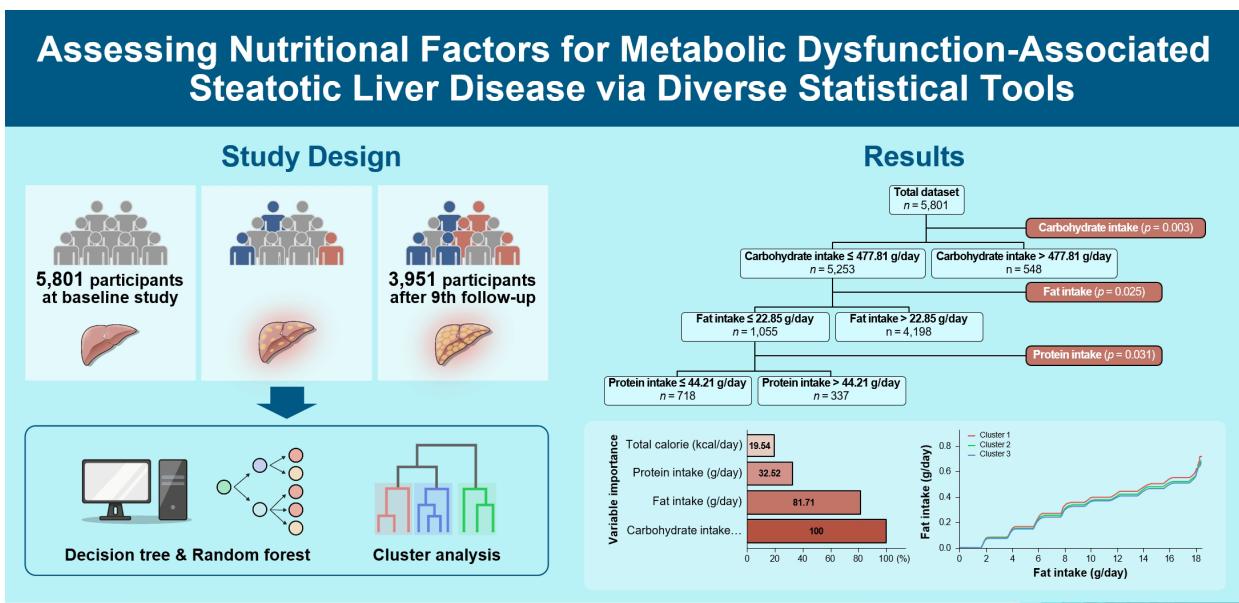


Assessing Nutritional Factors for Metabolic Dysfunction-Associated Steatotic Liver Disease via Diverse Statistical Tools

Yea-Chan Lee, Hye Sun Lee, Soyoung Jeon, Yae-Ji Lee, Yu-Jin Kwon, Ji-Won Lee

Diabetes Metab J 2026;50:178-189 | <https://doi.org/10.4093/dmj.2025.0026>



Conclusion Comprehensive analyses using decision tree, random forest, and clustering methods identified carbohydrate intake as the key dietary driver of MASLD, underscoring the pathogenic significance of macronutrient composition.



Highlights

- Nutritional strategies are essential for managing MASLD.
- The optimal macronutrient composition for MASLD remains unclear.
- Carbohydrate intake plays a key role in the development of MASLD.
- Optimizing macronutrients may be more effective than reducing energy intake alone.

How to cite this article:

Lee YC, Lee HS, Jeon S, Lee YJ, Kwon YJ, Lee JW. Assessing Nutritional Factors for Metabolic Dysfunction-Associated Steatotic Liver Disease via Diverse Statistical Tools. *Diabetes Metab J* 2026;50:178-189. <https://doi.org/10.4093/dmj.2025.0026>

Original Article

Lifestyle and Behavioral Interventions

Diabetes Metab J 2026;50:178-189

<https://doi.org/10.4093/dmj.2025.0026>

pISSN 2233-6079 · eISSN 2233-6087



Assessing Nutritional Factors for Metabolic Dysfunction-Associated Steatotic Liver Disease via Diverse Statistical Tools

Yea-Chan Lee^{1,*}, Hye Sun Lee^{2,*}, Soyoung Jeon², Yae-Ji Lee³, Yu-Jin Kwon⁴, Ji-Won Lee^{1,5}

¹Department of Family Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

²Biostatistics Collaboration Unit, Department of Research Affairs, Yonsei University College of Medicine, Seoul, Korea

³Department of Biostatistics and Computing, Yonsei University, Seoul, Korea

⁴Department of Family Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Korea

⁵Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea

Background: Lifestyle modifications are critical in addressing metabolic dysfunction-associated steatotic liver disease (MASLD); however, the specific macronutrients that most significantly influence the disease's progression are uncertain. In this study, we aimed to explore the role of carbohydrate, fat, and protein intake in MASLD development using decision trees, random forest models, and cluster analysis.

