Supplementary Materials

Myosteatosis mediates the link between specific dietary components and colorectal carcinogenesis: PPLSS multi-center study

Chun-wei Li PhD[1†], Yifang Hsieh MSc[2,3,4†], Chao Li MD[2,3], Yu Zhang MSc[5], Ling-juan Jiang PhD[6], Song-lin Yu PhD[7], Long-Chun Dong MSc[8], Zhi-quan Tan MSc[9], Judong Zhang BSc[2,3], Jing Xu MD[2,3], Kang Yu MD[5]

[1]*Department of Clinical Nutrition, Institute of Clinical Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*
[2]*Department of General Surgery, Tianjin Union Medical Center, The First Affiliated Hospital of Nankai University, Nankai University, Tianjin, China*
[3]*Tianjin Key Laboratory of General Surgery in construction, Tianjin Union Medical Center, Tianjin, China*
[4]*School of Medicine, Nankai University, Tianjin, China*
[5]*Department of Clinical Nutrition, Department of Health Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*
[6]*Biomarker Discovery and Validation Facility, Institute of Clinical Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*
[7]*Department of Clinical Laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*
[8]*Department of radiology, Tianjin Union Medicine Centre, The Affiliated Hospital of Nankai University, Nankai University, Tianjin, China*
[9]*Department of Information, Tianjin Union Medicine Centre, The Affiliated Hospital of Nankai University, Nankai University, Tianjin, China*
[†]*Both authors contributed equally to this manuscript*

Corresponding Author:
Prof Kang Yu, Department of Clinical Nutrition and Department of Health Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences (CAMS) and Peking Union Medical College (PUMC), No. 1 Shuaifuyuan, Dongcheng District, Beijing 100730, China.
Tel: +86-10-69155550. Fax: +86-10-65253037.
Email: yuk1997@sina.com
Dr Jing Xu, Department of General Surgery, Tianjin Union Medical Center, The Affiliated Hospital of Nankai University, Nankai University, No. 190 Jieyuan street, Hongqiao District, Tianjin 300121, China.
Tel: +86-10-27557221. Fax: +86-10-87721989.
Email: xujingdoc@126.com
Manuscript received 09 June 2025. Initial review completed 09 December 2025. Revision accepted 15 January 2025.
doi: 10.6133/apjcn.202604_35(2).0008

Supplementary Table 1. Nine nutrients derived from 10 food groups in statistical analyses with principal component analysis

Nutrients	Food/ food group
Total energy	Grain, Red meat, poultry, Fish, eggs, milk, Bean and its products, Vegetables, Fruits, Nuts
Total carbohydrate	Grain, milk, Bean and its products, Vegetables, Fruits, Nuts
Total protein	Grain, Red meat, poultry, Fish, eggs, milk, Bean and its products, Vegetables, Fruits, Nuts
Animal protein	Red meat, poultry, Fish, eggs, milk
Bean protein	Bean and its products
Total fat	Red meat, poultry, Fish, eggs, milk, Bean and its products,
Meat fat	Red meat, poultry
Total cholesterol	Red meat, poultry, Fish, eggs, milk
Meat cholesterol	Red meat, poultry
food group	Food
Grain	Rice, noodles, mantou, dumplings, Potatoes, corn
Red meat	Beef, lamb, pork
poultry	Chicken, duck
Fish	freshwater fish, marine fish
Eggs	Eggs
milk	Whole milk, skim milk, yogurt, sugar free yogurt, cheese, milk powder
Bean and its products	Bean, tofu, soybean milk, dried bean curd, dried beans, and other soy products
Vegetables	Legumes vegetables, leafy vegetables, solanaceous vegetables, bulb vegetables, melon vegetables
Fruits	Fruits
Nuts	Nuts

Supplementary Table 2. Body composition with CRC risks by sex

Variables	Men		Women		
	OR (95CI%)	*p*	Variables	*OR (95CI%)*	*p*
Model 1[†]					
NAMA, cm^2	1.015 (0.987, 1.044)	0.298	NAMA	1.002 (0.952, 1.051)	0.947
LAMA, cm^2	1.032 (0.985, 1.084)	0.188	LAMA	1.049 (0.980, 1.127)	0.181
SMD, HU	0.907 (0.772, 1.061)	0.228	SMD	0.927 (0.768, 1.117)	0.423
SAI, cm^2/m^2	0.970 (0.944, 0.994)	0.019	VAI	0.989 (0.971, 1.006)	0.198
Model 2[‡]					
NAMA, cm^2	1.002 (0.969, 1.037)	0.885	NAMA	1.004 (0.947, 1.061)	0.889
LAMA, cm^2	0.987 (0.932, 1.047)	0.668	LAMA	1.034 (0.954, 1.124)	0.427
SMD, HU	0.781 (0.631, 0.950)	0.017	SMD	0.880 (0.709, 1.088)	0.235
SAI, cm^2/m^2	0.964 (0.931, 0.994)	0.024	VAI	0.987 (0.966, 1.007)	0.232
Model 3[§]					
LDL, mmol/L	1.480 (0.542, 4.302)	0.452	LDL	0.609 (0.048, 5.776)	0.684
HDL, mmol/L	0.001 (0.00003, 0.016)	<0.001	HDL	0.0001 (0.00002, 0.004)	<0.001
FFA, μmol/ml	1.002 (1.000, 1.005)	0.104	FFA	1.003 (1.000, 1.007)	0.085
TG, mmol/L	0.257 (0.078, 0.653)	0.010	TG	0.172 (0.030, 0.705)	0.020
NAMA, cm^2	0.989 (0.936, 1.046)	0.705	NAMA	1.130 (1.002, 1.291)	0.051
LAMA, cm^2	0.937 (0.855, 1.021)	0.142	LAMA	0.916 (0.787, 1.059)	0.238
SMD, HU	0.713 (0.524, 0.935)	0.020	SMD	0.652 (0.425, 0.956)	0.035
SAI, cm^2/m^2	0.959 (0.906, 1.013)	0.141	VAI	1.027 (0.984, 1.075)	0.225

