Online methods

Study participants

In our previously reported Guangdong Gut Microbiome Project[1], we elaborated on the details of the participants and survey methods. Briefly, samples that were selected randomly under a standard protocol using a probability proportional to size (PPS) sampling method were collected from 14 different districts in Guangdong Province[2]. The 14 districts were divided into developed and less developed city groups, with Guangzhou, Foshan, and Shenzhen constituting the developed city group and the rest of the cities being classified as the less developed city groups. After removing the samples obtained from those who did not complete the questionnaire and those with faecal sample sequences of less than 10,000 reads. In addition, a total of 6,376 samples were retained in this study after a preprocessing procedure was performed on the data. We further selected 1083 healthy individuals from the cohort as healthy controls. The inclusion criteria for healthy individuals were as follows: no reported illness; fasting blood glucose (FBG) levels of <6.1; BMI of <24; no antibiotic use within 1 month of stool sample donation; no clinical parameters indicating diagnoses of hypertension, hypertriglyceridaemia, hypercholesterolemia, neurosis, or chronic fatigue syndrome (CFS); and no medication for metabolic diseases intake before donating the stool sample. As the age composition differed significantly between the healthy control group and the different disease groups, we age-matched the different disease groups to the healthy control group using chronological age to avoid its effect as a strong confounding factor. The matched disease groups were as follows: hypertriglyceridaemia (n=537), hypercholesterolemia (n=571), metabolic syndrome (MetS) (n=695), diabetes (n=445), hypertension (n=688) and obesity (n=585).

Sample processing

We described the sample processing method in our previous study[1]. In brief, we extracted total bacterial DNA from stool samples. The barcoded primers (5′ to 3′) by which the 16S rRNA gene V4 region was amplified were V4F, GTGYCAGCMGCCGCGGTAA and V4R, GGACTACNVGGGTWTCTAAT. The PCR amplification conditions were as follows: initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 30 s, extension at 72 °C for 45 seconds, and a final elongation at 72 °C for 5 min. Then, the PCR products were sent to Beijing Genome Institute (BGI; Beijing, China), and next-generation sequencing was performed on an Illumina HiSeq 2500 platform.

Bioinformatics and biostatistics

The method used for the preprocessing of the raw sequences was described in our previous study[1]. Briefly, using QIIME V.1.9.1, we merged the data, performed quality control assessments and analysed the data at the sequence level using the reference-free statistical denoising deblur method to prevent improperly identifying distinct operational taxonomic unit-based data originating from misclustered sequences[2, 3].

To establish a microbiome model to predict the age of the host's microbiota, we adopted the relative abundance data at the ASV level of 1083 relatively healthy people to train a random forest regression model by using the R packages ‘randomForest’ (version 4.6.12) and ‘caret’ (version 6.0.68) to tune hyperparameters and perform 10-fold cross-validation. The sample matching process was performed using the R package 'MatchIt'. It used the propensity score method (PSM) and greedy nearest neighbour matching to achieve pairwise matching of participants in different disease groups with their healthy control counterparts. Spearman's rank correlation test was used to analyse the correlation between amplicon sequence variant (ASVs) and host chronological age, and then $P \leq 0.05$ was used as the threshold to filter out the ASVs related to chronological age for modeling. The interaction effect between chronological age and different metabolic disorders was indicated by the significance of the interaction term in the multivariable linear regression model (Relative abundance of the ASV or the microbiota age ~ chronological age + disease group + chronological age*disease group). All plots in this study were drawn using the R package ‘ggplot2’ and ‘pheatmap’, and the statistical indicators in the plots were calculated using the R packages ‘ggpubr’ and ‘ggpmisc’.

Ethics approval

Ethical approvals were described previously[1]. The Guangdong gut microbiome project was approved by the ethical review committee of the Chinese Center for Disease Control and Prevention (No. 201519-A). Written consent was obtained from all participants.

Patient and public involvement

Neither patients nor the public were involved in the design, conduction, reporting, or dissemination plans of our research.