Methods: Participants ($n=3,951$) from the Korean Genome and Epidemiology Study were included. We used the classification and regression tree analysis to classify participants into subgroups based on variables associated with the incidence of new-onset MASLD. Random forest analyses were used to assess the relative importance of each variable. Participants were grouped into homogeneous clusters based on carbohydrate, protein, fat, and total caloric intake using hierarchical cluster analysis. Subsequently, we used the Cox proportional hazards regression models to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for MASLD risk across the clusters.

Results: Carbohydrate intake was identified as the most significant predictor of new-onset MASLD, followed by fat, protein, and total caloric intake. Participants in cluster 3, who consumed a lower proportion of carbohydrate but had higher total caloric, protein, and fat intake, had a lower risk of new-onset MASLD than those in cluster 1 after adjusting for confounders (cluster 1 as a reference; cluster 3: HR, 0.90; 95% CI, 0.82 to 0.99).

Conclusion: The study's results highlight the critical role of macronutrient composition, particularly carbohydrate intake, in MASLD development. The findings suggest that dietary strategies focusing on optimizing macronutrients, rather than simply reducing caloric intake, may be more effective in preventing MASLD.

Keywords: Dietary carbohydrates; Fatty liver; Nutrients

INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a recently proposed term replacing non-alcoholic fatty liver

disease (NAFLD). It emphasizes the metabolic dysfunction underlying the condition [1]. Furthermore, it is emerging as a major global health concern because of the increasing prevalence of atherosclerotic cardiovascular diseases [2], potential

Corresponding authors: Ji-Won Lee <https://orcid.org/0000-0002-2666-4249>
Department of Family Medicine, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
E-mail: IND15645@yuhs.ac

Yu-Jin Kwon <https://orcid.org/0000-0002-9021-3856>
Department of Family Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Gicheung-gu, Yongin 16995, Korea
E-mail: DIGDA3@yuhs.ac

*Yea-Chan Lee and Hye Sun Lee contributed equally to this study as first authors.

Received: Jan. 9, 2025; Accepted: Mar. 19, 2025

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

for severe liver diseases such as cirrhosis and hepatocellular carcinoma [3,4], and its contribution to significant morbidity and mortality [3,5]. Notably, obesity and metabolic syndrome significantly influence the development and progression of MASLD, with global prevalence estimates ranging from 20% to 25% [6]. Recent data from the Global Burden of Disease study indicate that MASLD affects approximately 25% to 38% of the global adult population, making it one of the most prevalent chronic liver diseases worldwide [7]. Furthermore, the annual prevalence of MASLD in South Korea rose from 15.69 per 1,000 population in 2010 to 34.23 per 1,000 in 2021, which was closely parallel to the increase in obesity and metabolic syndrome [8]. However, few targeted therapeutic strategies are available for managing this disease despite its widespread prevalence and substantial public health impact.

Notably, dietary interventions are crucial in managing MASLD. This involves addressing hepatic fat accumulation and metabolic dysfunction [9,10]. Furthermore, excessive calorie intake, particularly from unhealthy sources, promotes obesity and insulin resistance, influencing MASLD development [11]. High-calorie diets further damage the liver by increasing fat accumulation, inflammation, and oxidative stress [11,12]. Conversely, calorie restriction and modest weight loss (5% to 10%) reduce hepatic steatosis, improve liver function, and alleviate metabolic conditions such as type 2 diabetes mellitus (T2DM), dyslipidemia, and hypertension [12,13].

Additionally, adjusting macronutrient intake along with calorie control is important for preventing and managing MASLD. A high-carbohydrate diet increases MASLD risk, as shown in the National Health and Nutrition Examination Survey (NHANES) study, where patients with MASLD who have advanced fibrosis had higher carbohydrate intake [14]. Conversely, reducing free sugars decreased hepatic steatosis from 25% to 17% in adolescents with NAFLD [15]. In addition, fat intake worsens insulin resistance and inflammation, driving MASLD progression [16]. The NHANES study also reported higher saturated fat intake in patients with MASLD [16], whereas a 3-month low-fat diet led to over 9% weight loss and improved hepatic steatosis in Serbian men who were overweight [17]. Additionally, a Mediterranean diet rich in unsaturated fats reduced liver steatosis by 25% to 35% in patients with NAFLD [18]. The role of protein intake in MASLD remains unclear; however, it is crucial for preventing sarcopenia [19].

Notably, it is not well understood which macronutrient has the most profound impact on MASLD development. Addi-

tionally, most research performed, focused on Western populations, leaving a gap in understanding how dietary factors influence MASLD in Asian populations. Considering the differences in dietary patterns and metabolic responses across populations, identifying the specific macronutrients that most strongly affect MASLD risk in diverse populations is essential to creating more effective prevention and management strategies.

To address this gap, we use decision trees and random forest models to investigate the impact of carbohydrate, fat, and protein intake on new-onset MASLD in a large Korean cohort. In addition, we conducted a cluster analysis to explore the relationship between macronutrient proportions and MASLD.