NAMA: Normal Attenuation Muscle Area, LAMA: Low Attenuation Muscle Area, MA: Myosteatosis; SAI: Subcutaneous Adipose Index: Visceral adipose index, LDL: Low density lipoprotein, HDL: High density lipoprotein, FFA: Free fatty acid, TG: Triglyceride.

[†]Model 1 did not have any adjustments

[‡]Model 2 was adjusted for age, exercise, smoking, and drinking

[§]Model 3 was adjusted for age, exercise, smoking, drinking, BMI, metabolic syndrome, LDL, HDL, FFA, and TG

Supplementary Table 3. Dietary intake features of participants[†]

	Total			Men		
	Control (n=144)	CRC (n=163)	*p*	Control (n=61)	CRC (n=93)	*p*
Total energy, kcal/day	1655.98 (1413.16, 1880.15)	1421.47 (1195.91, 1566.15)	<0.001	1737.35 (1615.81, 1899.58)	1462.49 (1325.36, 1636.05)	<0.001
Total carbohydrate, g/day	254.01 (187.36, 267.58)	257.19 (178.54, 258.20)	0.037	262.42 (249.13, 269.49)	257.63 (196.38, 258.20)	<0.001
Total protein, g/day	72.40 (61.74, 85.41)	51.10 (40.78, 64.94)	<0.001	75.99 (67.99, 87.05)	54.678 (43.16, 69.13)	<0.001
Animal protein, g/day	44.70 (35.04, 56.41)	22.98 (12.61, 38.54)	<0.001	34.57 (39.14, 56.23)	17.01 (15.43, 42.46)	<0.001
Bean protein, g/day	2.57 (0.90, 2.57)	1.29 (0.30, 2.57)	<0.001	2.57 (0.6, 2.57)	1.29 (0.30, 2.57)	0.002
Total fat, g/day	33.03 (25.42, 42.32)	15.89 (9.31, 29.96)	<0.001	34.57 (28.46, 42.22)	17.01 (10.71, 30.70)	<0.001
Meat fat, g/day	15.43 (10.29, 18.86)	7.66 (4.23, 13.71)	<0.001	15.43 (10.29, 20.57)	8.57 (6.86, 13.71)	<0.001
Total cholesterol, mg/day	367.00 (319.09, 445.22)	210.16 (76.66, 303.07)	<0.001	375.79 (333.25, 432.69)	221.10 (98.05, 329.90)	<0.001
Meat cholesterol, mg/day	101.97 (69.69, 128.61)	49.26 (27.74, 91.21)	<0.001	101.97 (72.25, 144.49)	58.93 (43.04, 101.45)	<0.001
Wholegrain, n (%)			0.282			0.221
all	1 (0.69)	3 (1.84)		0 (0.00)	1 (1.08)	
all-2/3	2 (1.39)	5 (3.07)		1 (1.64)	3 (3.23)	
2/3-half	13 (9.03)	14 (8.59)		6 (9.84)	6 (6.45)	
half-1/3	42 (29.17)	33 (20.25)		18 (29.51)	14 (15.05)	
<1/3	68 (47.22)	77 (47.24)		26 (42.62)	45 (48.39)	
no	18 (12.50)	31 (19.02)		10 (16.39)	24 (25.81)	

	Women		
	Control (n=83)	CRC (n=70)	*p*
Total energy, kcal/day	1521.53 (1326.5, 1816.85)	1344.76 (1117.30, 1473.94)	<0.001
Total carbohydrate, g/day	196.95 (179.98, 258.52)	196.64 (177.63, 257.63)	0.315
Total protein, g/day	69.46 (56.45, 83.95)	46.60 (39.03, 61.06)	<0.001
Animal protein, g/day	43.81 (31.55, 56.48)	20.21 (11.48, 36.73)	<0.001
Bean protein, g/day	2.57 (1.29, 2.57)	1.29 (0.55, 2.57)	<0.001
Total fat, g/day	31.87 (23.04, 42.12)	14.04 (8.58, 26.5)	<0.001
Meat fat, g/day	13.71 (7.66, 17.14)	6.86 (4.23, 11.09)	<0.001
Total cholesterol, mg/day	349.49 (292.08, 468.03)	171.16 (70.12, 274.33)	<0.001
Meat cholesterol, mg/day	96.33 (49.26, 115.3)	48.16 (27.74, 70.79)	<0.001
Wholegrain, n (%)			0.896
all	1 (1.20)	2 (2.86)	
all-2/3	1 (1.20)	2 (2.86)	
2/3-half	7 (8.43)	8 (11.43)	
half-1/3	24 (28.92)	19 (27.14)	
<1/3	42 (50.60)	32 (45.71)	
no	8 (9.64)	7 (10.00)	

[†]Mann-Whitney test was used for continuous variables and Fisher's exact test was used for categorical variables.

