

## Appendix 1 Supplementary methods: power estimation

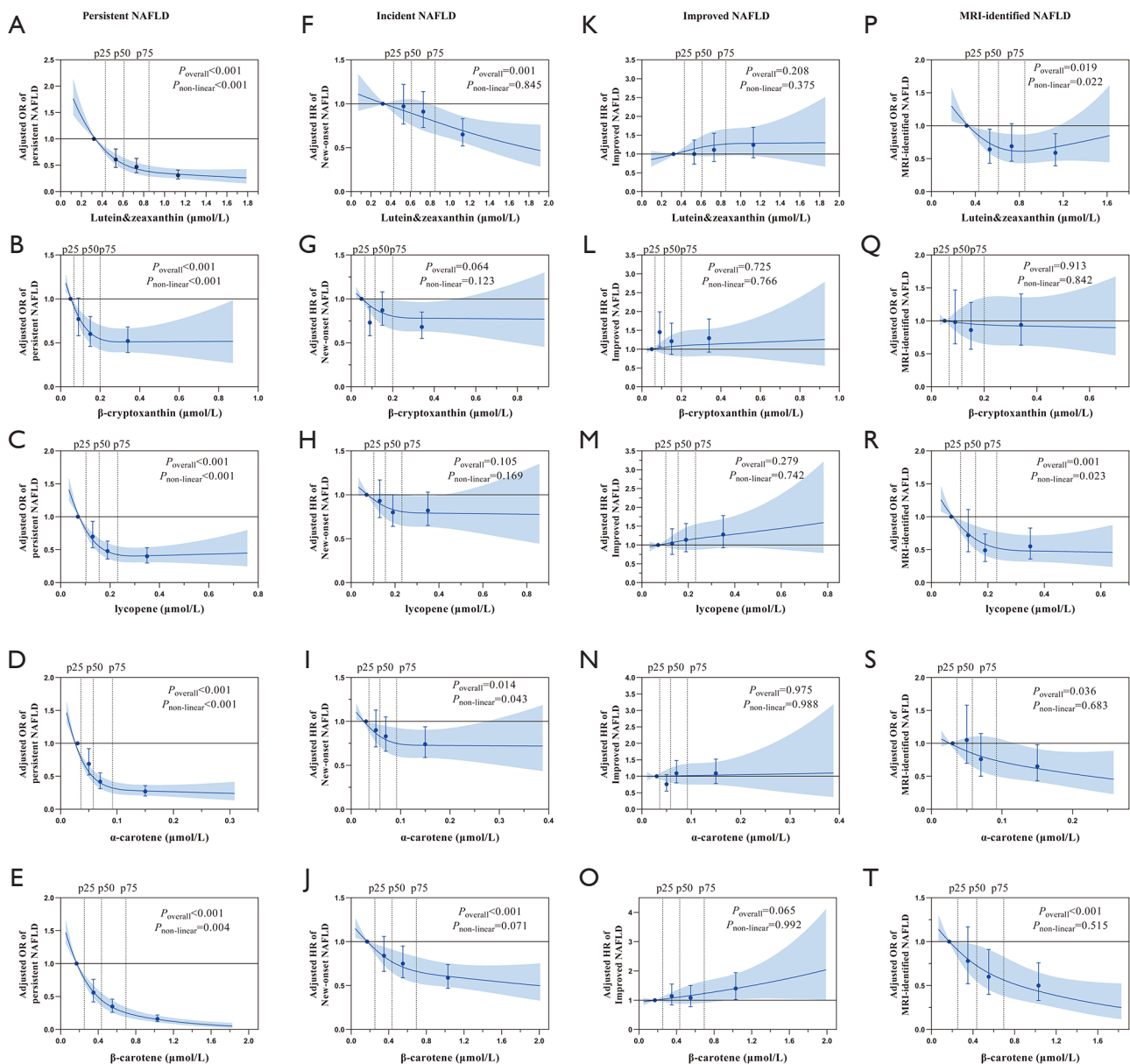
We considered two aspects for power estimation and used the distribution of the between-group rates in our actual analysis data to estimate the effect size. The detailed methods are as follows.

(I) Serum carotenoids and persistent-NAFLD association: We calculated power using multiple group proportion comparisons. Take persistent-NAFLD as an example, for total carotenoids, with odds ratios (ORs) of 0.56, 0.33, and 0.20 in quartiles Q2-Q4 compared to Q1, and considering the persistent-NAFLD (*vs.* NAFLD-free) cases/total N in each quartile (Q1: 325/434, Q2: 268/431, Q3: 217/435, Q4: 167/447), we achieved a power exceeding 0.96 for Q2-Q4 against Q1, at a 0.05 alpha level, as computed using PASS 15.

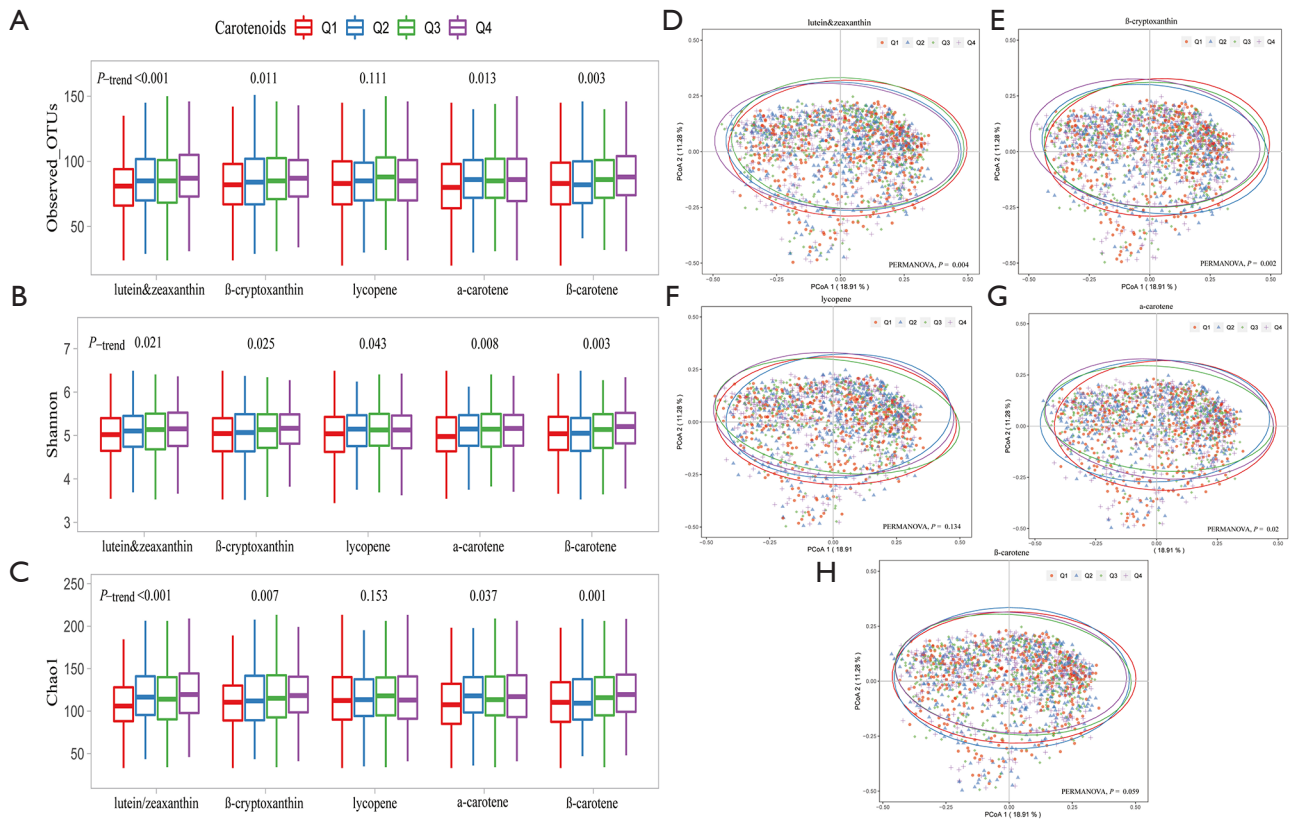
(II) Microbiome and metabolites analysis: For power estimation in the microbiome and metabolites analysis, we employed the Power Analysis Module on MetaboAnalyst (1). This method takes into account three critical factors: population effect size, sample size for detecting the effect, and the statistical significance threshold. Effect size and sample size were estimated from our omics data based on the distribution of NAFLD-free *vs.* persistent/incident-NAFLD group, with false discovery rate (FDR) serving as the significance criterion for high-dimensional omics data. At FDR =0.25 and with a maximum group sample size of 800 (NAFLD-free *vs.* incident/persistent-NAFLD group), we achieved a power of around 0.7 for 116 microbiome genera (total number in our study). With FDR =0.05 and a maximum group sample size of 400 (NAFLD-free *vs.* incident/persistent-NAFLD group), the power approximates 0.7 for metabolites with VIP values above 1.0 (all metabolites including in regression analysis).

## References

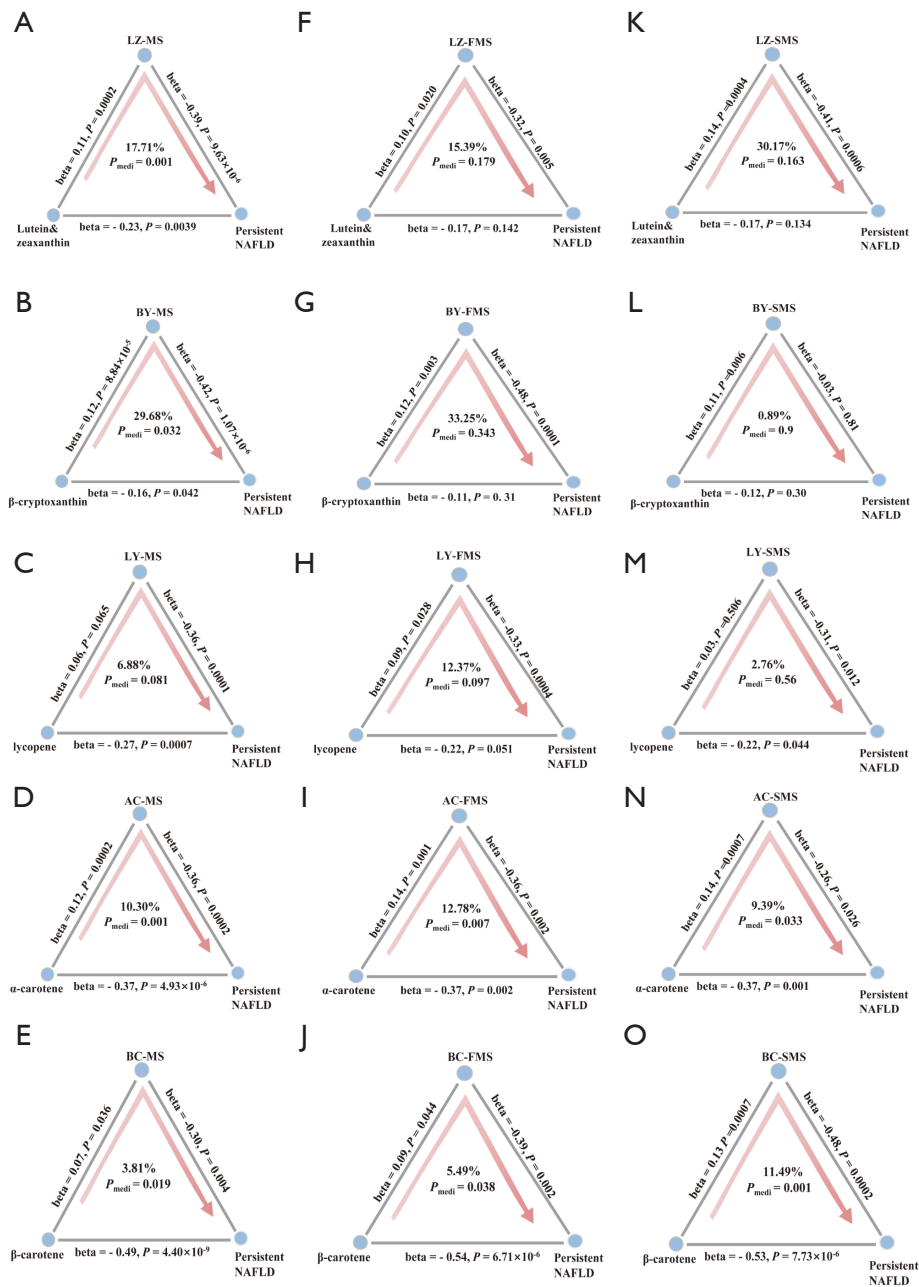
1. Pang Z, Zhou G, Ewald J, et al. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* 2022;17:1735-61.