METHODS

Study population

We used datasets obtained from the Ansung and Ansan cohort of the Korean Genome and Epidemiology Study (KoGES). These cohorts included community-dwelling individuals, comprising men and women who were aged ≥ 40 years at the time of enrollment. The baseline data were collected between 2001 and 2002. Furthermore, the 1st to 9th follow-up data were obtained between 2003 and 2020. The data collection in KoGES was meticulously standardized, with regular follow-up surveys and health examinations conducted by trained medical personnel.

Of the 10,030 participants, with ages ranging from 40 to 69 years, those who had no data on NAFLD liver fat score ($n=299$), metabolic syndrome criteria ($n=620$), or macronutrient intake ($n=31$) were excluded. Additionally, we excluded participants who had previous MASLD status at baseline ($n=2,089$) and were never followed up from 1st to 9th in the follow-up studies ($n=1,102$), and those who had total calorie intake <500 or $>5,000$ kcal/day ($n=88$). Overall, 5,801 participants were included in the baseline study. Furthermore, we included the participants in the study population every 2 years during the follow-up period. Finally, 3,951 participants were included in our study population after the 9th follow-up assessment (Fig. 1).

Definition of metabolic dysfunction-associated steatotic liver disease

We defined MASLD as the presence of hepatic steatosis accompanied by metabolic dysfunction [1]. Furthermore, hepatic steatosis was defined as a NAFLD liver fat score >-0.64 , which was calculated as liver fat score = $-2.89 + 1.18 \times$ metabolic syn-

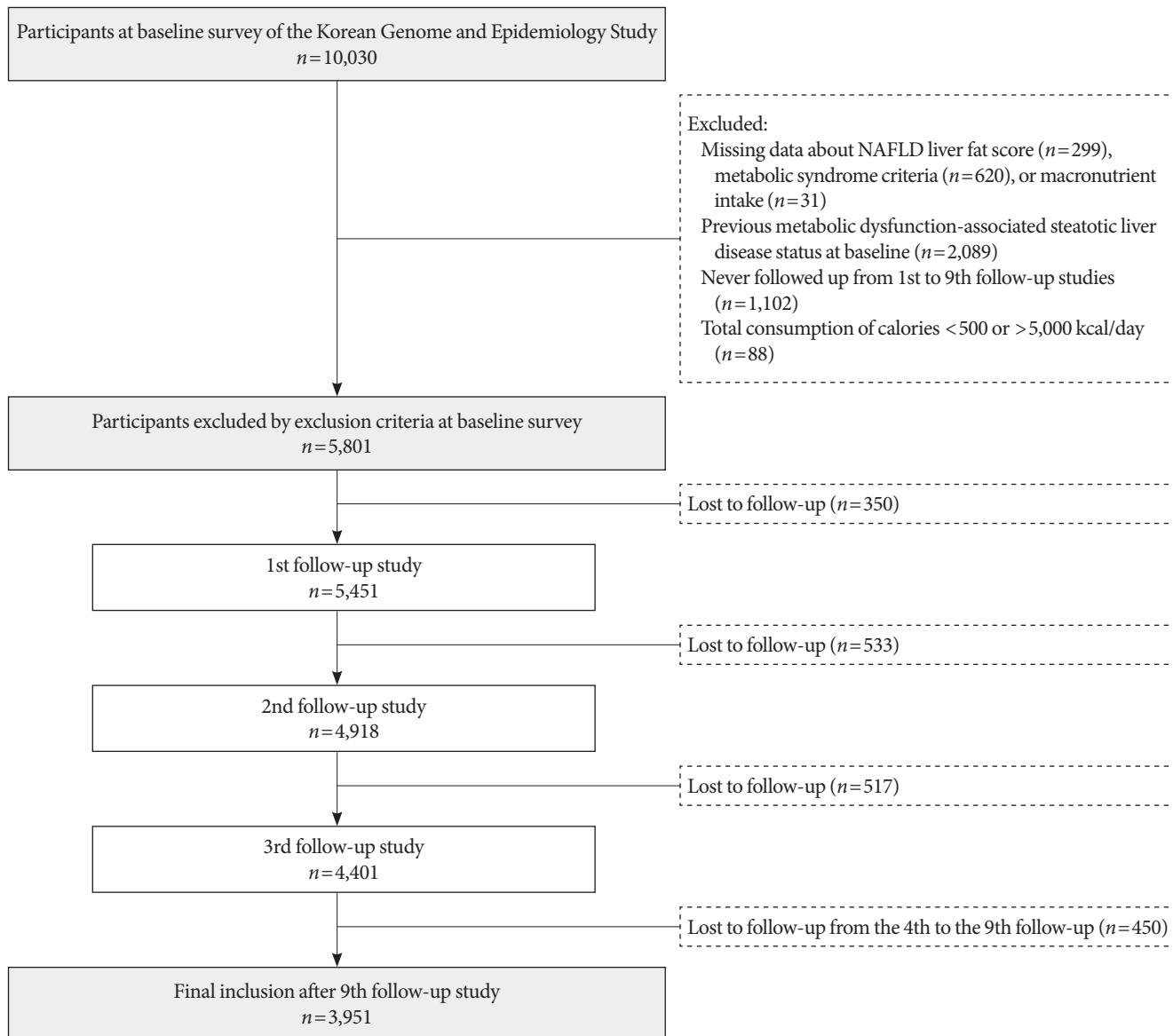


Fig. 1. Flow chart of study population selection. NAFLD, non-alcoholic fatty liver disease.