References for online methods

[1] Y. He, W. Wu, H.-M. Zheng, P. Li, D. McDonald, H.-F. Sheng, M.-X. Chen, Z.-H. Chen, G.-Y. Ji, Z.-D.-X. Zheng, P. Mujagond, X.-J. Chen, Z.-H. Rong, P. Chen, L.-Y. Lyu, X. Wang, C.-B. Wu, N. Yu, Y.-J. Xu, J. Yin, J. Raes, R. Knight, W.-J. Ma, and H.-W. Zhou, Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nature Medicine, 2018. 24(10): 1532-1535.

[2] Lao X Q , Ma W J , Sobko T , et al. Overall obesity is leveling-off while abdominal obesity continues to rise in a Chinese population experiencing rapid economic development: analysis of serial cross-sectional health survey data 2002–2010[J]. International Journal of Obesity.

[3] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Tumbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight, QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 2010. 7(5): 335-336.

[4] A. Amir, D. McDonald, J.A. Navas-Molina, E. Kopylova, J.T. Morton, Z. Zech Xu, E.P. Kightley, L.R. Thompson, E.R. Hyde, A. Gonzalez, and R. Knight, Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems, 2017. 2(2).

Supplementary tables

Table S1. Characteristics of the participants included in this study.

Characteristic	**Healthy**, N = 1,083	**Diabetes**, N = 555	**Hypercholesterolemia**, N = 775	**Hypertension**, N = 2,289	**Hypertriglyceridemia**, N = 674	**MetS**, N = 1,254	**Obesity**, N = 633	**p-value**[1]
Age, Median (IQR)	43 (31 – 54)	59 (52 – 66)	57 (49 – 65)	62 (53 – 70)	54 (45 – 63)	58 (49 – 66)	53 (44 – 62)	<0.001
Gender(Female), n (%)	638 (59)	298 (54)	431 (56)	1,194 (52)	297 (44)	601 (48)	379 (60)	<0.001
Weight(kg), Median (IQR)	53 (48 – 58)	61 (54 – 68)	58 (51 – 66)	59 (52 – 67)	64 (57 – 72)	66 (59 – 74)	74 (68 – 81)	<0.001
BMI, Median (IQR)	21.2 (19.6 – 22.5)	24.4 (22.3 – 27.1)	23.4 (21.3 – 25.9)	24.1 (21.8 – 26.5)	25.1 (22.9 – 27.6)	26.1 (24.1 – 28.4)	29.4 (28.6 – 31.2)	<0.001
Waist(cm), Median (IQR)	73 (68 – 78)	85 (79 – 92)	81 (75 – 88)	84 (77 – 90)	86 (80 – 92)	90 (85 – 95)	95 (90 – 100)	<0.001
FBG(mmol/L), Median (IQR)	5.04 (4.68 – 5.34)	8.06 (7.19 – 10.32)	5.47 (5.05 – 6.06)	5.52 (5.08 – 6.14)	5.74 (5.28 – 6.50)	6.11 (5.35 – 6.98)	5.52 (5.07 – 6.28)	<0.001
TCHO(mmol/L), Median (IQR)	5.00 (4.52 – 5.46)	5.39 (4.92 – 5.93)	6.56 (6.35 – 6.86)	5.36 (4.85 – 5.90)	5.40 (4.89 – 5.99)	5.32 (4.80 – 5.89)	5.28 (4.80 – 5.82)	<0.001
TG(mmol/L), Median (IQR)	0.81 (0.64 – 1.10)	1.53 (1.09 – 2.32)	1.30 (0.91 – 1.96)	1.22 (0.85 – 1.80)	3.15 (2.62 – 4.10)	2.06 (1.53 – 2.94)	1.48 (1.05 – 2.23)	<0.001
UA(μmol/L), Median (IQR)	304 (258 – 359)	322 (269 – 401)	329 (273 – 398)	344 (284 – 412)	390 (312 – 459)	368 (309 – 442)	365 (311 – 438)	<0.001
HDL(mmol/L), Median (IQR)	1.31 (1.10 – 1.56)	1.07 (0.89 – 1.34)	1.33 (1.13 – 1.59)	1.21 (0.99 – 1.48)	0.89 (0.73 – 1.08)	0.94 (0.81 – 1.08)	1.07 (0.87 – 1.29)	<0.001
LDL(mmol/L), Median (IQR)	2.88 (2.40 – 3.37)	3.39 (2.75 – 4.10)	4.31 (3.65 – 4.95)	3.41 (2.80 – 4.02)	3.53 (2.87 – 4.23)	3.43 (2.84 – 4.10)	3.49 (2.81 – 4.08)	<0.001
HbA1c(%), Median (IQR)	4.70 (4.30 – 5.10)	6.10 (5.20 – 7.60)	5.10 (4.60 – 5.50)	5.00 (4.60 – 5.50)	5.00 (4.60 – 5.60)	5.20 (4.70 – 5.80)	5.10 (4.60 – 5.50)	<0.001
Hb(g/L), Median (IQR)	142 (129 – 153)	143 (129 – 156)	144 (131 – 157)	143 (130 – 156)	148 (135 – 162)	147 (133 – 159)	145 (134 – 158)	<0.001
ALT(U/L), Median (IQR)	13 (10 – 18)	17 (12 – 26)	17 (12 – 23)	16 (12 – 23)	20 (14 – 30)	19 (13 – 27)	19 (14 – 29)	<0.001
BUN(mmol/L), Median (IQR)	4.82 (4.04 – 5.68)	5.36 (4.34 – 6.38)	5.42 (4.60 – 6.29)	5.39 (4.52 – 6.38)	5.08 (4.36 – 6.05)	5.13 (4.34 – 6.16)	5.03 (4.20 – 6.03)	<0.001
Antibotics use within 1 month, n (%)	0 (0)	42 (7.9)	48 (6.4)	119 (5.3)	52 (7.9)	79 (6.5)	38 (6.1)	<0.001
Medication use for metabolic diseases, n (%)	0 (0)	184 (34)	126 (17)	542 (24)	141 (22)	328 (27)	131 (21)	<0.001