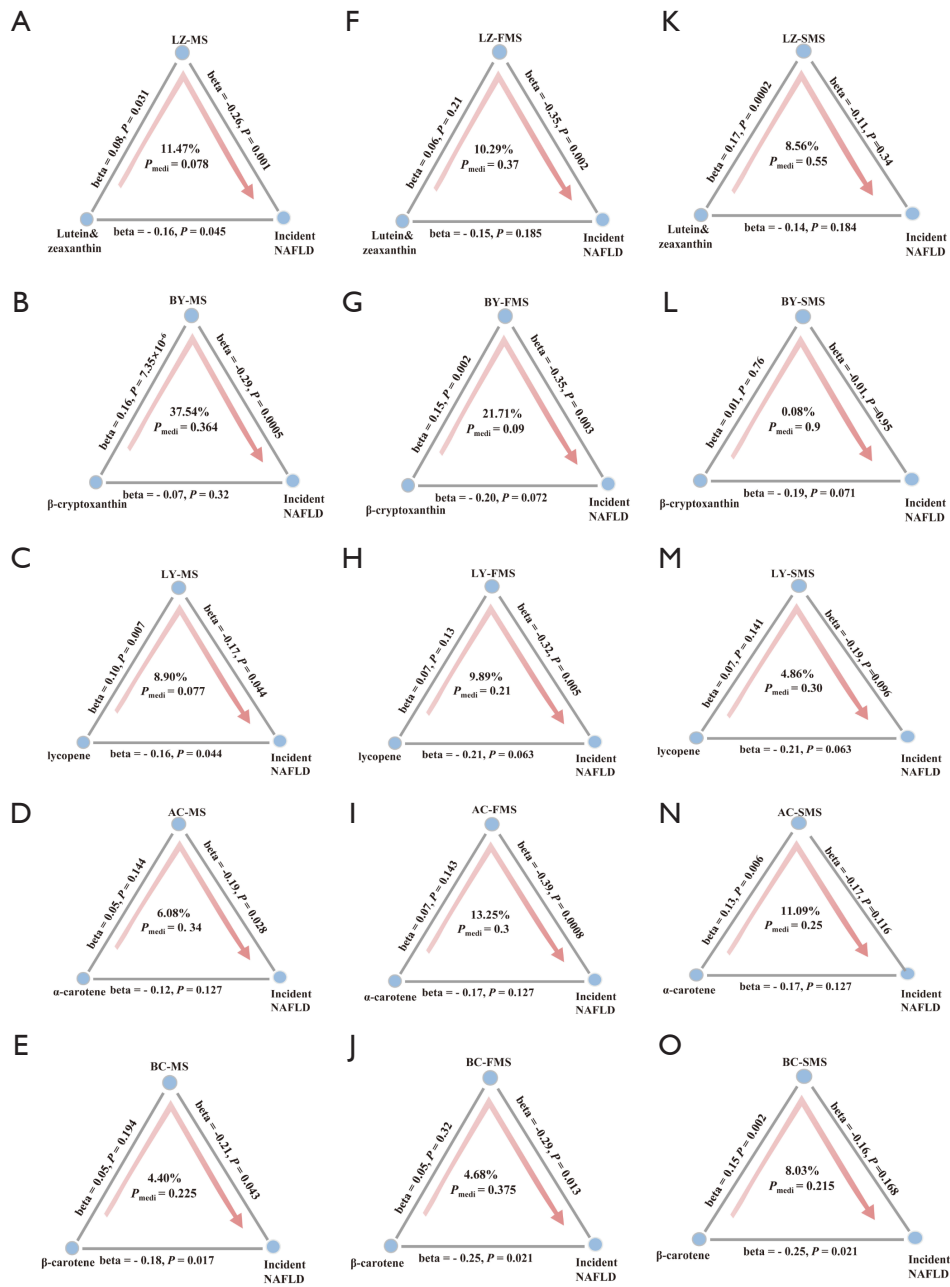
**Figure S1** Associations between serum carotenoids and odds/risks of persistent NAFLD, new-onset NAFLD, improved NAFLD, and MRI-identified NAFLD. Restricted cubic splines for the association between five serum carotenoids (lutein&zeaxanthin,  $\beta$ -cryptoxanthin, lycopene,  $\alpha$ -carotene, and  $\beta$ -carotene) and the odds/risks of persistent NAFLD (A-E), New-onset NAFLD (F-J), Improved NAFLD (K-O), and MRI-identified NAFLD (P-T), respectively. The blue lines and blue dots represent the estimated effects, while the light blue area and the error bars represent the 95% confidence intervals. Each column represents a NAFLD group type, and each row represents a kind of serum carotenoid. Covariates adjusted: age, sex, education, income, physical activity, multivitamin use, smoking, alcohol drinking, and dietary intake of energy, saturated fatty acids, and fiber. NAFLD, non-alcoholic fatty liver disease; MRI, magnetic resonance imaging.



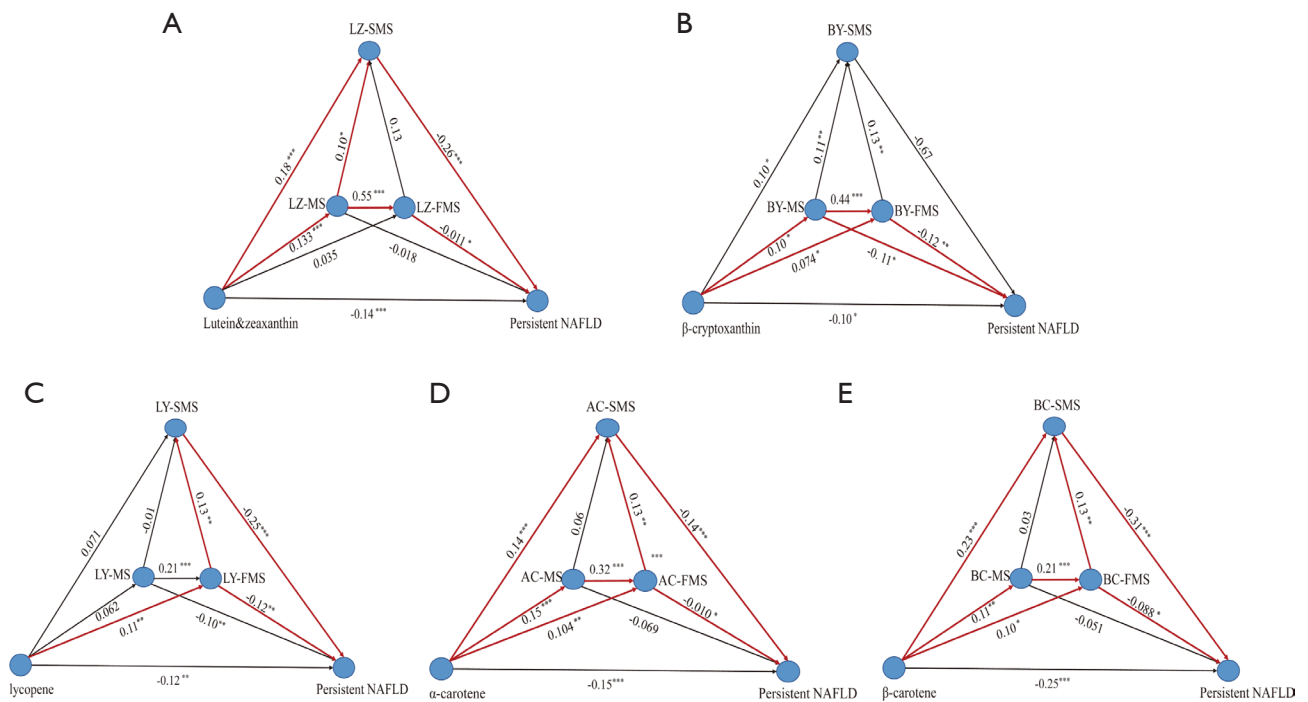
**Figure S2** Comparison of gut microbial  $\alpha$ -/ $\beta$ -diversity by categories of carotenoids. (A-C) Gut microbial  $\alpha$ -diversity (Observed OTUs, Shannon index, and Chao1 index) comparison in categories of five serum carotenoids (lutein&zeaxanthin,  $\beta$ -cryptoxanthin, lycopene,  $\alpha$ -carotene, and  $\beta$ -carotene). The *P-trend* value was calculated by multivariable linear regression. (D-H)  $\beta$ -diversity of gut microbiota by categories of five serum carotenoids (lutein&zeaxanthin,  $\beta$ -cryptoxanthin, lycopene,  $\alpha$ -carotene, and  $\beta$ -carotene), analyzed by PERMANOVA (999 permutations) based on Bray-Curtis distance. Covariates adjusted: age, sex, education, income, physical activity, multivitamin use, smoking, alcohol drinking, and dietary intake of energy, saturated fatty acids, and fiber.