drome (yes=1, no=0)+0.9×diabetes mellitus (yes=1, no=0)+0.15×serum insulin (μIU/mL)+0.04×serum aspartate aminotransferase (AST; U/L)–0.94×AST/alanine aminotransferase [20]. Metabolic dysfunction was indicated as meeting at least one of the following cardiometabolic risk factors: (1) body mass index ≥ 23 kg/m² or waist circumferences (WC) ≥ 94 cm in men and ≥ 80 cm in women; (2) fasting plasma glucose level ≥ 100 mg/dL, 2-hour postprandial glucose levels ≥ 140 mg/dL, glycosylated hemoglobin $\geq 5.7\%$, or receiving treatment for T2DM; (3) systolic blood pressure ≥ 130 mm Hg, diastolic blood pressure (DBP) ≥ 85 mm Hg, or receiving specific antihyper-

tensive drug treatment; (4) serum triglyceride (TG) level ≥ 150 mg/dL or receiving lipid-lowering treatment; and (5) serum high-density lipoprotein cholesterol level <40 mg/dL in men and <50 mg/dL in women [1].

Assessment of nutrition intake

Dietary intake was assessed using a semi-quantitative food frequency questionnaire (FFQ) consisting of 103 food items [21]. The FFQ was designed to estimate usual dietary intake by collecting information on the frequency and portion size of food consumption. To calculate total energy and nutrient intake, the

reported frequency of consumption for each food item was multiplied by the corresponding portion size and nutrient content, as specified in the Korean Food Composition Table (7th edition, The Korean Nutrition Society, 2000) [22]. The total daily nutrient intake for each participant was derived by summing the nutrient intakes from all reported food items. This process was conducted using the DS24 software, developed by the Human Nutrition Lab at Seoul National University in collaboration with the AI/DB Lab at Sookmyung Women's University (1996) [21]. Detailed information regarding the covariates is described in the Supplementary Methods.

Statistical analysis

We used the chi-square test to analyze categorical variables and the independent two-sample *t*-test for continuous variables to compare between the new-onset MASLD group and the not-developed MASLD group. Classification and regression tree (CART) analysis was used to divide subjects into subgroups based on optimal cut-off points for variables. The optimal cut-off points were determined based on the point at which the log-rank test statistic was maximized. Furthermore, a random forest model was used to assess the importance of each variable in the incidence of the new-onset MASLD. Additionally, hierarchical cluster analysis was used to classify the variables, including carbohydrate, protein, fat, and total calorie intake, and Ward's minimum variance method was used as the clustering criterion. All variables were standardized before clustering and two statistical values were estimated to determine the optimal number (*n*) of clusters: the smallest pseudo *t*² statistic and the largest decrease in semipartial *R*² at *n* clusters compared with *n*-1 clusters. A one-way analysis of variance and chi-square test were used to compare the clusters, for continuous and categorical variables, respectively. The cumulative incidence curve showed the incidence rate of new-onset MASLD in each cluster. Furthermore, the Cox proportional hazard regression model was performed to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) of MASLD for clusters. All statistical analyses were performed using the R package version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and SAS version 9.4 (SAS Institute, Cary, NC, USA). Statistical significance was set at *P*<0.05.

Ethics and consent

Ethical standards were rigorously upheld, with informed consent obtained from all participants. The study protocols were

conducted following the guidelines of the Declaration of Helsinki. The study was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine (IRB No: 4-2024-0866).

RESULTS

Demographic characteristics of the study population

Table 1 shows the demographic data of the participants based on the incidence of new-onset MASLD during the median follow-up for approximately 11.6 years (interquartile range [IQR], 5.7 to 17.6). Notably, 2,869 and 2,932 participants were included in the new-onset and not-developed MASLD groups, respectively. The mean ages were 51.4 ± 8.4 and 51.0 ± 9.0 years in new-onset and not-developed MASLD groups, respectively. Furthermore, participants in the not-developed MASLD group had lower rates of obesity and smaller WC than those in the new-onset MASLD group. Moreover, the prevalence of metabolic syndrome and other chronic diseases, such as diabetes, hypertension, and dyslipidemia were significantly lower in the not-developed MASLD group than in the new-onset MASLD. Notably, the not-developed MASLD group had a lower carbohydrate intake than the new-onset MASLD group (345.90 ± 97.44 g/day vs. 351.10 ± 103.74 g/day, *P*=0.049).

