

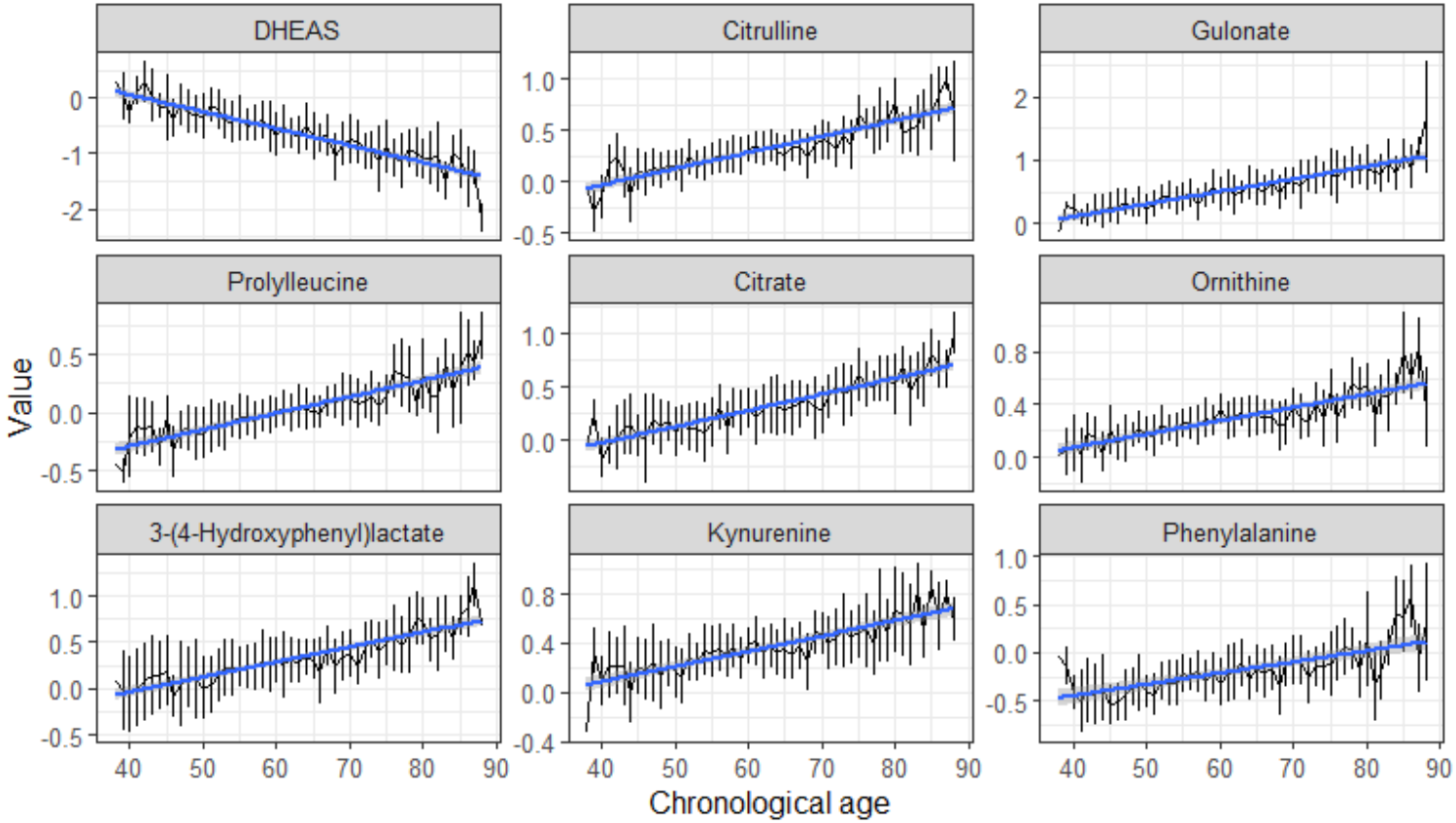
map04726: Serotonergic synapse	up	42	3	1	5-Hydroxyindoleacetate; KEGG ID: C05635	26	NA	19
map00330: Arginine and proline metabolism	up	78	34	9	cis-4-Hydroxy-D-proline; KEGG ID: C03440 5-Amino-2-oxopentanoic acid;5-Amino-2-oxopentanoate;2-Oxo-5-amino-pentanoate;2-Oxo-5-aminopentanoate;alpha-Keto-delta-aminopentanoate;2-Oxo-5-aminovalerate; KEGG ID: C01110 L-Glutamate 5-semialdehyde;L-Glutamate gamma-semialdehyde; KEGG ID: C01165 trans-3-Hydroxy-L-proline;trans-L-3-Hydroxyproline; KEGG ID: C05147 L-Ornithine;(S)-2,5-Diaminovaleric acid;(S)-2,5-Diaminopentanoic acid;(S)-2,5-Diaminopentanoate; KEGG ID: C00077 Hydroxyproline;L-Hydroxyproline;trans-4-Hydroxy-L-proline; KEGG ID: C01157 4-Acetamidobutanoate;N4-Acetylaminobutanoate; KEGG ID: C02946 cis-3-Hydroxy-L-proline;cis-3-Hydroxyproline; KEGG ID: C19706 Creatinine;1-Methylglycocyamidine; KEGG ID: C00791	27	NA	29
map00071: Fatty acid degradation	up	50	4	1	Glutarate;Glutaric acid;Pentanedioic acid;1,3-Propanedicarboxylic acid; KEGG ID: C00489	28	19	NA
map00650: Butanoate metabolism	up	42	13	1	(S)-2-Acetolactate;(S)-2-Hydroxy-2-methyl-3-oxobutanoate;(2S)-2-Hydroxy-2-methyl-3-oxobutanoate; KEGG ID: C06010	28	19	NA
map00750: Vitamin B6 metabolism	up	28	4	1	2-(Hydroxymethyl)-4-oxobutanoate;alpha-Hydroxymethyl succinate semialdehyde;2-Hydroxymethyl succinate semialdehyde; KEGG ID: C04106	28	19	NA