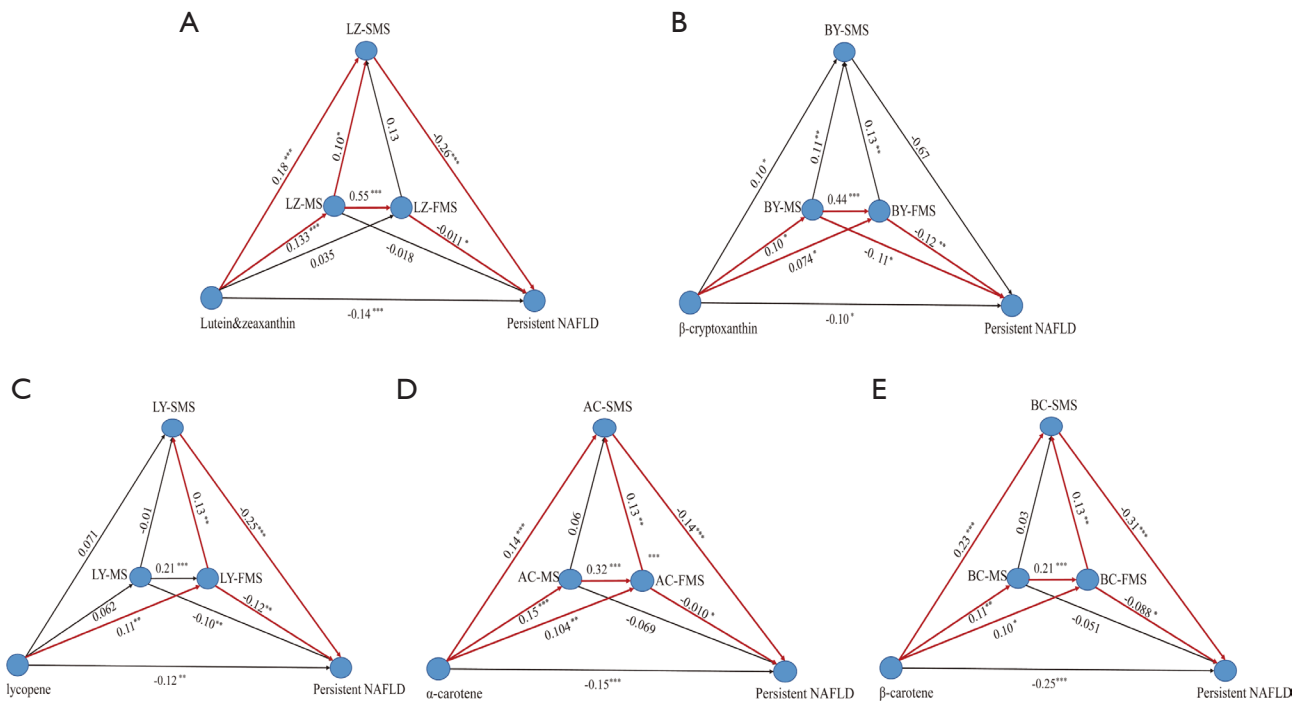
**Figure S3** The mediating effects of carotenoids-related microbial score (-MS) and metabolite score (-FMS and -SMS) on the associations between five serum carotenoids and persistent NAFLD. (A-E) The mediating effects of carotenoids-related microbial score (-MS) on the associations between five serum carotenoids (lutein/zeaxanthin,  $\beta$ -cryptoxanthin, lycopene,  $\alpha$ -carotene, and  $\beta$ -carotene) and persistent NAFLD. (F-J) The mediating effects of carotenoids-related fecal metabolite score (-FMS) on the associations between above five serum carotenoids and persistent NAFLD. (K-O) The mediating effects of carotenoids-related serum metabolite score (-SMS) on the associations between above five serum carotenoids and persistent NAFLD. The standardized beta coefficients on the lines were calculated by multivariable linear or logistic regression, adjusted for age, sex, education, income, physical activity, multivitamin use, smoking, alcohol drinking, energy, saturated fatty acids, and fiber intake. The number in the triangle is the proportion of mediation effect of the score in the association between carotenoids and NAFLD. NAFLD, non-alcoholic fatty liver disease; LZ, lutein&zeaxanthin; BY,  $\beta$ -cryptoxanthin; LY, lycopene; AC,  $\alpha$ -carotene; BC,  $\beta$ -carotene.



**Figure S4** The mediating effects of carotenoids-related microbial score(-MS) and metabolite score (-FMS and -SMS) on the associations between five serum carotenoids and Incident NAFLD. (A-E) The mediating effects of carotenoids-related microbial score (-MS) on the associations between five serum carotenoids (lutein&zeaxanthin, β-cryptoxanthin, lycopene, α-carotene, and β-carotene) and Incident NAFLD. (F-J) The mediating effects of carotenoids-related fecal metabolite score (-FMS) on the associations between above five serum carotenoids and Incident NAFLD. (K-O) The mediating effects of carotenoids-related serum metabolite score (-SMS) on the associations between above five serum carotenoids and incident NAFLD. The standardized beta coefficients on the lines were calculated by multivariable linear or logistic regression, adjusted for age, sex, education, income, physical activity, multivitamin use, smoking, alcohol drinking, energy, saturated fatty acids, and fiber intake. The number in the triangle is the proportion of mediation effect of the score in the association between carotenoids and NAFLD. NAFLD, non-alcoholic fatty liver disease; LZ, lutein&zeaxanthin; BY, β-cryptoxanthin; LY, lycopene; AC, α-carotene; BC, β-carotene.



**Figure S5** The path analysis of carotenoids-related microbial score (-MS) and metabolite score (-FMS and -SMS) on the associations between five serum carotenoids and Persistent NAFLD. a-e showed the pathways between the five carotenoids-related microbial score (-MS), carotenoids-related fecal (-FMS) and serum metabolite score (-SMS) on the association between serum carotenoids and Persistent NAFLD. The standardized beta coefficients were shown on the lines and the red arrow represents the direction of paths. NAFLD, non-alcoholic fatty liver disease; LZ, lutein&zeaxanthin; BY,  $\beta$ -cryptoxanthin; LY, lycopene; AC,  $\alpha$ -carotene; BC,  $\beta$ -carotene.



**Figure S6** The path analysis of carotenoids-related microbial score (-MS) and metabolite score (-FMS and -SMS) on the associations between five serum carotenoids and Incident NAFLD. (A-E) The pathways between the five carotenoids-related microbial score (-MS), carotenoids-related fecal (-FMS) and serum metabolite score (-SMS) on the association between serum carotenoids and Incident NAFLD. The standardized beta coefficients were shown on the lines and the red arrow represents the direction of paths. NAFLD, non-alcoholic fatty liver disease; LZ, lutein&zeaxanthin; BY,  $\beta$ -cryptoxanthin; LY, lycopene; AC,  $\alpha$ -carotene; BC,  $\beta$ -carotene.

**Table S1** The cross table of different methods to diagnosed NAFLD (N=828)

Methods to diagnosed NAFLD	Ultrasound		Total	Spearman correlation		Kappa	
	Non-NAFLD	NAFLD		Coefficients	P value	Coefficients	P value
MRI							
Non-NAFLD	310	176	486	0.46	3.63E-44	0.44	1.17E-39
NAFLD	60	282	342				
Total	370	458	828				

NAFLD, non-alcoholic fatty liver disease; MRI, magnetic resonance imaging.

**Table S2** Characteristics at baseline between participants included and excluded

Characteristics	Participants included at baseline (N=2, 921)	Participants excluded at baseline (N=1, 127)	P
Age (year)	58.55 (6.11)	58.62 (6.75)	0.730
Female, n (%)	1996 (68.3)	768 (68.2)	0.968
Waist circumference (cm)	83.22 (8.99)	83.42 (9.40)	0.543
BMI (kg/m <sup>2</sup> )	23.32 (3.05)	23.49 (3.32)	0.115
Education level, n (%)			0.002
Middle school	837 (28.7)	386 (34.3)	
High school or professional college	1348 (46.1)	485 (43.1)	
University	736 (25.2)	255 (22.6)	
Household income (Chinese Yuan /month/person), n (%)			<0.001
<1500	1377 (47.1)	621 (55.2)	
1,500-3,000	736 (25.2)	244 (21.7)	
≥3000	808 (27.7)	261 (23.2)	
Smoker <sup>†</sup> , n (%)	463 (15.9)	210 (18.7)	0.036
Alcohol drinker <sup>§</sup> , n (%)	198 (6.8)	82 (7.3)	0.619
Energy intake (kcal/day)	1779 (607)	1818 (593)	0.070
Fiber intake (g/day)	11.49 (7.18)	11.79 (6.30)	0.231
SFA intake (g/day)	14.05 (4.48)	14.09 (4.23)	0.801
Physical activity (MET·h/day)	26.18 (6.68)	24.57 (8.47)	<0.001
TG (mmol/L, N=2, 921/915)	1.58 (1.27)	1.65 (1.38)	0.158
TC (mmol/L, N=2, 921/1, 012)	5.53 (1.10)	5.39 (1.07)	<0.001
HDL-C (mmol/L, N=2, 921/915)	1.40 (0.34)	1.37 (0.34)	0.010
LDL-C (mmol/L, N=2, 921/1, 012)	3.59 (0.91)	3.55 (0.89)	0.315
FBG (mmol/L, N=2, 921/915)	4.89 (1.07)	4.99 (1.46)	0.022

Values are mean (SD) or n (%). <sup>†</sup>, smoker, 1 cigarette/d in the past year; <sup>§</sup>, alcohol drinker, 1 cup/week in the past year.



**Table S3** Association of serum carotenoids concentration with persistent NAFLD (N=1, 747)

Models	Quartile of serum carotenoids				Per quartile	P <sub>trend</sub>
	Quartile 1	Quartile 2	Quartile 3	Quartile 4		
<b>Lutein &amp; zeaxanthin</b>						
Case/all	309/439	257/435	217/412	194/461		
Model 1 <sup>a</sup>	Reference	0.61 (0.46, 0.81)	0.47 (0.36, 0.63)	0.31 (0.23, 0.4)	0.68 (0.63, 0.75)	1.00E-17
Model 2 <sup>b</sup>	Reference	0.61 (0.46, 0.81)	0.47 (0.36, 0.63)	0.31 (0.24, 0.41)	0.69 (0.63, 0.75)	9.72E-17
<b>B-cryptoxanthin</b>						
Case/all	283/434	256/442	225/430	213/441		
Model 1	Reference	0.74 (0.56, 0.97)	0.59 (0.45, 0.77)	0.5 (0.38, 0.66)	0.8 (0.73, 0.87)	2.15E-07
Model 2	Reference	0.77 (0.58, 1.01)	0.6 (0.46, 0.8)	0.52 (0.39, 0.68)	0.8 (0.73, 0.87)	7.62E-07
<b>Lycopene</b>						
Case/all	296/435	260/436	221/442	200/434		
Model 1	Reference	0.7 (0.53, 0.93)	0.47 (0.36, 0.62)	0.41 (0.31, 0.53)	0.73 (0.67, 0.8)	3.12E-12
Model 2	Reference	0.7 (0.53, 0.93)	0.48 (0.36, 0.63)	0.4 (0.3, 0.53)	0.73 (0.67, 0.8)	6.68E-12
<b>α-carotene</b>						
Case/all	314/444	273/434	218/434	172/435		
Model 1	Reference	0.7 (0.53, 0.94)	0.42 (0.32, 0.55)	0.27 (0.21, 0.36)	0.64 (0.59, 0.7)	1.63E-22
Model 2	Reference	0.69 (0.52, 0.92)	0.42 (0.31, 0.55)	0.27 (0.2, 0.36)	0.64 (0.58, 0.7)	8.79E-22
<b>β-carotene</b>						
Case/all	336/440	267/420	224/433	150/450		
Model 1	Reference	0.56 (0.42, 0.75)	0.34 (0.26, 0.46)	0.16 (0.12, 0.22)	0.55 (0.5, 0.6)	3.42E-37
Model 2	Reference	0.56 (0.42, 0.76)	0.35 (0.26, 0.46)	0.16 (0.12, 0.22)	0.55 (0.5, 0.6)	4.72E-36
<b>Total carotenoids</b>						
Case/all	325/434	268/431	217/435	167/447		
Model 1	Reference	0.56 (0.42, 0.75)	0.33 (0.25, 0.45)	0.2 (0.15, 0.27)	0.59 (0.54, 0.65)	2.50E-30
Model 2	Reference	0.55 (0.41, 0.74)	0.34 (0.25, 0.45)	0.2 (0.15, 0.27)	0.59 (0.54, 0.64)	2.97E-29

<sup>a</sup>Model 1: adjusted for age and sex. <sup>b</sup>Model 2: adjusted for variables in model 1 plus household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy intake, SFA intake, and fiber intake.