[1] Kruskal-Wallis rank sum test for continuous variables; Pearson's Chi-squared test for categorical variables

Table S2. The associations of potential covariates with microbiota age in the 1083 healthy participants were measured by the multivariable linear regression model.

Covariate	Forest Plot	Beta[1]	SE[2]	95% CI[2]	p-value
Chronological Age		0.37***	0.030	0.31, 0.43	<0.001
Sleep time per day		0.02	0.028	-0.04, 0.07	0.543
BMI		-0.01	0.028	-0.07, 0.04	0.636
Grains		0.06*	0.029	0.00, 0.11	0.050
Fruits		-0.11***	0.030	-0.16, -0.05	<0.001
Fruit drinks		-0.03	0.031	-0.09, 0.03	0.410
Vegetables		0.01	0.030	-0.05, 0.06	0.839
Livestock meat		-0.01	0.030	-0.07, 0.05	0.654
Carbonated beverage		0.04	0.031	-0.02, 0.10	0.233
High alcohol liquor		0.03	0.028	-0.02, 0.09	0.237
Low alcohol liquor		0.00	0.032	-0.07, 0.06	0.922
Beer		-0.03	0.031	-0.09, 0.03	0.337
Yellow rice wine		-0.03	0.028	-0.09, 0.02	0.251
Rice wine		0.07*	0.032	0.01, 0.13	0.028
Wine		-0.05	0.031	-0.11, 0.01	0.126

[1] Beta is the standardized regression coefficient in the linear regression model.
*p<0.05; **p<0.01; ***p<0.001