Supplementary Material 8

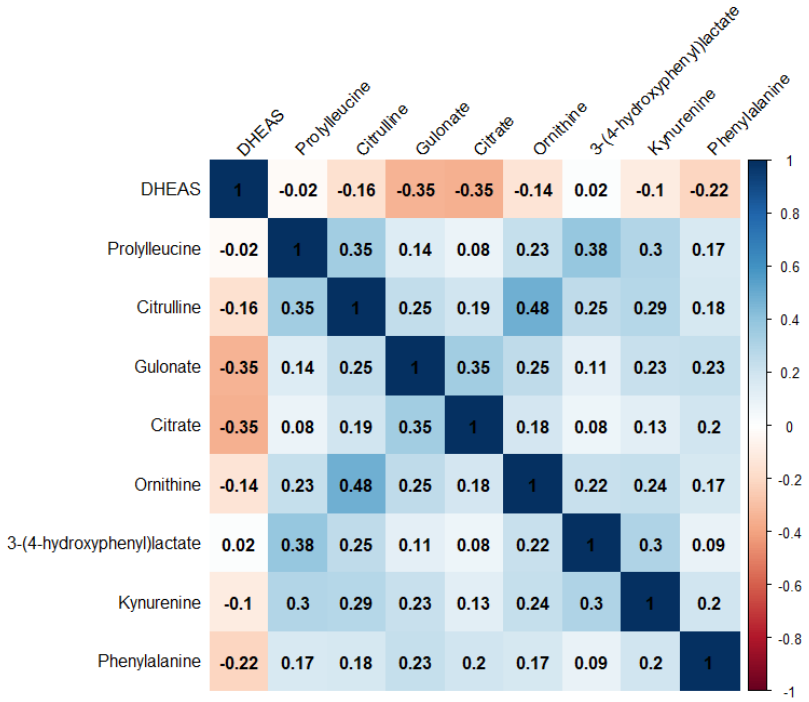
Elastic net regularized regression (ElasticNet) requires the following data prerequisites:

- Linearity: The relationship between the independent and dependent variables should be linear.
- Independence: The independent variables should be independent of each other.
- Normality: The dependent variable should be normally distributed.
- Homoscedasticity: The variance of the dependent variable should be equal across all levels of the independent variables.