Classification and regression tree analysis and random forest model

Fig. 2A illustrates the results from the CART algorithm, which analyzed the incidence of new-onset MASLD. The root node was initially split based on carbohydrate intake (*P*=0.003), with 548 individuals consuming over 477.81 g/day (node 1) and 5,253 consuming 477.81 g/day or less (node 1'). Node 1' was further divided by fat intake at a cut-off of 22.85 g/day (*P*=0.025), resulting in node 2 and node 2'. Of the 4,198 individuals in node 2' (fat intake ≤ 22.85 g/day), two more child nodes were created based on protein intake: node 3, with 337 individuals consuming over 44.21 g/day of protein, and node 4, with 718 individuals consuming 44.21 g/day or less (*P*=0.031). No further significant variables were found to split the subgroups; hence, in total, four nodes, including nodes 1, 2, 3, and 4, were used.

The variable importance of the selected splitters, as analyzed by the random forest model, is shown in Fig. 2B. The analysis for predicting new-onset MASLD identified carbohydrates, fat, protein, and total calorie intake as the most relevant variables.

Table 1. Demographic characteristics of study population

Characteristic	Total (n=5,801)	New-onset MASLD (n=2,869)	Not-developed MASLD (n=2,932)	P value
Age, yr	51.2±8.7	51.4±8.4	51.0±9.0	0.737
Male sex	2,714 (46.8)	1,353 (47.2)	1,361 (46.4)	0.572
Body weight, kg	61.3±9.3	63.7±9.1	58.9±8.8	<0.001
Height, cm	160.0±8.5	160.0±8.6	160.1±8.4	0.577
BMI, kg/m ²	23.9±2.8	24.9±2.7	22.9±2.6	<0.001
WC, cm	80.2±8.0	83.0±7.3	77.5±7.7	<0.001
SBP, mm Hg	117.8±17.1	120.8±17.0	114.9±16.6	<0.001
DBP, mm Hg	78.3±10.9	80.4±10.8	76.3±10.7	<0.001
Total cholesterol, mg/dL	188.8±34.0	192.9±35.0	184.9±32.5	<0.001
Triglyceride, mg/dL	138.1±74.5	155.1±84.8	121.4±58.3	<0.001
HDL-C, mg/dL	46.3±10.1	44.4±9.2	48.1±10.5	<0.001
LDL-C, mg/dL	115.0±31.6	117.5±33.0	112.4±29.9	<0.001
FPG, mg/dL	83.4±13.4	85.4±16.4	81.4±9.3	<0.001
HbA1c, %	5.56±0.53	5.65±0.66	5.47±0.35	<0.001
Insulin, μIU/mL	6.49±2.79	6.76±2.89	6.22±2.66	<0.001
AST, U/L	26.9±7.3	27.3±7.5	26.5±7.1	<0.001
ALT, U/L	22.8±9.3	24.6±10.0	21.1±8.2	<0.001
Current smoking (yes)	1,431 (24.8)	725 (25.4)	706 (24.3)	0.305
Alcohol intake, g/day	8.6±19.4	9.2±19.5	8.0±19.2	0.015
Physical activity				0.020
Low	407 (7.21)	179 (6.43)	228 (7.97)	
Moderate	3,524 (62.40)	1,724 (61.90)	1,800 (62.89)	
High	1,716 (30.39)	882 (31.67)	834 (29.14)	
Obesity (yes)	1,913 (33.2)	1,312 (46.1)	601 (20.6)	<0.001
Abdominal obesity (yes)	1,024 (17.7)	700 (24.4)	324 (11.1)	<0.0001
Metabolic syndrome (yes)	599 (10.3)	436 (15.2)	163 (5.6)	<0.001
Diabetes mellitus (yes)	194 (3.3)	150 (5.2)	44 (1.5)	<0.001
Hypertension (yes)	1,269 (21.9)	797 (27.8)	472 (16.1)	<0.001
Dyslipidemia (yes)	2,250 (38.8)	1,377 (48.0)	873 (29.8)	<0.001
Total calorie, kcal/day	2,092.3±653.1	2,104.1±667.9	2,080.8±638.2	0.174
Carbohydrate, g/day	348.47±100.63	351.10±103.74	345.90±97.44	0.049
Carbohydrate, %	67.49±7.50	67.59±7.51	67.39±7.50	0.299
Protein, g/day	70.67±28.31	71.12±28.61	70.24±28.02	0.241
Protein, %	13.32±2.31	13.33±2.29	13.30±2.33	0.614
Fat, g/day	43.43±24.02	43.41±24.15	43.44±23.90	0.958
Fat, %	17.91±6.10	17.82±6.11	18.01±6.10	0.235

Values are presented as mean±standard deviation or number (%).

MASLD, metabolic dysfunction-associated steatotic liver disease; BMI, body mass index; WC, waist circumference; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; FPG, fasting plasma glucose; HbA1c, glycosylated hemoglobin; AST, aspartate aminotransferase; ALT, alanine aminotransferase.