**Table S4** Association of serum carotenoids concentration with NAFLD incidence by Cox regression (N=1244)

Models	Quartile of serum carotenoids				Per quartile	P <sub>trend</sub>
	Quartile 1	Quartile 2	Quartile 3	Quartile 4		
<b>Lutein/zeaxanthin</b>						
Case/all	135/238	175/300	186/336	160/370		
Model 1 <sup>a</sup>	Reference	0.98 (0.78, 1.23)	0.93 (0.74, 1.16)	0.67 (0.53, 0.84)	0.88 (0.82, 0.94)	0.0003
Model 2 <sup>b</sup>	Reference	0.97 (0.77, 1.22)	0.91 (0.73, 1.14)	0.65 (0.52, 0.83)	0.87 (0.81, 0.94)	0.0002
<b>B-cryptoxanthin</b>						
Case/all	160/272	139/280	180/337	177/355		
Model 1	Reference	0.72 (0.58, 0.91)	0.86 (0.7, 1.07)	0.68 (0.55, 0.85)	0.91 (0.85, 0.97)	0.006
Model 2	Reference	0.73 (0.58, 0.92)	0.87 (0.7, 1.08)	0.68 (0.55, 0.85)	0.91 (0.85, 0.97)	0.008
<b>Lycopene</b>						
Case/all	142/247	161/290	175/353	178/354		
Model 1	Reference	0.92 (0.73, 1.15)	0.79 (0.63, 0.98)	0.82 (0.66, 1.03)	0.93 (0.87, 1)	0.046
Model 2	Reference	0.93 (0.74, 1.17)	0.8 (0.64, 1)	0.82 (0.65, 1.03)	0.93 (0.86, 1)	0.050
<b>α-carotene</b>						
Case/all	133/232	165/290	173/331	185/391		
Model 1	Reference	0.91 (0.72, 1.14)	0.84 (0.67, 1.06)	0.74 (0.59, 0.93)	0.91 (0.84, 0.97)	0.006
Model 2	Reference	0.9 (0.71, 1.13)	0.83 (0.66, 1.05)	0.74 (0.59, 0.94)	0.91 (0.84, 0.98)	0.009
<b>β-carotene</b>						
Case/all	125/201	158/270	180/342	193/431		
Model 1	Reference	0.85 (0.67, 1.08)	0.77 (0.61, 0.97)	0.59 (0.47, 0.74)	0.84 (0.78, 0.9)	1.48e-06
Model 2	Reference	0.84 (0.66, 1.06)	0.75 (0.59, 0.95)	0.59 (0.47, 0.74)	0.84 (0.78, 0.9)	2.08e-06
<b>Total carotenoids</b>						
Case/all	127/209	167/290	180/341	182/404		
Model 1	Reference	1.02 (0.8, 1.29)	0.81 (0.64, 1.02)	0.64 (0.51, 0.81)	0.85 (0.79, 0.91)	5.49e-06
Model 2	Reference	0.98 (0.78, 1.25)	0.79 (0.62, 0.99)	0.63 (0.5, 0.8)	0.85 (0.79, 0.91)	5.26e-06

<sup>a</sup>Model 1: adjusted for age and sex. <sup>b</sup>Model 2: adjusted for variables in model 1 plus household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy intake, SFA intake, and fiber intake.

**Table S5** Association of serum carotenoids concentration with Improved NAFLD by Cox regression (N=1278)

Models	Quartile of serum carotenoids				Per quartile	P <sub>trend</sub>
	Quartile 1	Quartile 2	Quartile 3	Quartile 4		
<b>Lutein/zeaxanthin</b>						
Improved/all	84/393	76/333	67/284	74/268		
Model 1 <sup>a</sup>	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.09 (0.98, 1.2)	0.103
Model 2 <sup>b</sup>	Reference	1 (0.73, 1.37)	1.11 (0.8, 1.55)	1.24 (0.9, 1.71)	1.08 (0.97, 1.2)	0.145
<b>B-cryptoxanthin</b>						
Improved/all	74/357	88/344	69/294	70/283		
Model 1	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.05 (0.95, 1.16)	0.373
Model 2	Reference	1.45 (1.06, 1.99)	1.21 (0.86, 1.69)	1.29 (0.92, 1.8)	1.06 (0.96, 1.18)	0.269
<b>Lycopene</b>						
Improved/all	83/379	72/332	72/293	74/274		
Model 1	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.07 (0.97, 1.18)	0.184
Model 2	Reference	1.04 (0.75, 1.43)	1.14 (0.82, 1.57)	1.28 (0.93, 1.78)	1.09 (0.98, 1.21)	0.117
<b>α-carotene</b>						
Improved/all	90/404	65/338	83/301	63/235		
Model 1	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.05 (0.95, 1.17)	0.317
Model 2	Reference	0.76 (0.55, 1.05)	1.09 (0.8, 1.48)	1.09 (0.78, 1.52)	1.06 (0.95, 1.18)	0.314
<b>β-carotene</b>						
Improved/all	96/432	80/347	60/284	65/215		
Model 1	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.11 (1, 1.23)	0.046
Model 2	Reference	1.14 (0.84, 1.55)	1.08 (0.78, 1.5)	1.4 (1.02, 1.94)	1.1 (0.99, 1.22)	0.074
<b>Total carotenoids</b>						
Improved/all	82/407	89/357	59/276	71/238		
Model 1	Reference	1.04 (0.76, 1.42)	1.12 (0.81, 1.55)	1.3 (0.95, 1.78)	1.11 (1.01, 1.23)	0.038
Model 2	Reference	1.4 (1.03, 1.91)	1.08 (0.76, 1.52)	1.53 (1.1, 2.12)	1.11 (1, 1.23)	0.054

<sup>a</sup>Model 1: adjusted for age and sex. <sup>b</sup>Model 2: adjusted for variables in model 1 plus household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy intake, SFA intake, and fiber intake.

**Table S6** Association of serum carotenoids concentration with MRI-identified NAFLD (N=828)

Models	Quartile of serum carotenoids				Per quartile	P <sub>trend</sub>
	Quartile 1	Quartile 2	Quartile 3	Quartile 4		
<b>Lutein/zeaxanthin</b>						
Case/all	98/200	80/210	87/208	77/210		
Model 1 <sup>a</sup>	Reference	0.64 (0.43, 0.95)	0.74 (0.5, 1.1)	0.6 (0.4, 0.89)	0.87 (0.77, 0.99)	0.030
Model 2 <sup>b</sup>	Reference	0.64 (0.43, 0.95)	0.69 (0.46, 1.03)	0.59 (0.39, 0.88)	0.86 (0.76, 0.98)	0.019
<b>B-cryptoxanthin</b>						
Case/all	84/200	87/205	87/218	84/205		
Model 1	Reference	1.02 (0.69, 1.51)	0.91 (0.62, 1.35)	0.96 (0.64, 1.42)	0.98 (0.86, 1.11)	0.694
Model 2	Reference	0.98 (0.65, 1.47)	0.86 (0.57, 1.28)	0.94 (0.63, 1.41)	0.97 (0.85, 1.1)	0.632
<b>Lycopene</b>						
Case/all	87/170	86/196	78/226	91/236		
Model 1	Reference	0.75 (0.49, 1.13)	0.5 (0.33, 0.75)	0.59 (0.4, 0.89)	0.83 (0.73, 0.94)	0.003
Model 2	Reference	0.72 (0.47, 1.11)	0.49 (0.32, 0.74)	0.55 (0.36, 0.83)	0.81 (0.71, 0.92)	0.001
<b>α-carotene</b>						
Case/all	81/180	97/208	81/208	83/232		
Model 1	Reference	1.07 (0.72, 1.6)	0.78 (0.52, 1.16)	0.68 (0.45, 1.01)	0.86 (0.76, 0.97)	0.016
Model 2	Reference	1.05 (0.7, 1.58)	0.76 (0.5, 1.15)	0.65 (0.43, 0.98)	0.85 (0.75, 0.97)	0.012
<b>β-carotene</b>						
Case/all	95/190	91/209	79/205	77/224		
Model 1	Reference	0.77 (0.52, 1.15)	0.62 (0.42, 0.93)	0.52 (0.35, 0.77)	0.8 (0.71, 0.91)	0.0007
Model 2	Reference	0.78 (0.52, 1.17)	0.6 (0.4, 0.91)	0.5 (0.33, 0.76)	0.79 (0.7, 0.9)	0.0004
<b>Total carotenoids</b>						
Case/all	90/185	92/223	72/191	86/229		
Model 1	Reference	0.71 (0.48, 1.05)	0.61 (0.4, 0.92)	0.6 (0.41, 0.9)	0.85 (0.75, 0.96)	0.011
Model 2	Reference	0.7 (0.47, 1.05)	0.6 (0.39, 0.91)	0.58 (0.39, 0.87)	0.84 (0.74, 0.96)	0.008

<sup>a</sup>Model 1: adjusted for age and sex. <sup>b</sup>Model 2: adjusted for variables in model 1 plus household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy intake, SFA intake, and fiber intake.

**Table S7** Association of serum carotenoids with microbiota  $\alpha$ -diversity of 16S rRNA sequencing (N=1, 661)<sup>a</sup>

Carotenoids	Observed OTUs				Shannon index				Chao1 index			
	B	95% CI		P	B	95% CI		P	B	95% CI		P
		Lower	Upper			Lower	Upper			Lower	Upper	
Lutein/zeaxanthin												
Q1 (N=400)	Ref				Ref				Ref			
Q2 (N=410)	0.23	0.10	0.37	0.001	0.16	0.02	0.30	0.025	0.27	0.14	0.40	5.28E-05
Q3 (N=427)	0.15	0.02	0.29	0.027	0.08	-0.06	0.22	0.267	0.17	0.04	0.30	0.009
Q4 (N=424)	0.30	0.16	0.44	2.33E-05	0.20	0.06	0.34	0.005	0.32	0.19	0.46	1.26E-06
Per quartile	0.08	0.04	0.12	3.03E-04	0.05	0.01	0.10	0.021	0.09	0.04	0.13	4.58E-05
B-cryptoxanthin												
Q1 (N=400)	Ref				Ref				Ref			
Q2 (N=410)	0.08	-0.06	0.22	0.259	0.07	-0.07	0.20	0.345	0.09	-0.04	0.22	0.159
Q3 (N=427)	0.10	-0.03	0.24	0.137	0.13	-0.003	0.27	0.055	0.09	-0.04	0.22	0.185
Q4 (N=424)	0.18	0.04	0.31	0.010	0.14	0.01	0.28	0.040	0.19	0.06	0.32	0.004
Per quartile	0.06	0.01	0.10	0.011	0.05	0.01	0.09	0.025	0.06	0.02	0.10	0.007
Lycopene												
Q1 (N=396)	Ref				Ref				Ref			
Q2 (N=418)	0.09	-0.05	0.22	0.211	0.14	0.002	0.28	0.046	0.06	-0.07	0.20	0.328
Q3 (N=412)	0.12	-0.02	0.26	0.103	0.14	0.001	0.28	0.049	0.07	-0.06	0.21	0.279
Q4 (N=435)	0.11	-0.03	0.25	0.111	0.16	0.02	0.30	0.028	0.10	-0.03	0.23	0.137
Per quartile	0.04	-0.01	0.08	0.111	0.05	0.002	0.09	0.043	0.03	-0.01	0.07	0.153
$\alpha$ -carotene												
Q1 (N=421)	Ref				Ref				Ref			
Q2 (N=418)	0.17	0.03	0.30	0.015	0.18	0.04	0.32	0.009	0.14	0.01	0.27	0.032
Q3 (N=403)	0.20	0.06	0.34	0.005	0.18	0.04	0.32	0.012	0.17	0.04	0.30	0.010
Q4 (N=419)	0.18	0.04	0.32	0.012	0.20	0.06	0.34	0.005	0.14	0.01	0.27	0.038
Per quartile	0.06	0.01	0.10	0.013	0.06	0.02	0.10	0.008	0.04	0.00	0.09	0.037
$\beta$ -carotene												
Q1 (N=395)	Ref				Ref				Ref			
Q2 (N=410)	0.04	-0.10	0.18	0.545	0.03	-0.10	0.17	0.631	0.05	-0.08	0.18	0.479
Q3 (N=430)	0.13	0.00	0.27	0.057	0.11	-0.03	0.24	0.130	0.12	-0.01	0.25	0.065
Q4 (N=426)	0.20	0.06	0.34	0.006	0.20	0.06	0.34	0.005	0.21	0.08	0.34	0.002
Per quartile	0.07	0.02	0.11	0.003	0.07	0.02	0.11	0.003	0.07	0.03	0.11	0.001
Total carotenoids												
Q1 (N=394)	Ref				Ref				Ref			
Q2 (N=415)	0.04	-0.10	0.18	0.568	0.08	-0.06	0.22	0.255	0.04	-0.09	0.17	0.543
Q3 (N=420)	0.14	0.00	0.28	0.049	0.14	0.01	0.28	0.042	0.15	0.02	0.28	0.026
Q4 (N=432)	0.22	0.08	0.36	0.003	0.22	0.08	0.36	0.003	0.23	0.09	0.36	0.001
Per quartile	0.08	0.03	0.12	0.001	0.07	0.03	0.12	0.002	0.08	0.04	0.12	2.28E-04