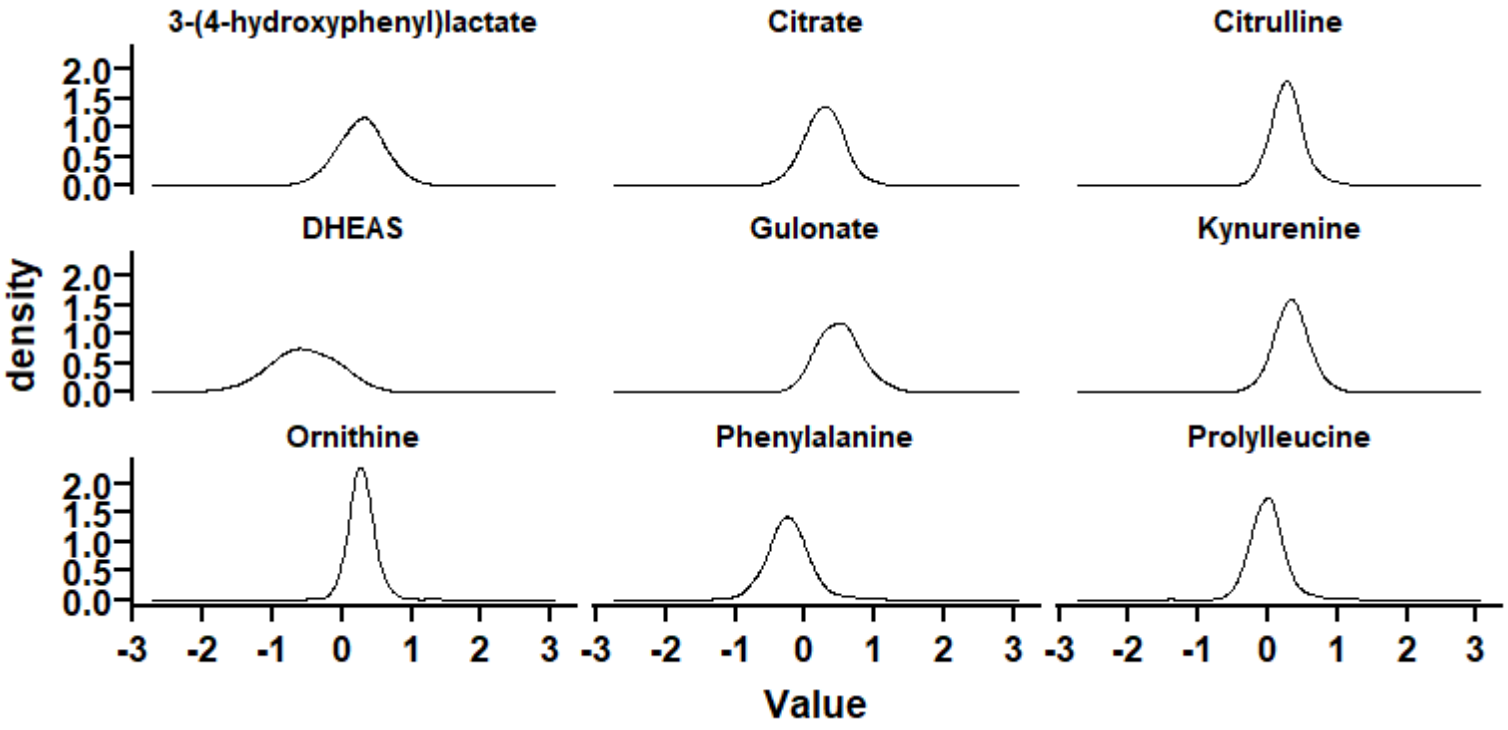
Supplementary Figure 5. Evidence to support the appropriateness to use of ElasticNet in this study.
A.. The validation for linearity has already been demonstrated in Figure 3C. All nine metabolic features used in the model exhibit good monotonicity.