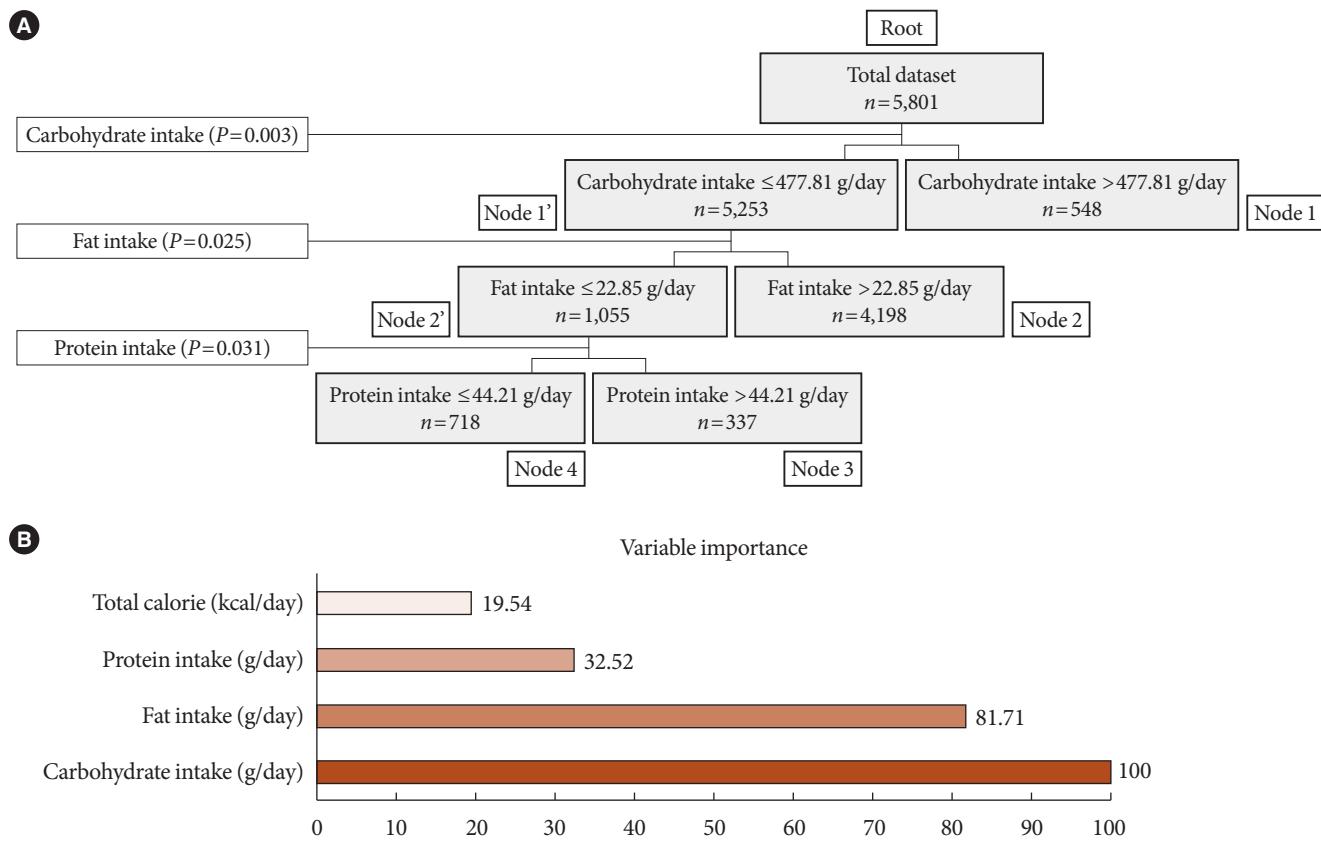


Fig. 2. Classification and decision tree presenting the splitting variables and comparison of the variable importance according to the incidence of new-onset metabolic dysfunction-associated steatotic liver disease. (A) Decision tree algorithm. (B) Random forest model.

Among them, carbohydrate intake had the highest relative importance, set at 100, followed by fat intake at 81.71, protein intake at 32.52, and total calories at 19.54.

Cluster analysis and Cox proportional hazard regression model

We classified the participants based on four variables using a hierarchical cluster analysis: the proportions of carbohydrate, protein, fat, and total calorie intake. Supplementary Table 1 shows the determination of the optimal number of clusters. We found that the pseudo t^2 statistic value of 545.8 was the smallest when the number of clusters was three, and the decrease in semipartial R^2 was largest at three clusters (0.050). Results from the hierarchical cluster analysis identified three clusters as the best number to represent the data provided for 0.532 R^2 values (Supplementary Fig. 1).