[2] SE = Standard Error, CI = Confidence Interval

Supplementary figures

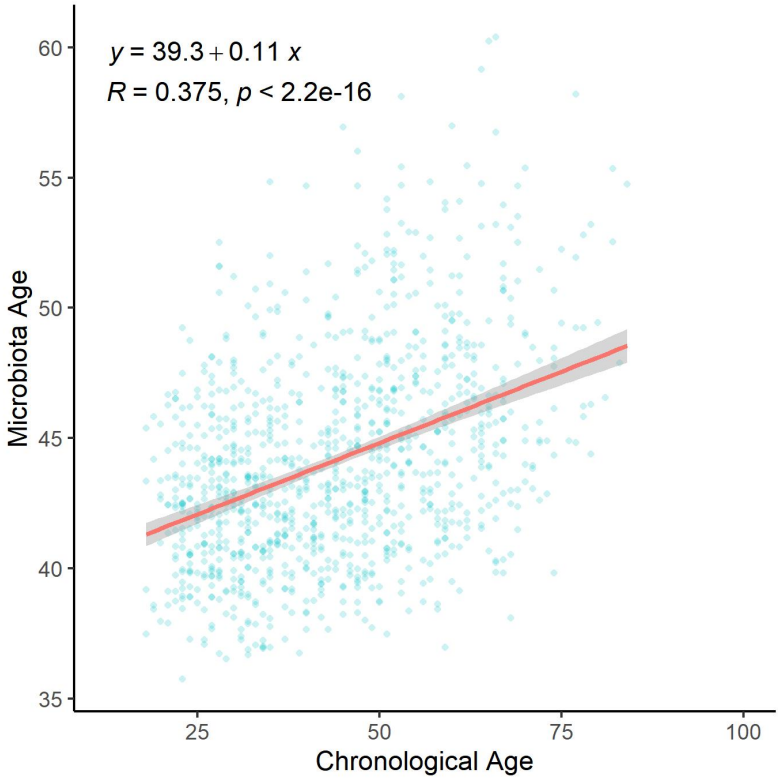

Figure S1. The scatter plot showed a significant and positive correlation between the microbiota age of 1083 healthy individuals predicted by the random forest model and their chronological age. R, Spearman correlation coefficient.

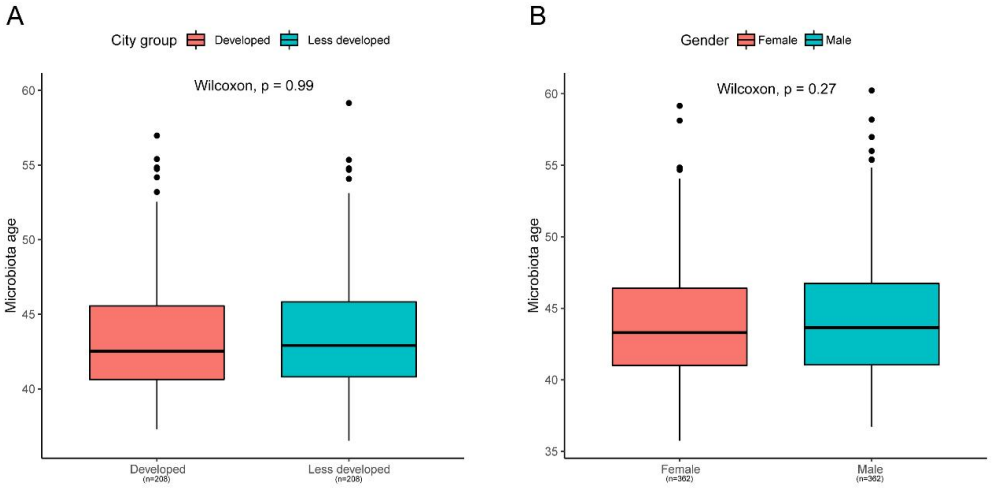

Figure S2. Comparing gut microbiota age between geographic areas (A) and gender (B) in the healthy individuals. Chronological ages were matched between different groups using the propensity score method.