<sup>a</sup>,  $\beta$  (95% CIs) were calculated for Q2–Q4 of the carotenoids using Q1 as the reference group. Multivariable linear regression was conducted with the standardized  $\alpha$ -diversity metrics as dependent variables and baseline quartiles of serum carotenoids as independent variables, adjusted for covariates in age, sex, education, income, physical activity, multivitamin use, smoking, alcohol drinking, energy, SFA and fiber intake, Bristol score, and sequencing parameters.

**Table S8** Results from PERMANOVA analysis of the associations of carotenoids (Q1–Q4) with microbiota  $\beta$ -diversity (N=1, 661)<sup>a</sup>

Carotenoids	DF	F	R <sup>2</sup>	P
Lutein + zeaxanthin	1	1.766312	0.003129	0.004
$\beta$ -cryptoxanthin	1	1.897346	0.00336	0.002
Lycopene	1	1.247866	0.002213	0.134
$\alpha$ -carotene	1	1.550915	0.002748	0.02
$\beta$ -carotene	1	1.352107	0.002397	0.059
Total carotenoids	1	1.934969	0.003427	0.001

<sup>a</sup>, the comparison of gut composition between carotenoids Q1–Q4 were assessed with PERMANOVA (999 permutations) based on the Bray-Curtis distance calculated at the genus level.

**Table S9** ANCOVA of long-term NAFLD groups with microbiota  $\alpha$ -diversity of 16S rRNA sequencing (N=1, 628)

$\alpha$ -diversity indices	NAFLD-free group (N=498)		Improved NAFLD group (N=174)		New-onset NAFLD group (N=323)		NAFLD-persistent group (N=633)		P <sub>trend</sub>
	Mean difference	SE	Mean difference	SE	Mean difference	SE	Mean difference	SE	
Observed OTUs	90.09	1.04	84.64	1.75	86.65	1.29	82.71***	0.92	1.84E-07
Shannon index	5.14	0.03	5.00	0.05	5.04	0.04	4.95***	0.03	3.37E-06
Chao1 index	123.66	1.51	116.72	2.55	118.28	1.88	112.82***	1.33	4.76E-09

Model adjusted for age, sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy intake, SFA intake, fiber intake and Bristol score, sequencing run, and sequencing depth. Benjamini-Hochberg method was used for poly nominal contrasts. Compared with long-term healthy group: \*\*\*, P<0.001; no significance of  $\alpha$ -diversity indices was observed between three NAFLD groups.

**Table S10** Results from PERMANOVA analysis of the associations of NAFLD groups with microbiota  $\beta$ -diversity (N=1, 628)<sup>a</sup>

NAFLD groups	DF	F	R <sup>2</sup>	P	FDR
Improved NAFLD group vs. NAFLD-free group	1	1.3996	0.0020	0.136	0.1632
New-onset NAFLD group vs. NAFLD-free group	1	2.1303	0.0025	0.019	0.057
NAFLD-persistent group vs. NAFLD-free group	1	5.2894	0.0046	0.001	0.006
Improved NAFLD group vs. new-onset NAFLD group	1	1.5178	0.0030	0.086	0.129
Improved NAFLD group vs. NAFLD-persistent group	1	1.1668	0.0014	0.287	0.287
New-onset NAFLD group vs. NAFLD-persistent group	1	1.6121	0.0017	0.079	0.158

<sup>a</sup>, the comparison of gut composition between different group of NAFLD status were assessed with PERMANOVA (999 permutations) based on the Bray-Curtis distance calculated at the genus level. P values were adjusted for multiple testing using Benjamini-Hochberg method.

**Table S11** The overlapping differential gut microbiota genera of serum carotenoids (Q4 vs. Q1) or NAFLD (NAFLD-free vs. persistent/incident NAFLD)<sup>a</sup>

Carotenoids	Phylum	Genus	Coefficient	P value	FDR
lutein/zeaxanthin	Firmicutes	Holdemania	0.0056	9.7E-06	0.002
	Firmicutes	Ruminococcaceae_uncultured	0.0045	9.8E-05	0.012
	Fusobacteria	Fusobacterium	-0.0189	0.0004	0.023
	Firmicutes	Ruminococcaceae UCG-005	0.0097	0.0005	0.027
	Firmicutes	Lachnoclostridium	-0.0195	0.0005	0.028
	Firmicutes	Ruminococcaceae UCG-014	0.0131	0.0012	0.053
	Firmicutes	Megamonas	-0.0330	0.0019	0.078
	Firmicutes	Ruminococcaceae UCG-002	0.0174	0.0023	0.089
	Firmicutes	Christensenellaceae R-7 group	0.0130	0.0089	0.184
β-cryptoxanthin	Fusobacteria	Fusobacterium	-0.0319	0.0004	0.031
	Firmicutes	Intestinibacter	0.0046	0.0011	0.058
	Firmicutes	Lachnoclostridium	-0.0166	0.0028	0.108
	Firmicutes	Megasphaera	-0.0222	0.0030	0.112
	Firmicutes	Ruminococcaceae UCG-002	0.0159	0.0038	0.131
	Firmicutes	Allisonella	-0.0041	0.0063	0.180
	Firmicutes	Faecalibacterium	0.0201	0.0077	0.199
	Firmicutes	Ruminococcaceae UCG-005	0.0069	0.0090	0.215
	Firmicutes	Ruminococcaceae UCG-014	0.0195	0.0089	0.215
Lycopene	Firmicutes	Christensenellaceae R-7 group	0.0132	0.0139	0.241
	Firmicutes	Ruminiclostridium 6	0.0098	0.0021	0.110
α-carotene	Firmicutes	Megasphaera	-0.0228	0.0032	0.147
	Firmicutes	Ruminococcaceae UCG-003	0.0075	7.4E-05	0.007
β-carotene	Firmicutes	Ruminiclostridium 6	0.0101	0.0015	0.068
	Firmicutes	Ruminococcaceae_uncultured	0.0033	0.0047	0.161
	Firmicutes	Megasphaera	-0.0193	0.0109	0.239
Total carotenoids	Firmicutes	Ruminiclostridium 6	0.0096	0.0022	0.101
	Firmicutes	Ruminococcaceae_uncultured	-0.0003	0.0015	0.077
	Fusobacteria	Fusobacterium	-0.0291	0.0018	0.088
	Firmicutes	Ruminiclostridium 6	0.0101	0.0022	0.107
NAFLD	Firmicutes	Megasphaera	-0.0196	0.0081	0.225
	Bacteroidetes	Alistipes	0.0165	0.0097	0.244
	Firmicutes	Ruminococcaceae UCG-002	-0.021	1.91E-06	2.06E-04
	Firmicutes	Ruminococcaceae UCG-014	-0.020	1.76E-04	0.007
	Firmicutes	Faecalibacterium	-0.018	0.003	0.056

**Table S11** (continued)

**Table S11** (continued)

Carotenoids	Phylum	Genus	Coefficient	P value	FDR
	Bacteroidetes	Alistipes	-0.016	0.001	0.029
	Firmicutes	Christensenellaceae R-7 group	-0.011	0.004	0.079
	Firmicutes	Ruminiclostridium 6	-0.008	7.11E-04	0.019
	Firmicutes	Ruminococcaceae UCG-005	-0.007	2.99E-04	0.011
	Firmicutes	Ruminococcaceae UCG-003	-0.005	6.51E-04	0.019
	Firmicutes	Ruminococcaceae_uncultured	-0.003	3.02E-04	0.011
	Firmicutes	Intestinibacter	-0.003	0.005	0.091
	Firmicutes	Holdemania	-0.003	0.004	0.068
	Firmicutes	Megasphaera	0.018	0.002	0.036
	Firmicutes	Megamonas	0.025	0.003	0.052
	Fusobacteria	Fusobacterium	0.022	0.003	0.052
	Firmicutes	Lachnoclostridium	0.012	0.004	0.079
	Firmicutes	Allisonella	0.003	0.024	0.243

<sup>a</sup>, the  $\beta$  coefficients were calculated by Multivariate Analysis by Linear Models (MaAsLin) (FDR <0.25 was considered statistically significant) after transformed the relative abundance of gut microbes by arcsine square root, and the false discovery rate (FDR) was controlled using Benjamini-Hochberg method.



**Table S12** The association of overlapping microbiome genera with NAFLD groups

Phylum	Family	Genus	NAFLD persistent and incident			New-onset NAFLD		MRI-identified NAFLD	
			Case/all=956/1, 454			Case/all=323/821		Case/all=265/632	
			OR (95% CI)	P	FDR	OR (95% CI)	P	OR (95% CI)	P
Bacteroidetes	Rikenellaceae	Alistipes	0.84 (0.76, 0.94)	0.002	0.0046	0.89 (0.77, 1.04)	0.140	0.79 (0.67, 0.93)	0.005
Firmicutes	Christensenellaceae	Christensenellaceae R-7 group	0.86 (0.77, 0.96)	0.005	0.0062	0.84 (0.72, 0.97)	0.021	0.79 (0.67, 0.94)	0.007
Firmicutes	Peptostreptococcaceae	Intestinibacter	0.86 (0.77, 0.96)	0.006	0.0071	0.94 (0.82, 1.08)	0.398	1.04 (0.88, 1.23)	0.655
Firmicutes	Ruminococcaceae	Faecalibacterium	0.85 (0.76, 0.95)	0.004	0.0053	0.87 (0.75, 1.02)	0.085	0.87 (0.73, 1.03)	0.109
Firmicutes	Ruminococcaceae	Ruminiclostridium 6	0.83 (0.75, 0.93)	8.04E-04	0.0026	0.91 (0.8, 1.04)	0.164	0.71 (0.58, 0.87)	0.001
Firmicutes	Ruminococcaceae	Ruminococcaceae UCG-002	0.77 (0.69, 0.86)	2.81E-06	4.49E-05	0.83 (0.72, 0.96)	0.012	0.76 (0.64, 0.9)	0.001
Firmicutes	Ruminococcaceae	Ruminococcaceae UCG-003	0.84 (0.75, 0.93)	0.001	0.0034	0.89 (0.77, 1.03)	0.108	0.74 (0.62, 0.87)	4.72E-04
Firmicutes	Ruminococcaceae	Ruminococcaceae UCG-005	0.82 (0.74, 0.92)	4.04E-04	0.0022	0.86 (0.74, 0.99)	0.040	0.75 (0.64, 0.88)	0.001
Firmicutes	Ruminococcaceae	Ruminococcaceae UCG-014	0.81 (0.73, 0.91)	2.01E-04	0.0016	0.86 (0.75, 1)	0.044	0.8 (0.67, 0.95)	0.011
Firmicutes	Ruminococcaceae	Ruminococcaceae_ uncultured	0.82 (0.74, 0.92)	4.06E-04	0.0016	0.83 (0.72, 0.97)	0.019	0.8 (0.67, 0.96)	0.014
Firmicutes	Erysipelotrichaceae	Holdemania	0.85 (0.76, 0.95)	0.004	0.0054	0.94 (0.82, 1.08)	0.391	0.89 (0.75, 1.07)	0.216
Firmicutes	Lachnospiraceae	Lachnoclostridium	1.16 (1.04, 1.31)	0.011	0.0116	1.1 (0.96, 1.27)	0.172	1.27 (1.06, 1.51)	0.008
Firmicutes	Veillonellaceae	Allisonella	1.15 (1.02, 1.29)	0.023	0.0225	1.07 (0.92, 1.24)	0.383	1.05 (0.9, 1.22)	0.541
Firmicutes	Veillonellaceae	Megamonas	1.2 (1.06, 1.35)	0.003	0.0051	1.08 (0.92, 1.26)	0.372	1.26 (1.06, 1.49)	0.009
Firmicutes	Veillonellaceae	Megasphaera	1.22 (1.08, 1.38)	0.002	0.0044	1.09 (0.93, 1.28)	0.298	1.18 (1, 1.39)	0.049
Fusobacteria	Fusobacteriaceae	Fusobacterium	1.2 (1.06, 1.35)	0.003	0.005	1.12 (0.95, 1.31)	0.168	1.17 (0.99, 1.4)	0.070