Supplementary Table 4. Factor loadings of nutrient groups for dietary intake

Variables	PC1	PC2	PC3
Total			
Total carbohydrate		0.946†	0.157
Total fat	0.418		
Total cholesterol	0.395		
Total protein	0.426	0.177	
Meat fat	0.382	-0.168	
Meat cholesterol	0.381	-0.167	
Animal protein	0.428		
Bean protein		0.125	-0.980
Male			
Total carbohydrate		0.843	0.399
Total fat	0.413		0.104
Total cholesterol	0.389		0.154
Total protein	0.422	0.132	0.108
Meat fat	0.385	-0.182	-0.124
Meat cholesterol	0.383	-0.191	-0.126
Animal protein	0.423		
Bean protein	0.141	0.445	-0.868
Female			
Total carbohydrate		0.773	0.537
Total fat	0.421		
Total cholesterol	0.401		0.143
Total protein	0.428	0.112	0.170
Meat fat	0.379		-0.296
Meat cholesterol	0.379		-0.297
Animal protein	0.433		
Bean protein		-0.618	0.693

†Only those with absolute factor loads greater than 0.1 are listed

Supplementary Table 5. Associations of daily nutrient intakes with dietary patterns between the low (Q1–Q3) and high (Q4) groups in general[†]

	High animal-derived nutrients (PC1)			High carbohydrate (PC2)		
	Low (Q1-Q3)	High (Q4)	*p*	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1425.5 (1250.24, 1569.37)	1884.38 (1701.38, 1988.49)	<0.001	1448.15 (1254.00, 1634.39)	1793.57 (1462.5, 1920.53)	<0.001
Total carbohydrate (g/d)	255.06 (180.31, 258.2)	258.74 (196.95, 276.08)	<0.001	196.95 (178.32, 257.63)	268.45 (258.2, 276.95)	<0.001
Total protein (g/d)	55.33 (43.17, 65.82)	87.05 (80.59, 95.96)	<0.001	60.70 (46.79, 73.91)	70.95 (49.54, 82.52)	0.013
Animal protein (g/d)	27.29 (16.22, 38.26)	57.22 (52.9, 68.18)	<0.001	35.93 (20.98, 45.89)	38.12 (17.35, 50.22)	0.586
Bean protein (g/d)	1.29 (0.60, 2.57)	2.57 (1.29, 2.57)	0.002	1.29 (0.60, 2.57)	2.57 (1.29, 2.57)	<0.001
Total fat (g/d)	19.85 (11.40, 28.82)	43.08 (38.74, 51.2)	<0.001	25.25 (14.67, 34.45)	31.18 (13.13, 39.83)	0.149
Meat fat (g/d)	7.66 (4.23, 13.59)	20.57 (15.43, 24.00)	<0.001	12.00 (6.86, 18.6)	10.29 (4.23, 13.71)	0.001
Total cholesterol (mg/d)	252.32 (114.25, 329.98)	453.25 (408.05, 528.95)	<0.001	289.61 (181.09, 372.65)	333.25 (196.04, 409.96)	0.179
Meat cholesterol (mg/d)	49.26 (27.74, 89.80)	144.49 (104.53, 166.28)	<0.001	76.88 (48.16, 122.32)	69.69 (27.74, 101.45)	0.001

	Low bean protein (PC3)		
	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1509.86 (1319.73, 1765.44)	1472.46 (1338.22, 1737.35)	0.831
Total carbohydrate (g/d)	257.63 (182.99, 259.45)	257.06 (186.38, 257.93)	0.151
Total protein (g/d)	62.38 (47.9, 76.98)	63.59 (46.42, 78.27)	0.990
Animal protein (g/d)	36.03 (20.53, 46.44)	37.59 (18.53, 53.36)	0.518
Bean protein (g/d)	2.57 (1.29, 2.57)	0.3 (0.00, 0.30)	<0.001
Total fat (g/d)	25.93 (14.9, 35.75)	27.00 (12.55, 39.36)	0.898
Meat fat (g/d)	10.29 (6.86, 15.43)	13.71 (6.86, 17.14)	0.402
Total cholesterol (mg/d)	297.45 (181.79, 374.39)	330.27 (180.86, 405.8)	0.552
Meat cholesterol (mg/d)	69.69 (48.16, 104.53)	91.21 (48.16, 117.85)	0.342

[†]Mann-Whitney test was used for continuous variables.