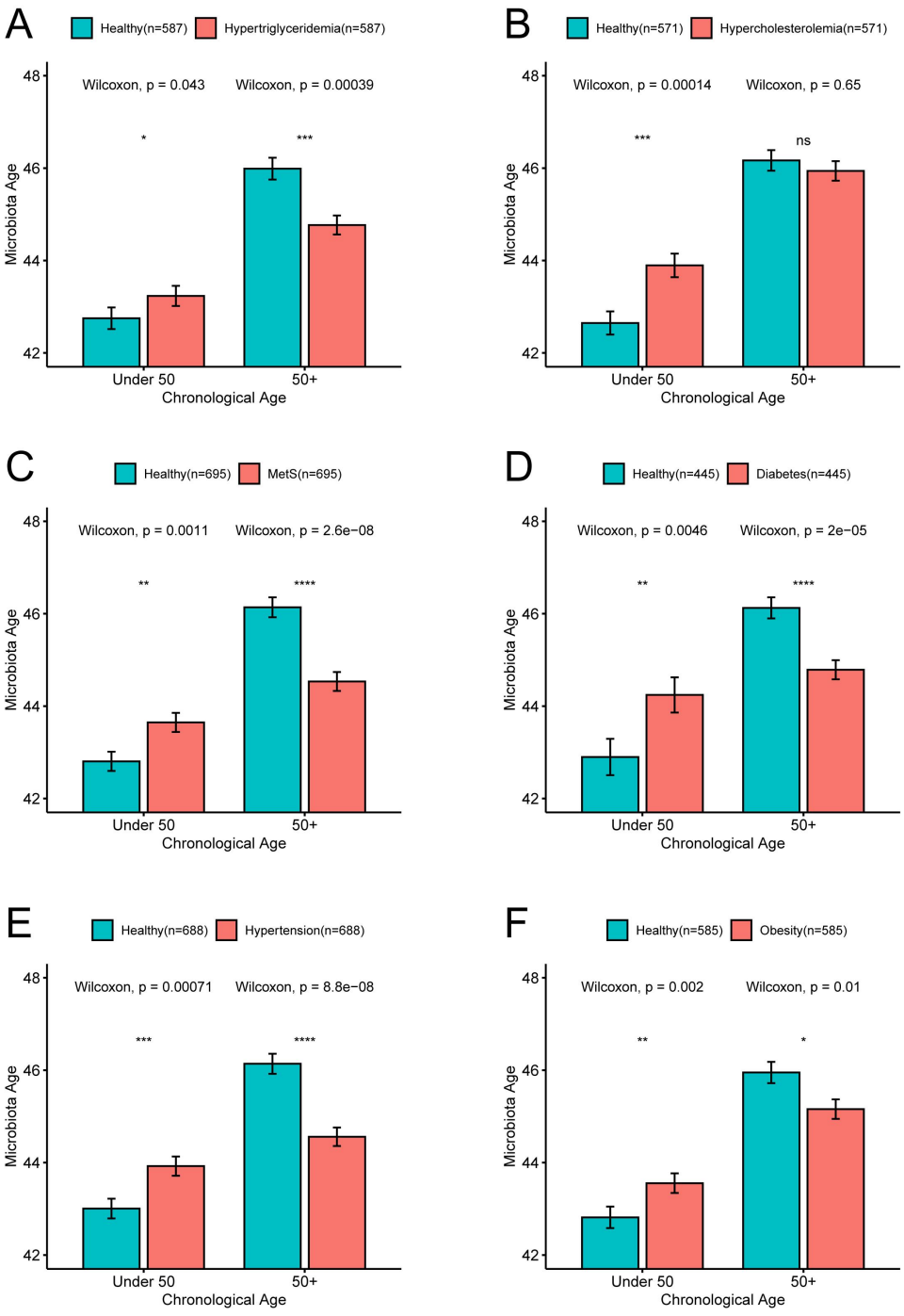

Figure S3. Comparing microbiota age between people with different metabolic disorders and healthy individuals. Figures S2(A) to S2(F) illustrate the comparisons between groups with hypertriglyceridaemia, hypercholesterolemia, metabolic syndrome (MetS), diabetes, hypertension, obesity and healthy individuals, respectively. The "under50" and "50+" labels indicate that the chronological age of the participant in their corresponding group is less than 50 years and greater than or