Multivariable logistic regression to determine the associations between genera biomarkers and NAFLD groups, adjusted for age and sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, total energy intake, SFA intake, fiber intake, Bristol score, and sequencing parameters.

**Table S13** The association of carotenoids with deferential fecal metabolites (N=893)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Lutein/zeaxanthin						
Q2/Q1	Bile acids	HMDB0000637	-0.201	0.092	0.0298	1.10
Q3/Q1		Glycochenodeoxycholic acid	-0.218	0.093	0.0199	
Q4/Q1			-0.246	0.095	0.0096	
Per quartile			-0.079	0.031	0.0094	
Q2/Q1	Bile acids	HMDB0000467	-0.107	0.092	0.2447	2.65
Q3/Q1		7ketoLCA	-0.113	0.093	0.2243	
Q4/Q1			-0.190	0.095	0.0450	
Per quartile			-0.053	0.030	0.0807	
Q2/Q1	Bile acids	HMDB0000506	-0.144	0.093	0.1225	2.17
Q3/Q1		a-muricholic acid and b-muricholic acid	-0.134	0.094	0.1552	
Q4/Q1			-0.233	0.096	0.0148	
Per quartile			-0.069	0.031	0.0251	
Q2/Q1	Bile acids	HMDB0000518	-0.166	0.093	0.0739	2.45
Q3/Q1		Chenodeoxycholic acid	-0.173	0.094	0.0658	
Q4/Q1			-0.219	0.095	0.0218	
Per quartile			-0.060	0.031	0.0513	
Q2/Q1	Bile acids	HMDB0000664	-0.122	0.092	0.1866	2.73
Q3/Q1		b-Hyodeoxycholic acid	-0.110	0.094	0.2399	
Q4/Q1			-0.239	0.095	0.0118	
Per quartile			-0.064	0.031	0.0363	
Q2/Q1	Bile acids	HMDB0000708	-0.159	0.091	0.0814	1.29
Q3/Q1		Glycoursodeoxycholic acid	-0.136	0.092	0.1405	
Q4/Q1			-0.252	0.093	0.0071	
Per quartile			-0.077	0.030	0.0101	
Q2/Q1	Bile acids	HMDB0000811	-0.120	0.091	0.1902	2.59
Q3/Q1		Muro-cholic acid	-0.103	0.093	0.2676	
Q4/Q1			-0.206	0.094	0.0284	
Per quartile			-0.056	0.030	0.0657	
Q2/Q1	Fatty acids	HMDB0000555	0.306	0.093	0.0010	1.36
Q3/Q1		3-Methyladipic acid	0.091	0.094	0.3306	
Q4/Q1			0.307	0.095	0.0013	
Per quartile			0.063	0.031	0.0397	

**Table S13** (continued)

Table S13 (continued)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Q2/Q1	Carnitines	HMDB0005065	-0.204	0.093	0.0287	2.11
Q3/Q1		Oleylcarnitine C18:1	-0.192	0.094	0.0419	
Q4/Q1			-0.192	0.096	0.0453	
Per quartile			-0.039	0.031	0.2017	
Q2/Q1	Phenylpropanoic acids	HMDB0000764	0.102	0.092	0.2675	1.92
Q3/Q1		Hydrocinnamic acid	0.094	0.093	0.3154	
Q4/Q1			0.206	0.095	0.0297	
Per quartile			0.067	0.030	0.0286	
Q2/Q1	Phenylpropanoic acids	HMDB0011743	0.090	0.093	0.3341	2.03
Q3/Q1		2-Phenylpropionate	0.079	0.094	0.3964	
Q4/Q1			0.206	0.095	0.0307	
Per quartile			0.067	0.031	0.0294	
$\beta$ -cryptoxanthin						
Q2/Q1	Fatty acids	HMDB0000448	-0.047	0.093	0.6153	1.97
Q3/Q1		Adipic acid	0.048	0.093	0.6055	
Q4/Q1			0.257	0.095	0.0070	
Per quartile			0.080	0.030	0.0084	
Q2/Q1	Bile acids	HMDB0000502	0.078	0.094	0.4061	1.96
Q3/Q1		3-DHCA	0.006	0.093	0.9484	
Q4/Q1			-0.228	0.095	0.0170	
Per quartile			-0.070	0.030	0.0224	
Q2/Q1	Phenylpropanoic acids	HMDB0000764	0.133	0.093	0.1521	2.22
Q3/Q1		Hydrocinnamic acid	0.141	0.093	0.1285	
Q4/Q1			0.276	0.095	0.0036	
Per quartile			0.086	0.030	0.0043	
Q2/Q1	Indoles	HMDB0002302	0.010	0.091	0.9146	1.77
Q3/Q1		3-Indolepropionic acid	0.030	0.091	0.7458	
Q4/Q1			0.318	0.093	0.0007	
Per quartile			0.094	0.030	0.0017	
Q2/Q1	Phenylpropanoic acids	HMDB0011743	0.142	0.093	0.1288	2.16
Q3/Q1		2-Phenylpropionate	0.145	0.093	0.1196	
Q4/Q1			0.272	0.095	0.0043	
Per quartile			0.085	0.030	0.0051	

Table S13 (continued)

**Table S13** (continued)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Lycopene						
Q2/Q1	Carnitines	HMDB0000222	-0.097	0.096	0.3146	1.95
Q3/Q1		Palmitoylcarnitine	-0.197	0.097	0.0429	
Q4/Q1			-0.199	0.097	0.0398	
Per quartile				-0.057	0.031	
Q2/Q1	Fatty acids	HMDB0000448	0.161	0.097	0.0951	1.85
Q3/Q1		Adipic acid	0.102	0.098	0.2958	
Q4/Q1			0.218	0.097	0.0252	
Per quartile				0.054	0.031	
Q2/Q1	Bile acids	HMDB0000502	-0.156	0.097	0.1068	2.31
Q3/Q1		3-DHCA	-0.202	0.098	0.0389	
Q4/Q1			-0.275	0.097	0.0048	
Per quartile				-0.085	0.031	
Q2/Q1	Fatty acids	HMDB0000784	0.102	0.097	0.2927	1.67
Q3/Q1		Azelaic acid	0.139	0.098	0.1534	
Q4/Q1			0.209	0.097	0.0319	
Per quartile				0.059	0.031	
Q2/Q1	Fatty acids	HMDB0000792	0.123	0.095	0.1985	1.92
Q3/Q1		Sebacic acid	0.068	0.096	0.4813	
Q4/Q1			0.233	0.096	0.0154	
Per quartile				0.057	0.031	
Q2/Q1	Fatty acids	HMDB0000893	0.076	0.097	0.4297	1.55
Q3/Q1		Suberic acid	0.137	0.098	0.1610	
Q4/Q1			0.217	0.097	0.0258	
Per quartile				0.064	0.031	
Q2/Q1	Carnitines	HMDB0005065	-0.075	0.097	0.4414	1.63
Q3/Q1		Oleylcarnitine C18:1	-0.253	0.098	0.0099	
Q4/Q1			-0.226	0.097	0.0205	
Per quartile				-0.071	0.031	
$\alpha$ -carotene						
Q2/Q1	Carnitines	HMDB0000222	-0.124	0.093	0.1831	1.94
Q3/Q1		Palmitoylcarnitine	-0.160	0.097	0.1003	
Q4/Q1			-0.195	0.097	0.0440	

**Table S13** (continued)

Table S13 (continued)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Per quartile			-0.050	0.031	0.1058	
Q2/Q1	Fatty acids	HMDB0000555	0.117	0.094	0.2133	1.51
Q3/Q1		3-Methyladipic acid	0.305	0.098	0.0018	
Q4/Q1			0.234	0.097	0.0168	
Per quartile			0.082	0.031	0.0090	
Q2/Q1	Phenylpropanoic acids	HMDB0000764	0.296	0.092	0.0014	2.35
Q3/Q1		Hydrocinnamic acid	0.322	0.096	0.0009	
Q4/Q1			0.322	0.096	0.0008	
Per quartile			0.103	0.031	0.0008	
Q2/Q1	Indoles	HMDB0002302	0.174	0.091	0.0576	2.09
Q3/Q1		3-Indolepropionic acid	0.248	0.095	0.0094	
Q4/Q1			0.316	0.095	0.0009	
Per quartile			0.103	0.030	0.0007	
Q2/Q1	Carnitines	HMDB0005065	-0.027	0.094	0.7730	1.64
Q3/Q1		Oleylcarnitine C18:1	-0.127	0.098	0.1950	
Q4/Q1			-0.219	0.098	0.0253	
Per quartile			-0.061	0.031	0.0481	
Q2/Q1	Phenylpropanoic acids	HMDB0011743	0.294	0.093	0.0016	2.35
Q3/Q1		2-Phenylpropionate	0.317	0.097	0.0011	
Q4/Q1			0.314	0.096	0.0012	
Per quartile			0.101	0.031	0.0012	
$\beta$ -carotene						
Q2/Q1	Fatty acids	HMDB0000448	0.128	0.094	0.1732	2.21
Q3/Q1		Adipic acid	0.100	0.094	0.2852	
Q4/Q1			0.221	0.099	0.0254	
Per quartile			0.055	0.032	0.0830	
Q2/Q1	Fatty acids	HMDB0000555	0.196	0.094	0.0385	1.86
Q3/Q1		3-Methyladipic acid	0.049	0.094	0.6051	
Q4/Q1			0.249	0.099	0.0121	
Per quartile			0.045	0.032	0.1598	
Q2/Q1	Fatty acids	HMDB0000784	0.139	0.094	0.1397	2.15
Q3/Q1		Azelaic acid	0.179	0.094	0.0556	
Q4/Q1			0.195	0.099	0.0481	
Per quartile			0.052	0.032	0.1000	

Table S13 (continued)