Supplementary Table 6. Associations of daily nutrient intakes with dietary patterns between the low (Q1–Q3) and high (Q4) groups in men[†]

	High animal-derived nutrients (PC1)			High carbohydrate (PC2)		
	Low (Q1-Q3)	High (Q4)	*p*	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1491.00 (1363.24, 1663.25)	1884.38 (1739.47, 1977.24)	<0.001	1538.57 (1380.16, 1709.63)	1796.24 (1511.7, 1905.67)	<0.001
Total carbohydrate (g/d)	257.57 (195.88, 258.98)	260.42 (253.58, 276.08)	0.003	257.19 (187.7, 258.2)	268.15 (258.46, 276.95)	<0.001
Total protein (g/d)	58.11 (46.12, 69.24)	86.37 (80.11, 95.54)	<0.001	64.46 (48.11, 76.57)	71.12 (53.91, 82.33)	0.088
Animal protein (g/d)	27.64 (16.27, 39.95)	56.23 (50.72, 65.51)	<0.001	36.83 (21.61, 46.78)	39.14 (21.09, 49.89)	0.664
Bean protein (g/d)	1.29 (0.30, 2.57)	2.57 (1.29, 2.57)	0.006	1.29 (0.30, 2.57)	2.57 (2.57, 2.57)	<0.001
Total fat (g/d)	19.63 (11.41, 30.17)	42.22 (37.39, 47.45)	<0.001	26.55 (14.98, 34.65)	31.77 (15.37, 38.59)	0.206
Meat fat (g/d)	10.29 (6.83, 13.71)	21.43 (18.43, 26.57)	<0.001	13.71 (6.86, 19.89)	10.29 (6.86, 14.57)	0.121
Total cholesterol (mg/d)	247.85 (127.36, 333.96)	432.50 (389.46, 499.29)	<0.001	315.66 (198.57, 385.28)	326.69 (188.05, 403.00)	0.559
Meat cholesterol (mg/d)	69.69 (43.04, 91.21)	144.49 (125.79, 180.75)	<0.001	91.21 (48.16, 134.12)	69.69 (45.6, 101.71)	0.083

	Low bean protein (PC3)		
	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1552.50 (1380.16, 1770.26)	1650.18 (1448.15, 1857.05)	0.069
Total carbohydrate (g/d)	257.63 (187.70, 260.88)	257.19 (257.10, 276.01)	0.244
Total protein (g/d)	66.75 (51.45, 78.87)	67.02 (47.66, 79.45)	0.564
Animal protein (g/d)	37.07 (21.61, 46.88)	40.50 (18.62, 50.28)	0.593
Bean protein (g/d)	2.57 (1.29, 2.57)	0.30 (0.00, 0.60)	<0.001
Total fat (g/d)	27.71 (15.05, 35.30)	28.48 (12.48, 39.43)	0.520
Meat fat (g/d)	10.29 (6.86, 20.57)	12.00 (6.86, 16.29)	0.524
Total cholesterol (mg/d)	316.10 (188.10, 380.83)	330.01 (216.65, 393.20)	0.565
Meat cholesterol (mg/d)	72.25 (48.16, 139.64)	75.33 (48.16, 113.88)	0.523

[†]Mann-Whitney test was used for continuous variables.

Supplementary Table 7. Associations of daily nutrient intakes with dietary patterns between the low (Q1–Q3) and high (Q4) groups in women[†]

	High animal-derived nutrients (PC1)			High carbohydrate (PC2)		
	Low (Q1-Q3)	High (Q4)	*p*	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1346.19 (1164.68, 1475.01)	1857.86 (1578.63, 2002.36)	<0.001	1347.41 (1164.68, 1531)	1705.17 (1433.57, 1890.66)	<0.001
Total carbohydrate (g/d)	196.32 (178.2, 257.63)	257.63 (188.02, 276.48)	0.002	183.85 (177.71, 215.7)	258.26 (257.63, 275.94)	<0.001
Total protein (g/d)	51.05 (41.53, 62.58)	87.49 (80.82, 96.12)	<0.001	56.93 (43.90, 69.72)	70.16 (45.47, 83.41)	0.046
Animal protein (g/d)	26.04 (15.39, 36.84)	59.03 (53.93, 68.92)	<0.001	32.97 (20.19, 44.89)	40.93 (17.12, 55.64)	0.452
Bean protein (g/d)	1.29 (1.29, 2.57)	2.57 (0.94, 2.57)	0.230	2.57 (1.29, 2.57)	1.29 (0.00, 1.29)	<0.001
Total fat (g/d)	19.85 (11.13, 27.32)	45.14 (39.29, 53.01)	<0.001	24.56 (14.2, 34.67)	27.89 (12.05, 39.06)	0.568
Meat fat (g/d)	7.26 (4.23, 12.8)	18.80 (15.43, 23.14)	<0.001	10.29 (5.34, 15.43)	10.29 (6.86, 15.43)	0.821
Total cholesterol (mg/d)	252.32 (102.32, 313.47)	486.72 (419.42, 554.70)	<0.001	282.00 (169.94, 365.90)	334.31 (154.78, 412.52)	0.486
Meat cholesterol (mg/d)	48.16 (27.74, 83.88)	122.32 (101.97, 152.96)	<0.001	69.69 (38.87, 101.97)	69.69 (47.16, 103.25)	0.816