Table 2 shows the results of cluster analysis for demographic variables based on the clusters. Approximately 1,568 partici-

pants were included in cluster 1, 2,414 in cluster 2, and 1,819 in cluster 3. Cluster 3 participants had lower WC, higher high-density lipoprotein levels, and lower prevalence of abdominal obesity, metabolic syndrome, hypertension, and dyslipidemia than those in other clusters. Their carbohydrate consumption was relatively low despite their higher total caloric intake. Notably, cluster 3 participants also consumed 1.30 ± 0.36 g/kcal of dietary fiber, which was more than the 1.14 ± 0.31 g/kcal in cluster 1 or 1.23 ± 0.32 g/kcal in cluster 2 ($P < 0.001$). Additionally, this group consumed more proteins and fats, with a notably lower n-6 to n-3 polyunsaturated fatty acid ratio.

Fig. 3 presents the Kaplan–Meier analysis for the cumulative incidence of new-onset MASLD. Over a median 11.6-year follow-up period (IQR, 5.7 to 17.6), the cumulative incidence of new-onset MASLD was highest in cluster 1, followed by cluster 2, and lowest in cluster 3. The calculated cumulative incidence values were 0.68 for cluster 1, 0.65 for cluster 2, and 0.64 for cluster 3. However, this difference was not statistically signifi-

Table 2. Baseline characteristics of clusters with dataset defined by the Ward's minimum variance method

Characteristic	Cluster 1 (n=1,568)	Cluster 2 (n=2,414)	Cluster 3 (n=1,819)	P value	P value between each pair of clusters		
					1 vs. 2	1 vs. 3	2 vs. 3
Demographic variable							
Age, yr	54.4±9.0	50.9±8.7	49.0±7.8	<0.001	a	a	a
Male sex	597 (38.1)	1,200 (49.7)	917 (50.4)	<0.001	a	a	ns
Body weight, kg	59.3±9.2	61.9±9.0	62.3±9.4	<0.001	a	a	ns
Height, cm	157.7±8.5	160.7±8.3	161.3±8.4	<0.001	a	a	a
BMI, kg/m ²	23.8±3.0	23.9±2.8	23.9±2.7	0.412	ns	ns	ns
WC, cm	80.8±8.3	80.2±7.9	79.7±7.8	0.001	a	a	a
SBP, mm Hg	120.9±17.8	117.1±16.8	116.0±16.5	<0.001	a	a	ns
DBP, mm Hg	79.5±10.8	78.0±10.9	77.8±11.0	<0.001	a	a	ns
Total cholesterol, mg/dL	186.8±32.9	188.4±35.0	191.1±33.6	<0.001	ns	a	a
Triglyceride, mg/dL	140.6±71.6	138.9±77.0	134.9±73.7	0.065	ns	ns	ns
HDL-C, mg/dL	45.6±10.0	46.0±10.0	47.1±10.1	<0.001	ns	a	a
LDL-C, mg/dL	113.1±30.8	114.6±32.5	117.0±30.9	0.001	ns	a	a
FPG, mg/dL	82.7±12.3	83.1±13.6	84.4±14.1	0.001	ns	a	a
HbA1c, %	5.58±0.54	5.55±0.48	5.56±0.59	0.256	ns	ns	ns
Insulin, µIU/mL	6.45±2.70	6.51±2.83	6.49±2.80	0.757	ns	ns	ns
AST, U/L	27.1±7.2	27.0±7.5	26.6±7.3	0.131	ns	ns	ns
ALT, U/L	22.1±8.8	23.1±9.6	23.1±9.2	0.004	a	a	ns
Current smoking (yes)	335 (21.6)	604 (25.2)	492 (27.2)	0.001	a	a	ns
Current drinking (yes)	596 (38.2)	1,184 (49.2)	1,050 (58.0)	<0.001	a	a	a
Physical activity							
Low	135 (8.91)	165 (6.99)	107 (6.04)				
Moderate	759 (50.10)	1,532 (64.89)	1,233 (69.62)				
High	621 (40.99)	664 (28.12)	431 (24.34)				
Obesity (yes)	508 (32.7)	823 (34.3)	582 (32.2)	0.344	ns	ns	ns
Abdominal obesity (yes)	352 (22.5)	407 (16.9)	265 (14.6)	<0.0001	a	a	a
Metabolic syndrome (yes)	240 (15.31)	237 (9.82)	122 (6.71)	<0.001	a	a	a
Diabetes mellitus (yes)	58 (3.7)	69 (2.9)	67 (3.7)	0.221	ns	ns	ns
Hypertension (yes)	441 (28.1)	485 (20.1)	343 (18.9)	<0.001	a	a	ns
Dyslipidemia (yes)	621 (39.60)	974 (40.35)	655 (36.01)	0.012	ns	a	a
Selected variable							
Total calorie, kcal/day	1,646.6±320.0	2,126.1±589.2	2,431.6±723.7	<0.001	a	a	a
Carbohydrate, g	310.9±59.5	364.9±109.5	359.1±108.0	<0.001	a	a	ns
Carbohydrate, %	75.72±3.81	68.43±3.28	59.14±4.83	<0.001	a	a	a
Protein, g	46.6±12.2	69.0±19.5	93.7±29.9	<0.0001	a	a	a
Protein, %	11.24±1.51	13.05±1.57	15.47±1.83	<0.001	a	a	a
Fat, g	20.8±7.6	40.7±12.4	66.5±24.5	<0.0001	a	a	a
Fat, %	11.16±3.12	17.34±2.83	24.49±4.10	<0.001	a	a	a
Nutrient details per calorie, g/kcal							
Dietary fiber	1.14±0.31	1.23±0.32	1.30±0.36	<0.001	a	a	a
Saturated fatty acid	1.07±0.85	1.40±0.87	1.50±0.80	<0.001	a	a	a
MUFA	0.48±0.18	0.64±0.21	0.80±0.27	<0.001	a	a	a
PUFA	0.64±0.24	0.81±0.27	0.93±0.31	<0.001	a	a	a
n-3 PUFA	0.05±0.03	0.06±0.03	0.07±0.04	<0.001	a	a	a
n-6 PUFA	0.41±0.17	0.47±0.20	0.48±0.24	<0.001	a	a	a
n-6 to n-3 ratio	9.60±3.58	8.38±3.21	7.18±2.63	<0.001	a	a	a