Table S13 (continued)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Q2/Q1	Fatty acids	HMDB0000792	0.016	0.093	0.8622	2.11
Q3/Q1		Sebacic acid	0.167	0.092	0.0707	
Q4/Q1			0.220	0.097	0.0245	
Per quartile			0.069	0.031	0.0281	
Q2/Q1	Fatty acids	HMDB0000857	-0.017	0.094	0.8583	1.60
Q3/Q1		Pimelic acid	0.155	0.094	0.0986	
Q4/Q1			0.236	0.099	0.0174	
Per quartile			0.082	0.032	0.0103	
Q2/Q1	Fatty acids	HMDB0000893	-0.002	0.094	0.9857	2.15
Q3/Q1		Suberic acid	0.122	0.094	0.1932	
Q4/Q1			0.216	0.099	0.0290	
Per quartile			0.064	0.032	0.0449	
Q2/Q1	Indoles	HMDB0002302	0.010	0.092	0.9093	1.89
Q3/Q1		3-Indolepropionic acid	0.133	0.091	0.1468	
Q4/Q1			0.294	0.096	0.0024	
Per quartile			0.103	0.031	0.0010	
Total carotenoids						
Q2/Q1	Fatty acids	HMDB0000448	0.083	0.092	0.3661	2.26
Q3/Q1		Adipic acid	0.179	0.095	0.0605	
Q4/Q1			0.263	0.096	0.0064	
Per quartile			0.083	0.031	0.0080	
Q2/Q1	Bile acids	HMDB0000502	-0.071	0.092	0.4411	1.61
Q3/Q1		3-DHCA	-0.157	0.096	0.1009	
Q4/Q1			-0.213	0.097	0.0277	
Per quartile			-0.071	0.031	0.0238	
Q2/Q1	Fatty acids	HMDB0000555	0.114	0.093	0.2201	1.82
Q3/Q1		3-Methyladipic acid	0.184	0.096	0.0560	
Q4/Q1			0.201	0.097	0.0388	
Per quartile			0.057	0.031	0.0703	
Q2/Q1	Phenylpropanoic acids	HMDB0000764	0.138	0.091	0.1329	2.03
Q3/Q1		Hydrocinnamic acid	0.144	0.095	0.1292	
Q4/Q1			0.204	0.096	0.0341	
Per quartile			0.070	0.031	0.0248	
Q2/Q1	Indoles	HMDB0002302	-0.029	0.090	0.7434	1.77

Table S13 (continued)

Table S13 (continued)

Carotenoids	Fecal metabolites		$\beta$	SE	P value	VIP
	Class	HMDB				
Q3/Q1		3-Indolepropionic acid	0.161	0.093	0.0850	
Q4/Q1			0.291	0.094	0.0020	
Per quartile			0.108	0.031	0.0004	
Q2/Q1	Phenylpropanoic acids	HMDB0011743	0.135	0.092	0.1410	1.98
Q3/Q1		2-Phenylpropionate	0.137	0.095	0.1518	
Q4/Q1			0.201	0.096	0.0372	
Per quartile			0.069	0.031	0.0271	

Orthogonal partial least squares discrimination analysis (variable importance in the projection, VIP >1) and multivariable linear regression were used to identify potential fecal metabolites related to carotenoids, adjusted for age and sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, total energy intake, SFA intake, and fiber intake.

**Table S14** The association of overlapping fecal metabolites with NAFLD

HMDB	Name	Class	VIP	NAFLD persistent and incident			New-onset NAFLD		MRI-identified NAFLD	
				Case/all=533/787			Case/all=180/434		Case/all=199/449	
				OR (95% CI)	P	FDR	OR (95% CI)	P	OR (95% CI)	P
HMDB0000448	Adipic acid	Fatty acids	2.19	0.74 (0.63, 0.86)	1.29E-04	0.004	0.76 (0.62, 0.94)	0.010	0.91 (0.75, 1.11)	0.351
HMDB0000792	Sebacic acid	Fatty acids	2.10	0.76 (0.64, 0.91)	0.003	0.017	0.78 (0.62, 0.98)	0.029	0.69 (0.54, 0.87)	0.002
HMDB0000893	Suberic acid	Fatty acids	2.31	0.76 (0.65, 0.89)	5.24E-04	0.007	0.81 (0.66, 0.99)	0.039	0.75 (0.61, 0.91)	0.005
HMDB0000784	Azelaic acid	Fatty acids	2.23	0.77 (0.66, 0.9)	8.66E-04	0.007	0.82 (0.67, 1)	0.049	0.74 (0.6, 0.9)	0.003
HMDB0000857	Pimelic acid	Fatty acids	1.66	0.8 (0.69, 0.94)	0.007	0.026	0.91 (0.74, 1.11)	0.350	0.86 (0.7, 1.05)	0.142
HMDB0000555	3-Methyladipic acid	Fatty acids	2.06	0.79 (0.68, 0.92)	0.003	0.017	0.83 (0.68, 1.02)	0.075	0.79 (0.65, 0.96)	0.018
HMDB0000764	Hydrocinnamic acid	Phenylpropanoic acids	1.97	0.8 (0.68, 0.94)	0.007	0.026	0.77 (0.62, 0.95)	0.016	1 (0.82, 1.22)	0.975
HMDB0011743	2-Phenylpropionate	Phenylpropanoic acids	1.94	0.8 (0.68, 0.94)	0.008	0.027	0.76 (0.62, 0.95)	0.014	1 (0.82, 1.22)	0.965
HMDB0002302	3-Indolepropionic acid	Indoles	1.37	0.81 (0.68, 0.96)	0.017	0.044	0.81 (0.64, 1.01)	0.056	1.13 (0.91, 1.4)	0.281
HMDB0000502	3-DHCA	Bile acids	1.22	1.21 (1.03, 1.41)	0.018	0.045	1.13 (0.93, 1.38)	0.230	1.3 (1.08, 1.58)	0.007
HMDB0000506	a-muricholic acid and b-muricholic acid	Bile acids	1.68	1.21 (1.04, 1.4)	0.015	0.043	1.24 (1.01, 1.53)	0.039	1.29 (1.07, 1.56)	0.009
HMDB0000811	Muro-cholic acid	Bile acids	1.74	1.21 (1.04, 1.41)	0.015	0.043	1.16 (0.95, 1.41)	0.154	1.34 (1.1, 1.63)	0.003
HMDB0000708	Glycoursodeoxycholic acid	Bile acids	1.60	1.21 (1.04, 1.42)	0.015	0.043	1.25 (1.03, 1.53)	0.025	1.2 (0.99, 1.47)	0.069
HMDB0000637	Glycochenodeoxycholic acid	Bile acids	1.49	1.23 (1.05, 1.43)	0.010	0.034	1.25 (1.03, 1.52)	0.022	1.15 (0.94, 1.4)	0.171
HMDB0000664	bHyodeoxycholic acid	Bile acids	1.88	1.24 (1.06, 1.44)	0.006	0.026	1.17 (0.96, 1.42)	0.121	1.3 (1.07, 1.58)	0.007
HMDB0000467	7ketoLCA	Bile acids	1.83	1.27 (1.1, 1.48)	0.002	0.013	1.26 (1.03, 1.54)	0.027	1.3 (1.07, 1.57)	0.008
HMDB0000518	Chenodeoxycholic acid	Bile acids	1.70	1.31 (1.12, 1.52)	5.86E-04	0.007	1.3 (1.06, 1.59)	0.012	1.28 (1.06, 1.56)	0.012
HMDB0005065	Oleylcarnitine C18:1	Carnitines	1.81	1.31 (1.13, 1.52)	4.21E-04	0.007	1.33 (1.07, 1.65)	0.011	1.41 (1.15, 1.73)	0.001
HMDB0000222	Palmitoylcarnitine	Carnitines	1.92	1.35 (1.16, 1.57)	1.22E-04	0.004	1.44 (1.15, 1.8)	0.001	1.38 (1.13, 1.68)	0.002

Orthogonal partial least squares discrimination analysis (variable importance in the projection, VIP >1) and multivariable logistic regression were used to identify potential fecal metabolites related to NAFLD, adjusted for age, sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy, saturated fatty acids and fiber intake.



**Table S15** The association of carotenoids with overlapping differential serum metabolites (N=896)

Carotenoids	Serum metabolites		$\beta$	SE	P value	VIP
	Class	HMDB/name				
Lutein/zeaxanthin						
Q2/Q1	Organic acids	HMDB0000208	0.029	0.091	0.747	1.53
Q3/Q1		Oxoglutaric acid	-0.189	0.093	0.042	
Q4/Q1			-0.234	0.094	0.013	
Per quartile			-0.053	0.029	0.073	
Q2/Q1	Carnitines	HMDB0000688	-0.027	0.089	0.765	3.2
Q3/Q1		Isovalerylcarnitine	-0.193	0.090	0.033	
Q4/Q1			-0.439	0.092	2.02E-06	
Per quartile			-0.133	0.029	7.75E-06	
Q2/Q1	Carnitines	HMDB0000848	0.268	0.089	0.003	1.59
Q3/Q1		Stearoylcarnitine	0.198	0.090	0.028	
Q4/Q1			0.220	0.091	0.016	
Per quartile			0.045	0.029	0.129	
Q2/Q1	Fatty acids	HMDB0033724	0.046	0.090	0.608	2.21
Q3/Q1		Undecylenic acid	0.051	0.092	0.580	
Q4/Q1			0.243	0.093	0.009	
Per quartile			0.067	0.030	0.025	
$\beta$ -cryptoxanthin						
Q2/Q1	Carnitines	HMDB0000688	-0.025	0.091	0.783	2.16
Q3/Q1		Isovalerylcarnitine	-0.169	0.091	0.064	
Q4/Q1			-0.228	0.093	0.014	
Per quartile			-0.075	0.029	0.010	
Q2/Q1	Fatty acids	HMDB0000893	0.026	0.095	0.779	1.19
Q3/Q1		Suberic acid	0.118	0.094	0.212	
Q4/Q1			0.242	0.097	0.012	
Per quartile			0.085	0.031	0.006	
Q2/Q1	Benzenoids	HMDB0001856	0.108	0.094	0.250	1.3
Q3/Q1		Protocatechuic acid	0.138	0.094	0.141	
Q4/Q1			0.259	0.096	0.007	
Per quartile			0.072	0.030	0.018	
Q2/Q1	Indoles	HMDB0002302	0.048	0.093	0.608	1.26
Q3/Q1		3-Indolepropionic acid	0.084	0.093	0.362	
Q4/Q1			0.231	0.095	0.015	
Per quartile			0.069	0.030	0.023	

Table S15 (continued)