	Low bean protein (PC3)		
	Low (Q1-Q3)	High (Q4)	*p*
Total energy (kcal/d)	1351.92 (1164.68, 1574)	1542.66 (1407.48, 1893.04)	<0.001
Total carbohydrate (g/d)	184.97 (177.71, 257.16)	258.2 (232.70, 268.45)	<0.001
Total protein (g/d)	57.61 (43.88, 70.51)	63.10 (50.99, 80.82)	0.039
Animal protein (g/d)	32.97 (19.29, 44.89)	35.48 (20.81, 50.15)	0.474
Bean protein (g/d)	1.29 (0.30, 2.57)	2.57 (2.57, 2.57)	<0.001
Total fat (g/d)	24.56 (13.00, 34.55)	30.00 (15.05, 39.01)	0.267
Meat fat (g/d)	10.69 (6.86, 15.43)	10.29 (3.85, 15.43)	0.118
Total cholesterol (mg/d)	281.24 (145.92, 363.06)	333.84 (238.26, 430.55)	0.148
Meat cholesterol (mg/d)	70.79 (48.16, 103.89)	69.69 (25.05, 101.97)	0.142

[†]Mann-Whitney test was used for continuous variables.

Supplementary Table 8. Dietary intake and CRC risk

	All participants	
Variables	*OR (95CI%)*	*p*
Model 1[†]		
PC1[¶]	1.086 (0.935, 1.260)	0.281
PC2[††]	3.657 (2.795, 4.784)	<0.001
PC3[‡‡]	2.341 (1.755, 3.123)	<0.001
Model 2[‡]		
PC1[¶]	0.855 (0.671, 1.090)	0.206
PC2[††]	2.239 (1.428, 3.508)	<0.001
PC3[‡‡]	2.302 (1.651, 3.208)	<0.001
Model 3[§]		
LDL, mmol/L	1.793 (0.918, 3.503)	0.087
HDL, mmol/L	0.001 (0.0003, 0.002)	<0.001
FFA, μmol/L	1.003 (1.001, 1.005)	0.003
TG, mmol/L	0.335 (0.146, 0.767)	0.010
PC1[¶]	1.795 (1.192, 2.703)	0.005
PC2[††]	8.421 (3.808, 18.62)	<0.001
PC3[‡‡]	2.974 (1.912, 4.626)	<0.001

LDL: Low density lipoprotein, HDL: High density lipoprotein, FFA: Free fatty acid, TG: Triglyceride.
[†]Model 1 No adjustments
[‡]Model 2 was adjusted for age, exercise, smoking, and drinking
[§]Model 3 was adjusted for age, exercise, smoking, drinking, BMI, metabolic syndrome, LDL, HDL, FFA, and TG
[¶]PC1: Principal component 1 associated with high animal-derived nutrients intake
[††]PC2: Principal component 2 linked to high carbohydrate intake
[‡‡]PC3: Principal component 3 related to low bean protein intake

Supplementary Table 9. The mediation effect of the dietary factors and body composition for colorectal cancer

Variables	Total effect	Direct effect	Mediated effect	Proportion mediated (%)	*p*
Men					
PC1†					
NAMA, cm^2	-0.144 (-0.211, -0.096)	-0.137 (-0.206, -0.091)	-0.007 (-0.018,0.002)	4.700	0.144
LAMA, cm^2	-0.144 (-0.216, -0.089)	-0.134 (-0.202, -0.089)	-0.010 (-0.025, 0.002)	6.700	0.090
SMI, cm^2/m^2	-0.144 (-0.207, -0.096)	-0.144 (-0.207, -0.096)	0.001 (-0.006, 0.007)	0.531	0.794
SMD, HU	-0.142 (-0.211, -0.096)	-0.128 (-0.199, -0.083)	-0.014 (-0.029, -0.003)	9.700	0.012
SAI, cm^2/m^2	-0.144 (-0.203, -0.094)	-0.143 (-0.202, -0.092)	-0.001 (-0.008, 0.004)	0.681	0.648
VAI, cm^2/m^2	-0.143 (-0.209, -0.092)	-0.143 (-0.210, -0.090)	0.0001 (-0.005, 0.004)	0.064	0.932
PC2‡					
NAMA, cm^2	-0.066 (-0.156, 0.016)	-0.060 (-0.149, 0.020)	-0.006 (-0.028, 0.009)	9.010	0.448
LAMA, cm^2	-0.060 (-0.151, 0.014)	-0.069 (-0.157, 0.005)	0.009 (-0.012, 0.032)	15.197	0.488
SMI, cm^2/m^2	-0.066 (-0.161, 0.009)	-0.064 (-0.159, 0.014)	-0.002 (-0.014, 0.005)	2.643	0.672
SMD, HU	-0065 (-0.154, 0.006)	-0.060 (-0.148, 0.007)	-0.005 (-0.031, 0.020)	7.797	0.728
SAI, cm^2/m^2	-0.067 (-0.164, 0.012)	-0.066 (-0.162, 0.015)	-0.001 (-0.011, 0.007)	1.805	0.832
VAI, cm^2/m^2SAI, cm^2/m^2	-0.066 (-0.161, 0.014)	-0.067 (-0.160, 0.017)	0.0002 (-0.007, 0.008)	0.362	0.996
PC3§					
NAMA, cm^2	0.005 (-0.087, 0.085)	-0.008 (-0.104, 0.074)	0.013 (-0.002, 0.041)	261.182	0.912
LAMA, cm^2	0.005 (-0.070, 0.090)	0.022 (-0.054, 0.102)	-0.017 (-0.042, 0.007)	323.611	0.920
SMI, cm^2/m^2	0.005 (-0.085, 0.089)	0.0003 (-0.092, 0.090)	0.004 (-0.008, 0.027)	92.802	0.978
SMD, HU	0.007 (0.081, 0.096)	0.011 (-0.069, 0.091)	-0.004 (-0.035, 0.023)	61.663	0.756
SAI, cm^2/m^2	0.005 (-0.088, 0.095)	0.005 (-0.087, 0.095)	-0.0003 (-0.014, 0.011)	5.850	0.926
VAI, cm^2/m^2	0.005 (-0.078, 0.092)	0.004 (-0.075, 0.091)	0.001 (-0.009, 0.009)	11.621	0.960