equal to 50 years, respectively. Significantly different groups are indicated with **** for $P \leqslant 0.0001$, *** for $P \leqslant 0.001$, ** for $P \leqslant 0.01$, * for $P \leqslant 0.05$ or ns for $P > 0.05$ (two-sided Wilcoxon test).

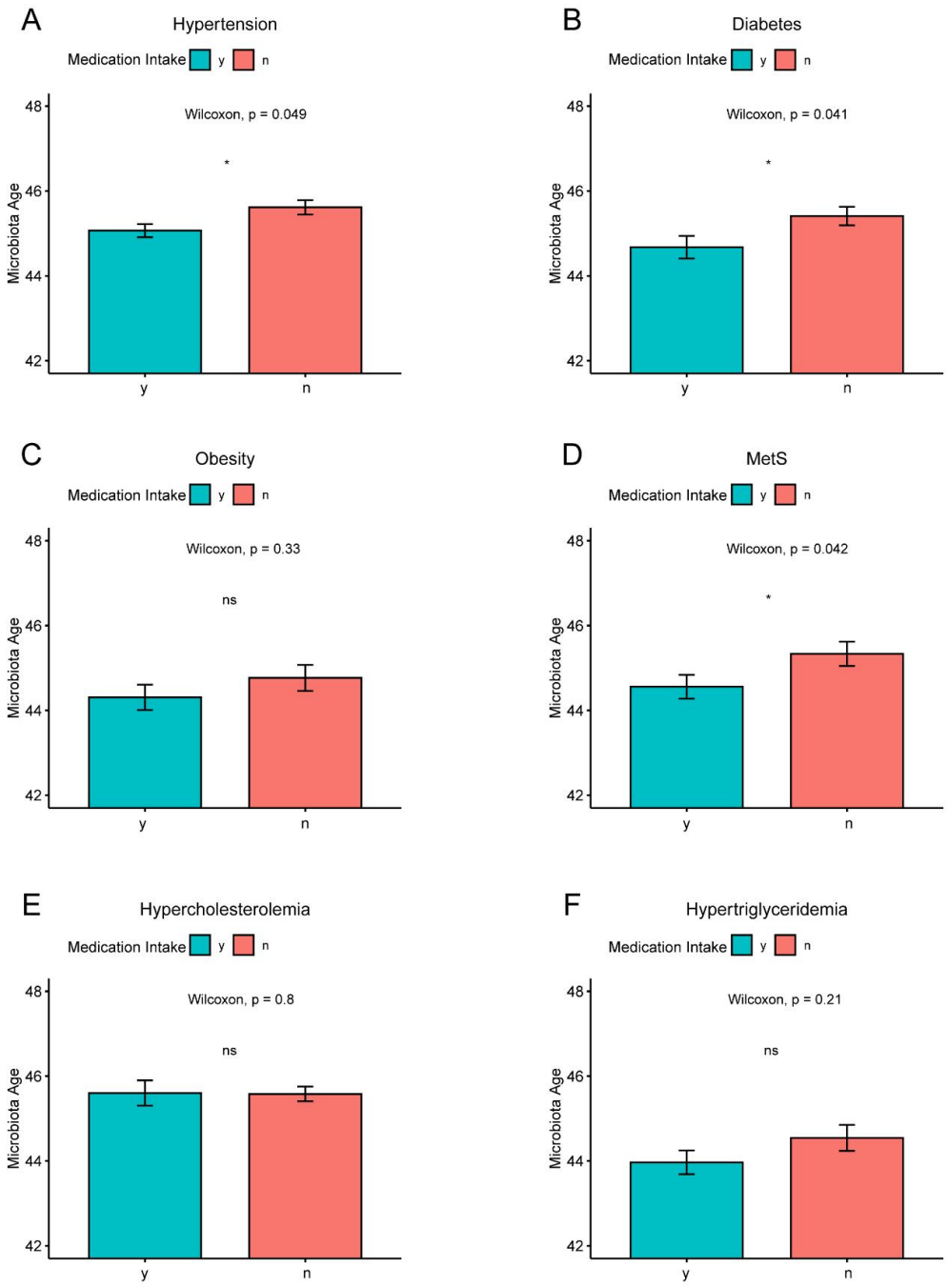

Figure S4. Comparison of gut microbiota age between metabolic diseases patients who took medication and those who are unmedicated. Figures S3(A) to S3(F) show the differences in microbiota age of individuals with hypertension, diabetes, obesity, metabolic syndrome (MetS), hypercholesterolemia and hypertriglyceridemia respectively after matching for chronological age between the two groups. Significantly different groups are indicated with **** for P⩽ 0.0001, *** for P ⩽ 0.001, ** for P ⩽ 0.01, * for P ⩽ 0.05 or ns for P > 0.05 (two-sided Wilcoxon test).