**Table S15** (continued)

Carotenoids	Serum metabolites		$\beta$	SE	P value	VIP
	Class	HMDB/name				
lycopene						
Q2/Q1	Amino acids	HMDB0000148	-0.140	0.094	0.136	1.87
Q3/Q1		L-Glutamic acid	-0.206	0.095	0.030	
Q4/Q1			-0.286	0.094	0.003	
Per quartile			-0.071	0.030	0.017	
Q2/Q1	Amino acids	HMDB0000159	-0.083	0.094	0.378	2.2
Q3/Q1		L-Phenylalanine	-0.146	0.095	0.123	
Q4/Q1			-0.348	0.094	0.000	
Per quartile			-0.090	0.030	0.002	
Q2/Q1	Organic acids	HMDB0000190	0.062	0.094	0.510	1.38
Q3/Q1		L-Lactic acid	-0.063	0.095	0.512	
Q4/Q1			-0.204	0.095	0.032	
Per quartile			-0.057	0.030	0.057	
Q2/Q1	Organic acids	HMDB0000243	-0.193	0.095	0.041	2.12
Q3/Q1		Pyruvic acid	-0.110	0.095	0.248	
Q4/Q1			-0.219	0.095	0.022	
Per quartile			-0.032	0.030	0.282	
Q2/Q1	Amino acids	HMDB0000267	-0.109	0.095	0.250	1.73
Q3/Q1		Pyroglutamic acid	-0.163	0.096	0.090	
Q4/Q1			-0.206	0.095	0.031	
Per quartile			-0.044	0.030	0.142	
$\alpha$ -carotene						
Q2/Q1	Amino acids	HMDB0000123	0.108	0.092	0.243	1.21
Q3/Q1		Glycine	0.217	0.096	0.024	
Q4/Q1			0.268	0.096	0.005	
Per quartile			0.077	0.030	0.011	
Q2/Q1	Amino acids	HMDB0000168	0.110	0.089	0.215	1.03
Q3/Q1		L-Asparagine	0.151	0.093	0.104	
Q4/Q1			0.312	0.093	0.001	
Per quartile			0.088	0.029	0.003	
Q2/Q1	Amino acids	HMDB0000641	0.113	0.091	0.216	1.22
Q3/Q1		L-Glutamine	0.098	0.095	0.304	
Q4/Q1			0.268	0.095	0.005	
Per quartile			0.066	0.030	0.029	

**Table S15** (continued)

Table S15 (continued)

Carotenoids	Serum metabolites		$\beta$	SE	P value	VIP
	Class	HMDB/name				
Q2/Q1	Carnitines	HMDB0000848	0.092	0.090	0.304	1.1
Q3/Q1		Stearoylcarnitine	0.139	0.094	0.139	
Q4/Q1			0.287	0.093	0.002	
Per quartile			0.078	0.030	0.008	
Q2/Q1	Indoles	HMDB0002302	0.112	0.093	0.227	1.24
Q3/Q1		3-Indolepropionic acid	0.225	0.097	0.020	
Q4/Q1			0.295	0.096	0.002	
Per quartile			0.095	0.031	0.002	
Q2/Q1	Fatty acids	HMDB0033724	0.043	0.091	0.635	1.68
Q3/Q1		Undecylenic acid	0.066	0.095	0.491	
Q4/Q1			0.235	0.095	0.014	
Per quartile			0.069	0.030	0.023	
$\beta$ -carotene						
Q2/Q1	Amino acids	HMDB0000123	0.107	0.092	0.245	1.55
Q3/Q1		Glycine	0.318	0.092	0.001	
Q4/Q1			0.343	0.097	0.000	
Per quartile			0.103	0.031	0.001	
Q2/Q1	Amino acids	HMDB0000159	-0.044	0.092	0.635	1.44
Q3/Q1		L-Phenylalanine	-0.174	0.091	0.057	
Q4/Q1			-0.234	0.096	0.015	
Per quartile			-0.043	0.030	0.153	
Q2/Q1	Organic acids	HMDB0000208	0.014	0.093	0.883	1.3
Q3/Q1		Oxoglutaric acid	-0.120	0.092	0.194	
Q4/Q1			-0.212	0.097	0.030	
Per quartile			-0.027	0.031	0.384	
Q2/Q1	Organic acids	HMDB0000243	-0.073	0.092	0.426	2.17
Q3/Q1		Pyruvic acid	-0.193	0.091	0.035	
Q4/Q1			-0.357	0.096	0.000	
Per quartile			-0.075	0.030	0.014	
Q2/Q1	Organic acids	HMDB0000426	-0.114	0.092	0.215	2.03
Q3/Q1		Citramalic acid	-0.148	0.091	0.107	
Q4/Q1			-0.321	0.096	0.001	
Per quartile			-0.061	0.031	0.048	
Q2/Q1	Carnitines	HMDB0000688	-0.125	0.091	0.173	2.13

Table S15 (continued)

Table S15 (continued)

Carotenoids	Serum metabolites		$\beta$	SE	P value	VIP
	Class	HMDB/name				
Q3/Q1		Isovalerylcarnitine	-0.163	0.091	0.073	
Q4/Q1			-0.311	0.096	0.001	
Per quartile			-0.079	0.031	0.011	
Q2/Q1	Carnitines	HMDB0000848	-0.081	0.090	0.368	1.64
Q3/Q1		Stearoylcarnitine	0.155	0.089	0.083	
Q4/Q1			0.249	0.094	0.008	
Per quartile			0.076	0.030	0.013	
Q2/Q1	Fatty acids	HMDB0033724	0.084	0.092	0.358	2.36
Q3/Q1		Undecylenic acid	0.074	0.091	0.418	
Q4/Q1			0.288	0.096	0.003	
Per quartile			0.078	0.031	0.012	
Total carotenoids						
Q2/Q1	Amino acids	HMDB0000123	0.098	0.091	0.282	1.64
Q3/Q1		Glycine	0.210	0.094	0.026	
Q4/Q1			0.188	0.095	0.048	
Per quartile			0.045	0.031	0.141	
Q2/Q1	Organic acids	HMDB0000190	0.005	0.090	0.956	1.02
Q3/Q1		L-Lactic acid	-0.023	0.093	0.807	
Q4/Q1			-0.203	0.095	0.032	
Per quartile			-0.031	0.030	0.300	
Q2/Q1	Organic acids	HMDB0000243	-0.142	0.090	0.115	1.53
Q3/Q1		Pyruvic acid	-0.211	0.093	0.025	
Q4/Q1			-0.202	0.095	0.033	
Per quartile			-0.027	0.030	0.362	
Q2/Q1	Carnitines	HMDB0000688	-0.148	0.089	0.098	2.72
Q3/Q1		Isovalerylcarnitine	-0.122	0.092	0.187	
Q4/Q1			-0.401	0.093	0.000	
Per quartile			-0.101	0.030	0.001	
Q2/Q1	Amino acids	HMDB0000812	0.084	0.088	0.343	1.25
Q3/Q1		N-Acetyl-L-aspartic acid	0.254	0.092	0.006	
Q4/Q1			0.206	0.093	0.027	
Per quartile			0.064	0.030	0.032	
Q2/Q1	Carnitines	HMDB0000848	0.175	0.088	0.047	2.15
Q3/Q1		Stearoylcarnitine	0.185	0.092	0.044	

Table S15 (continued)

**Table S15** (continued)

Carotenoids	Serum metabolites		$\beta$	SE	P value	VIP
	Class	HMDB/name				
Q4/Q1			0.234	0.093	0.012	
Per quartile			0.052	0.030	0.083	
Q2/Q1	Indoles	HMDB0002302	0.069	0.091	0.449	1.19
Q3/Q1		3-Indolepropionic acid	0.222	0.094	0.019	
Q4/Q1			0.253	0.096	0.008	
Per quartile			0.085	0.031	0.006	
Q2/Q1	Fatty acids	HMDB0033724	0.097	0.090	0.282	2.26
Q3/Q1		Undecylenic acid	0.087	0.093	0.348	
Q4/Q1			0.264	0.094	0.005	
Per quartile			0.073	0.030	0.017	

Orthogonal partial least squares discrimination analysis (variable importance in the projection, VIP >1) and multivariable linear regression were used to identify potential serum metabolites related to carotenoids, adjusted for age and sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy, saturated fatty acids and fiber intake.

**Table S16** The association of overlapping serum metabolites with NAFLD groups

HMDB	Name	Class	VIP	NAFLD persistent and incident			New-onset NAFLD		MRI-identified NAFLD	
				Case/all=535/789			Case/all=180/434		Case/all=199/450	
				OR (95% CI)	P	FDR	OR (95% CI)	P	OR (95% CI)	P
HMDB0000812	N-Acetyl-L-aspartic acid	Amino acids	1.99	0.73 (0.61, 0.86)	2.43E-04	0.001	0.66 (0.52, 0.83)	5.95E-04	0.92 (0.74, 1.13)	0.408
HMDB0000123	Glycine	Amino acids	2.52	0.74 (0.64, 0.87)	2.06E-04	0.001	0.82 (0.67, 1.01)	0.062	0.76 (0.62, 0.93)	0.007
HMDB0000641	L-Glutamine	Amino acids	2.29	0.79 (0.67, 0.92)	0.003	0.010	0.77 (0.62, 0.94)	0.012	0.97 (0.8, 1.18)	0.749
HMDB0001856	Protocatechuic acid	Benzenoids	1.68	0.76 (0.63, 0.93)	0.007	0.024	0.9 (0.67, 1.2)	0.475	0.79 (0.63, 0.99)	0.040
HMDB0000848	Stearoylcarnitine	Carnitines	2.38	0.78 (0.66, 0.91)	0.002	0.009	0.85 (0.69, 1.05)	0.124	0.78 (0.63, 0.95)	0.016
HMDB0033724	Undecylenic acid	Fatty acids	1.97	0.82 (0.7, 0.96)	0.016	0.042	0.91 (0.75, 1.12)	0.382	0.85 (0.7, 1.03)	0.093
HMDB0002302	3-Indolepropionic acid	Indoles	1.22	0.82 (0.7, 0.97)	0.019	0.046	0.85 (0.68, 1.05)	0.126	0.9 (0.74, 1.1)	0.316
HMDB0000159	L-Phenylalanine	Amino acids	1.46	1.42 (1.21, 1.67)	1.49E-05	1.97E-04	1.2 (0.98, 1.48)	0.083	1.55 (1.27, 1.9)	2.31E-05
HMDB0000267	Pyroglutamic acid	Amino acids	1.79	1.55 (1.32, 1.82)	1.14E-07	3.36E-06	1.31 (1.06, 1.62)	0.012	1.66 (1.34, 2.04)	2.50E-06
HMDB0000148	L-Glutamic acid	Amino acids	2.03	1.63 (1.38, 1.92)	8.57E-09	5.66E-07	1.41 (1.14, 1.75)	0.002	1.69 (1.37, 2.09)	1.15E-06
HMDB0000688	Isovalerylcarnitine	Carnitines	1.29	1.21 (1.05, 1.4)	0.010	0.031	1.09 (0.9, 1.32)	0.399	1.12 (0.87, 1.44)	0.381
HMDB0000893	Suberic acid	Fatty acids	1.25	1.27 (1.09, 1.48)	0.002	0.009	1.19 (0.97, 1.46)	0.100	1.1 (0.91, 1.32)	0.326
HMDB0000190	L-Lactic acid	Organic acids	1.42	1.35 (1.15, 1.58)	1.67E-04	0.001	1.25 (1.01, 1.54)	0.038	1.35 (1.11, 1.65)	0.003
HMDB0000243	Pyruvic acid	Organic acids	2.57	1.41 (1.2, 1.65)	2.75E-05	2.61E-04	1.2 (0.98, 1.47)	0.082	1.35 (1.11, 1.65)	0.003
HMDB0000208	Oxoglutaric acid	Organic acids	1.95	1.53 (1.29, 1.81)	7.53E-07	1.24E-05	1.27 (1.01, 1.59)	0.040	1.7 (1.37, 2.1)	1.16E-06
HMDB0000426	Citramalic acid	Organic acids	2.79	1.56 (1.32, 1.85)	1.53E-07	3.36E-06	1.13 (0.92, 1.38)	0.235	1.81 (1.44, 2.27)	2.63E-07

Orthogonal partial least squares discrimination analysis (variable importance in the projection, VIP >1) and multivariable logistic regression were used to identify potential serum metabolites related to NAFLD groups, adjusted for age and sex, household income, physical activity, multivitamin use, smoking, alcohol drinking, education level, energy, saturated fatty acids and fiber intake.