NAMA: Normal Attenuation Muscle Area, LAMA: Low Attenuation Muscle Area, SMI: Skeletal Muscle Index, SMD: Skeletal Muscle Density, SAI: Subcutaneous Adipose Index, VAI: Visceral adipose index.

†PC1: Principal component 1 associated with high animal-derived nutrients intake

‡PC2: Principal component 2 related to low bean protein intake and high carbohydrate intake

§PC3: Principal component 3 linked to high carbohydrate and bean protein intake

Supplementary Table 9. The mediation effect of the dietary factors and body composition for colorectal cancer (cont)

Variables	Total effect	Direct effect	Mediated effect	Proportion mediated (%)	*p*
Women					
PC1†					
NAMA, cm^2	-0.135 (-0.178, -0.099)	-0.127 (-0.171, -0.089)	-0.008 (-0.019, 0.0003)	6.048	0.032
LAMA, cm^2	-0.141 (-0.181, -0.100)	-0.127 (-0.166, -0.088)	-0.014 (-0.030, -0.0004)	9.782	0.038
SMI, cm^2/m^2	-0.137 (-0.177, -0.098)	-0.137 (-0.177, -0.099)	0.0001 (-0.003, 0.004)	0.097	0.842
SMD, HU	-0.131 (-0.179, -0.098)	-0.115 (-0.162, -0.079)	-0.016 (-0.031, -0.004)	11.934	0.004
SAI, cm^2/m^2	-0.136 (-0.175, -0.101)	-0.136 (-0.176, 0.102)	0.0005 (-0.006, 0.008)	0.339	0.864
VAI, cm^2/m^2	-0.140 (-0.178, -0.100)	-0.141 (-0.181, -0.102)	0.001 (-0.004, 0.009)	0.891	0.686
PC2‡					
NAMA, cm^2	0.130 (0.054, 0.211)	0.127 (0.054, 0.209)	0.002 (-0.026, 0.025)	1.834	0.906
LAMA, cm^2	0.127 (0.057, 0.211)	0.117 (0.049, 0.197)	0.010 (0.019, 0.042)	7.559	0.550
SMI, cm^2/m^2	0.127 (0.048, 0.209)	0.127 (0.047, 0.210)	-0.00002 (0.006, 0.007)	0.017	0.986
SMD, HU	0.131 (0.051, 0.211)	0.116 (0.039, 0.194)	0.015 (-0.021, 0.049)	11.536	0.382
SAI, cm^2/m^2	0.127 (0.049, 0.212)	0.126 (0.050, 0.212)	0.001 (-0.010, 0.011)	0.426	0.940
VAI, cm^2/m^2	0.127 (0.057, 0.212)	0.124 (0.052, 0.210)	0.003 (-0.004, 0.018)	2.746	0.462
PC3§					
NAMA, cm^2	-0.096 (-0.184, -0.024)	-0.085 (-0.171, -0.009)	-0.011 (-0.034, 0.006)	11.606	0.224
LAMA, cm^2	-0.097 (-0.186, -0.023)	-0.098 (-0.179, -0.029)	0.001 (-0.029, 0.032)	1.180	0.986
SMI, cm^2/m^2	-0.098 (-0.174, -0.028)	-0.098 (-0.175, -0.024)	0.0005 (-0.008, 0.009)	0.485	0.968
SMD, HU	-0.097 (-0.172, -0.021)	-0.091 (-0.167, -0.022)	-0.006 (-0.032, 0.024)	6.445	0.702
SAI, cm^2/m^2	-0.098 (-0.184, -0.020)	-0.098 (-0.184, -0.022)	0.0005 (-0.009, 0.015)	0.467	0.888
VAI, cm^2/m^2	-0.096 (-0.182, -0.023)	-0.103 (-0.192, -0.028)	0.007 (-0.004, 0.021)	6.995	0.274

NAMA: Normal Attenuation Muscle Area, LAMA: Low Attenuation Muscle Area, SMI: Skeletal Muscle Index, SMD: Skeletal Muscle Density, SAI: Subcutaneous Adipose Index, VAI: Visceral adipose index.