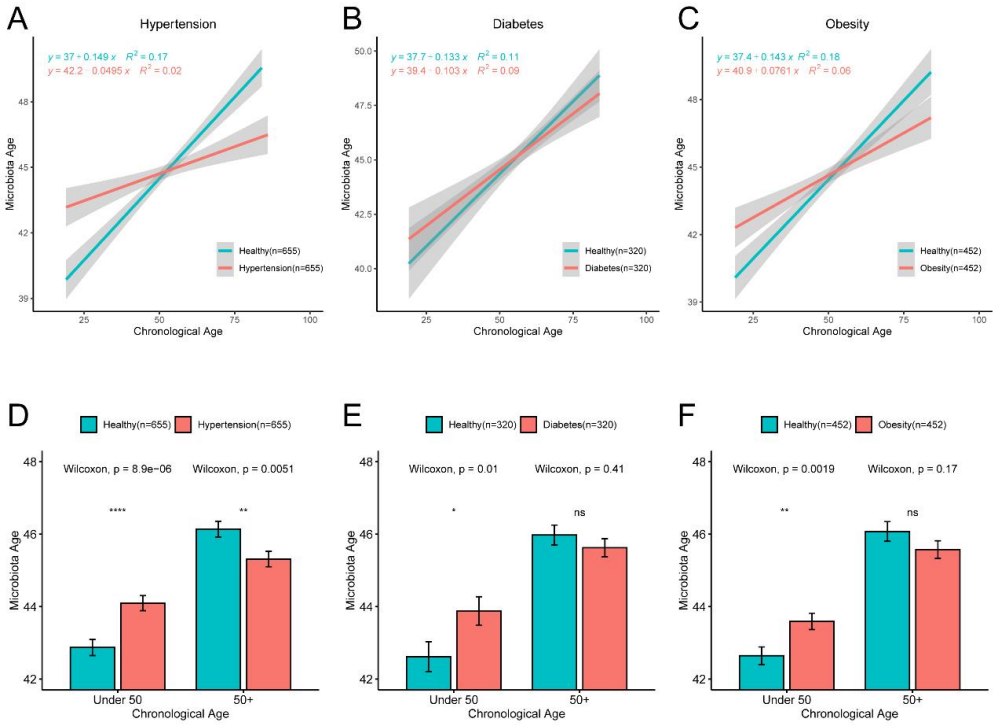

Figure S5. Reproducing the intersecting pattern between gut microbiota age and metabolic diseases after excluding medicated patients. In order to avoid the effect of antibiotic use and intake of medication on our results, we excluded subjects who used antibiotics within one month and used medication for metabolic diseases before sample collection in different disease groups. Meanwhile, patients' chronological age was matched with healthy individuals. Figures S3(A) to S3(C) illustrate the microbiota ageing trajectories of chronological age-matched participants in hypertension, diabetes and obesity respectively. Significantly different groups are indicated with **** for P⩽ 0.0001, *** for P ⩽ 0.001, ** for P ⩽ 0.01, * for P ⩽ 0.05 or ns for P > 0.05 (two-sided Wilcoxon test).