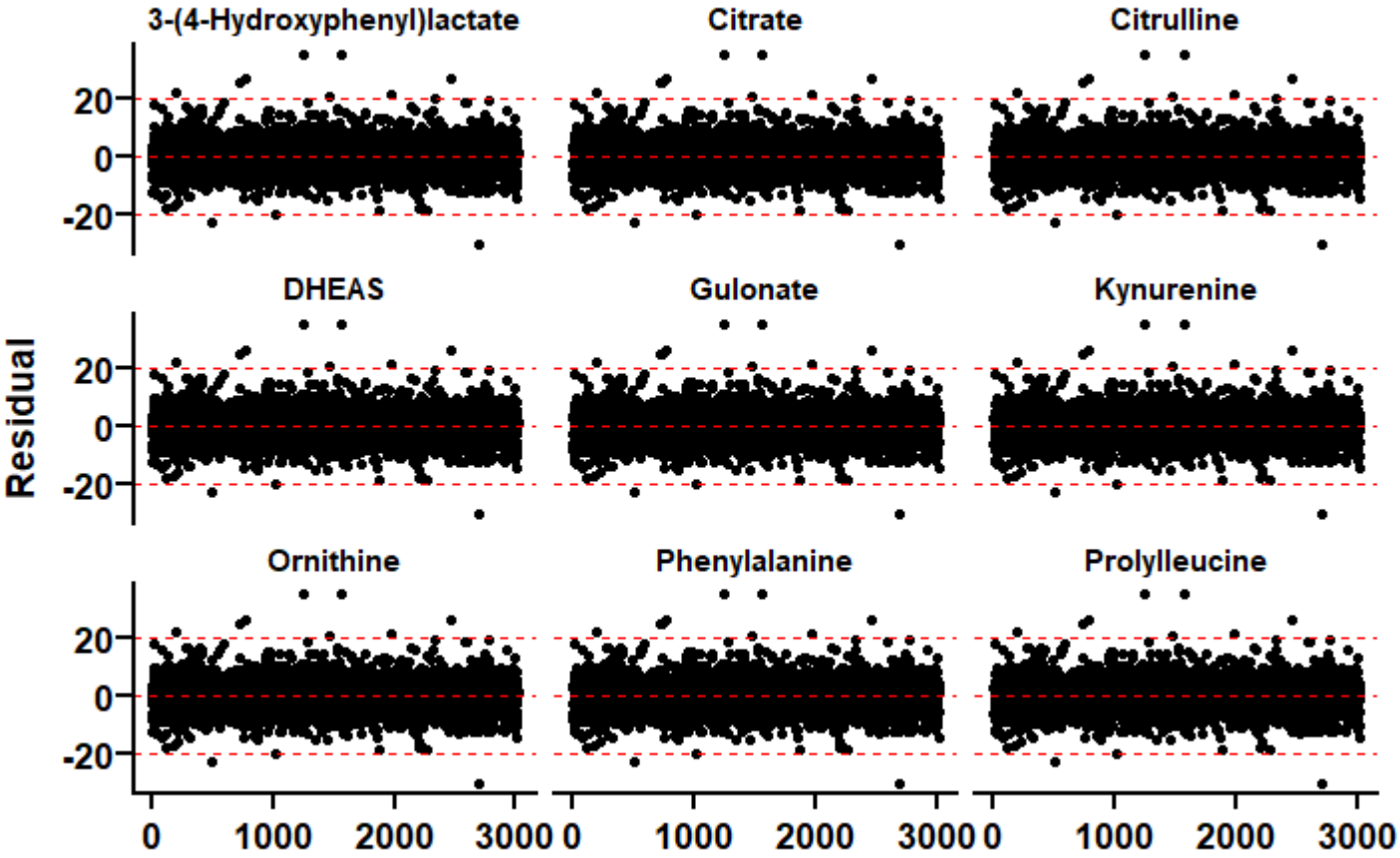
B. As shown in the heatmap below, the correlation coefficients between the features are all less than 0.5, indicating no significant inter-feature correlations and verifying the independence among the features. Furthermore, Elastic Net combines both L1 regularization (Lasso) and L2 regularization (Ridge), allowing it to simultaneously address variable selection (feature selection) and control model complexity. The presence of these two regularization terms enables Elastic Net to maintain model sparsity in the presence of multicollinearity. It selects the most important independent variables and pushes their coefficients towards zero while reducing the instability in coefficient estimation caused by multicollinearity.



C. For the distribution of features, we used density plots for validation. Our plots show that each of the distributions of the nine features closely follows a normal distribution.



D. To address feature homoscedasticity, we applied a logarithmic transformation during data standardization to ensure the homoscedasticity of features. Additionally, we plotted residual graphs for each feature. From these graphs, it can be observed that the residuals for each feature exhibit random distribution, constant variance, and no apparent trends. Consequently, our dataset meets the requirement of equal variance.



E. Regarding the issue of different gender distributions between the two datasets, we took into account the influence of gender on the model before modeling. Consequently, we independently established models for both the male and female populations and validated them on the cancer dataset. Therefore, we believe that the differing gender distributions will not affect the predictive results.