†PC1: Principal component 1 associated with high animal-derived nutrients intake

‡PC2: Principal component 2 related to low bean protein intake and high carbohydrate intake

§PC3: Principal component 3 linked to high carbohydrate and bean protein intake

Supplementary Table 10. The python code of the machine learning algorithms

```
#Import Lib
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score, accuracy_score, roc_curve

#Define function
def get_Xy(df, flag=True, is_all=False):
    list1 = ["NAMA", "LAMA", "SMD.HU.", "SAI", "VAI" ,"SMI"]#体成分
    list2 = ["Comp.1", "Comp.2", "Comp.3"]#主成分
    list3 = ['total.CHO', 'totalprote', 'total.fat', 'animalprot', 'bean.protein',
      'total.CH', 'meatfat3', 'meatCH3']#单独膳食
    list4 = ['age', 'smoking', 'drinking', 'exercise']#校正因素 remove'totalenergy'
    list5 = ["sex"]#all data need
    if flag:
        if is_all:
            X = df[list1+list2+list4+list5]
        else:
            X = df[list1+list2+list4]
    else:
        if is_all:
            X = X = df[list1+list3+list4+list5]
        else:
            X = df[list1+list3+list4]

    y = df[["zgroup"]].values.ravel()
    return X, y

def get_auc_acc(X, y, clf):
    y_pred_prob = clf.predict_proba(X)[:, 1]
    y_predict = clf.predict(X)
    auc_score = roc_auc_score(y, y_pred_prob)
    acc_score = accuracy_score(y, y_predict)
    return auc_score, acc_score

def custom_sort(order_list, target_list):
    # 创建一个字典，记录 order_list 中元素在 target_list 中的相对顺序
    order_dict = {elem: i for i, elem in enumerate(order_list)}
    # 使用 lambda 表达式进行排序，按照 order_dict 中的相对顺序对子表进行排序
    sorted_list = sorted(target_list, key=lambda x: order_dict[x])
    return sorted_list

def plot_roc(fpr, tpr, auc_score, title):
    xylabel_font_dict=dict(fontsize=20,
                           fontproperties='Times New Roman',
                       #weight='bold'
            )
```

Supplementary Table 10. The python code of the machine learning algorithms (cont.)

```
    title_font_dict=dict(fontsize=24,
                         fontproperties='Times New Roman',
                        #weight='bold'
                )
    plt.figure(figsize=(12,8))
    plt.plot(fpr, tpr, label='ROC curve (area= %0.3f)'% auc_score)
    plt.plot([0, 1], [0, 1], 'k--', label='Random Guess')
    plt.xlabel('False Positive Rate', fontdict=xylabel_font_dict)
    plt.ylabel('True Positive Rate', fontdict=xylabel_font_dict)
    plt.title(title, fontdict=title_font_dict)
    plt.yticks(fontproperties='Times New Roman', size=15,weight='bold')
    plt.xticks(fontproperties='Times New Roman', size=15,weight='bold')
    legend_font = {
    'family': 'Times New Roman',  # 字体
    'style': 'normal',
    'size': 15,  # 字号
    'weight': "normal",  # 是否加粗，不加粗
    }

    plt.legend(prop=legend_font)
    plt.savefig(title+str(".pdf"), format="pdf")
    plt.show()

def Get_trained_GSRF(X_train, y_train):
    rf_param_grid = {
        'n_estimators':[i for i in range(10, 150, 2)],
        'max_depth':[i for i in range(2, 8)],
        'min_samples_split': [i for i in range(2, 8)],          # 分裂一个节点需要的最小样本数
        'min_samples_leaf': [1, 2, 4],
    }

    gs = GridSearchCV(RandomForestClassifier(class_weight="balanced", random_state=1),
                      rf_param_grid, cv=10, verbose=1, n_jobs=-1, scoring='roc_auc')

    gs.fit(X_train, y_train) #Run fit with all sets of parameters.
    return gs
def train_and_draw(Name, X_train, y_train, X_test, y_test):
    gsrf = Get_trained_GSRF(X_train, y_train)
#     print('最优参数: ', gsrf.best_params_)
#     print('最佳 auc: ', gsrf.best_score_)
    bestrf = gsrf.best_estimator_
    do_featuresimp(bestrf, X_train, y_train, "%s feature importance" %(Name))
    X_train_train, X_val, y_train_train, y_val = train_test_split(X_train, y_train, test_size=0.2, random_state=1)

    bestrf.fit(X_train_train, y_train_train)
    train_auc, train_acc = get_auc_acc(X=X_val, y=y_val, clf=bestrf)
    print("%s train:\nauc=%.3f\nacc=%.3f" %(Name, train_auc, train_acc))
    draw_roc(X_val, y_val, bestrf, train_auc, "%s Train_set" %(Name) )

    bestrf.fit(X_train, y_train)
    test_auc, test_acc = get_auc_acc(X=X_test, y=y_test, clf=bestrf)
    print("%s test:\nauc=%.3f\nacc=%.3f" %(Name, test_auc, test_acc))
    draw_roc(X_test, y_test, bestrf, test_auc, "%s Test_set" %(Name) )

def do_featuresimp(clf, X_train, y_train, title):
    feature_importances = clf.feature_importances_
    # 创建特征名列表
    feature_names = clf.feature_names_in_
    # 创建一个 DataFrame，包含特征名和其重要性得分
    feature_importances_df = pd.DataFrame({'feature': feature_names, 'importance': feature_importances})
    # 对特征重要性得分进行排序
    feature_importances_df = feature_importances_df.sort_values('importance', ascending=False)
    draw_featureimp(feature_importances, X_train, title)
```