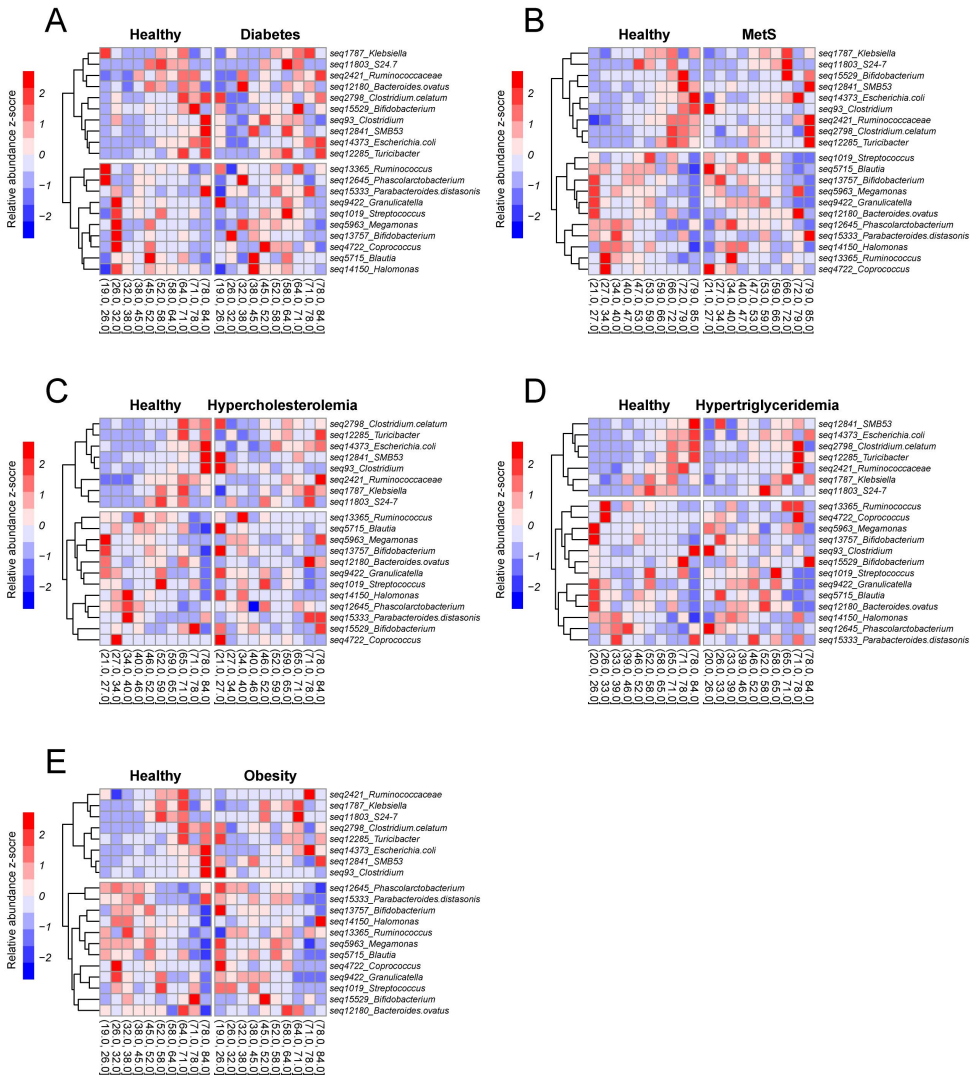

Figure S6. The aging trajectories of gut microbes in participants with different metabolic diseases and healthy individuals. The healthy and diseased groups were matched by chronological ages.