Supplementary Table 10. The python code of the machine learning algorithms (cont.)

```
def train_order_features(features_import_lt, X_train, y_train, clf):
    #保持 order train
    for i in range(len(features_import_lt)):
        #print(features_import_lt[:i+1])
        order_features = custom_sort(order_list=list(X_train.columns), target_list=features_import_lt[:i+1])
        select_x = X_train[order_features]
        #print(select_x)
        cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=1)
        cv_acc= cross_val_score(estimator=clf, X=select_x, y=y_train, scoring='accuracy', cv=10)
        cv_auc = get_cv_auc(cf=clf, X=select_x, y=y_train, cv=cv)
        print("n_featres:{} 训练集 10-fold 交叉验证 acc:{:.4f} auc:{:.4f}".format(i+1, cv_acc.mean(),
cv_auc.mean()))

## 计算特征重要性排序
def draw_featureimp(feature_importances, X_train, title):
    import seaborn as sns
    sns.set(font='SimHei',font_scale=1.5)
    indices = np.argsort(feature_importances)[::-1]
    print("Feature ranking")
    featurerank=[]
    for f in range(X_train.shape[1]):
        featurerank.append(X_train.columns[indices[f]])
        print("%d. %s (%f)" % (f +1,X_train.columns[indices[f]],feature_importances[indices[f]]))

    plt.figure(figsize=(20,16))
    feature_imp=pd.Series(feature_importances,index=X_train.columns).sort_values(ascending=False)
    sns.barplot(x=feature_imp,y=feature_imp.index)
    xylabel_font_dict=dict(fontsize=32,
                           fontproperties='Times New Roman',
                           weight='bold'
                )
    title_font_dict=dict(fontsize=40,
                         fontproperties='Times New Roman',
                         weight='bold'
                )
    plt.xlabel('Feature importance score', fontdict=xylabel_font_dict)
    plt.ylabel('Features', fontdict=xylabel_font_dict)
    plt.title(title, fontdict=title_font_dict)
    plt.yticks(fontproperties='Times New Roman', size=15,weight='bold')
    plt.xticks(fontproperties='Times New Roman', size=15,weight='bold')
    plt.savefig(title+str(".pdf"), format="pdf")
    plt.show()

def draw_roc(X_data, y, clf, auc_score, title):
    y_pred_prob = clf.predict_proba(X_data)[:, 1]
    fpr, tpr, thresholds = roc_curve(y, y_pred_prob)
    plot_roc(fpr, tpr, auc_score, "%s ROC Curve" %(title))

#Load Data
female_data = pd.read_excel("data/CRCPCAF.xlsx", sheet_name="dpF")#(153, 38)
male_data = pd.read_excel("data/CRCPCAM.xlsx", sheet_name="dpM") #(154, 38)
all_data = pd.read_excel("data/CRCPCA.xlsx")#(307, 38)
all_data.rename(columns={"WC": "Wcfefine"}, inplace=True)

#V1 data get[X y](Female, Male, All)
#female
X_female1, y_female = get_Xy(df=female_data, flag=True, is_all=False)
#male
X_male1, y_male = get_Xy(df=male_data, flag=True, is_all=False)
#all_data
X1, y = get_Xy(df=all_data, flag=True, is_all=True)
```

Supplementary Table 10. The python code of the machine learning algorithms (cont.)

```
#V1 train_test_split (train:test = 7:3)
#train:test = 7:3
X_female_train1, X_female_test1, y_female_train, y_female_test = train_test_split(X_female1, y_female,
test_size=0.3, random_state=42)
X_male_train1, X_male_test1, y_male_train, y_male_test = train_test_split(X_male1, y_male, test_size=0.3,
random_state=42)
X_train1, X_test1, y_train, y_test = train_test_split(X1, y, test_size=0.3, random_state=42)


train_and_draw("Famale V1-noenergy", X_female_train1, y_female_train, X_female_test1, y_female_test)
train_and_draw("Male V1-noenergy", X_male_train1, y_male_train, X_male_test1, y_male_test)
# train_and_draw("All V1", X_train1, y_train, X_test1, y_test)
```