Values are presented as mean±standard deviation or number (%).

ns, not significant; BMI, body mass index; WC, waist circumference; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; FPG, fasting plasma glucose; HbA1c, glycosylated hemoglobin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acid.

^aP value of less than 0.05.

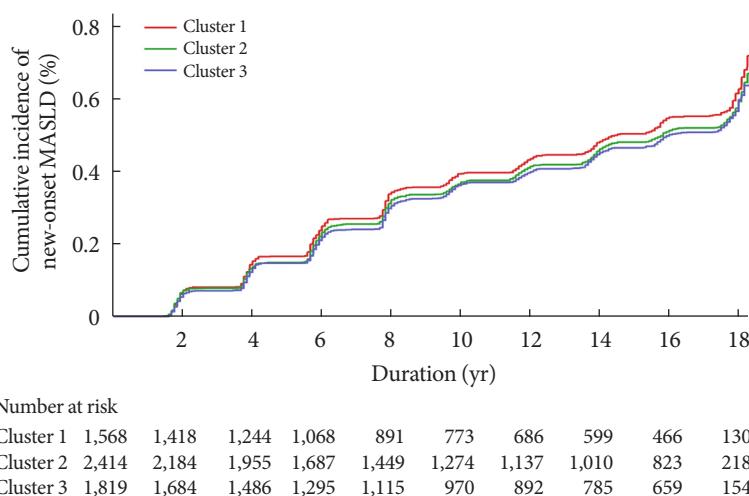


Fig. 3. Kaplan-Meier analysis for the cumulative incidence of new-onset metabolic dysfunction-associated steatotic liver disease (MASLD) (above) and number at risk table (below) (log-rank test, $P=0.113$).

Table 3. Hazard ratios and 95% confidence intervals for the presence of new-onset MASLD according to clusters

	Cluster 1 ($n=1,568$)	Cluster 2 ($n=2,414$)	Cluster 3 ($n=1,819$)
Unadjusted model	Ref	0.92 (0.84–1.01)	0.89 (0.81–0.98)
Adjusted model	Ref	0.93 (0.85–1.02)	0.90 (0.82–0.99)

Adjusted variables: age, sex, current smoking, alcohol intake (g/day), and physical activity.

MASLD, metabolic dysfunction-associated steatotic liver disease.

cant (log-rank test, $P=0.113$).

Moreover, Table 3 presents the HR and 95% CI of clusters for the incidence of new-onset MASLD. The HR and 95% CI for the incidence of new-onset MASLD in cluster 3 was 0.89 (95% CI, 0.81 to 0.98) compared with that in cluster 1 in the unadjusted model. However, the HR and 95% CI for the presence of new-onset MASLD in cluster 3 was 0.90 (95% CI, 0.82 to 0.99) compared with that in cluster 1, after adjusting for age, sex, current smoking and alcohol drinking, and physical activity.

DISCUSSION

In this study, we investigated the relationship between macronutrient intake and the incidence of new-onset MASLD in a large Korean cohort using CART and random forest models. From the analysis, we identified carbohydrate, fat, and protein intake as key contributors to MASLD risk, with carbohydrate intake having the greatest influence. Notably, hierarchical clus-

ter analysis revealed distinct dietary patterns across the three clusters. Participants in cluster 3, who had lower carbohydrate intake and higher intake of healthy fats and proteins, had a 10% reduced risk of developing MASLD compared with those in cluster 1, even after adjusting for confounders.

A high-carbohydrate diet promotes *de novo* lipogenesis, increasing triglyceride production and hepatic fat accumulation, hence contributing to hepatic steatosis [23,24]. Consistent with our findings, several studies emphasize the critical role of carbohydrate intake in managing liver diseases, beyond simply reducing calories [25–27]. Afsharfar et al. [25] demonstrated that total carbohydrate intake is significantly associated with the risk of hepatic steatosis in NAFLD patients. Additionally, findings from a 2-week randomized clinical trial revealed that a low-carbohydrate diet reduced hepatic steatosis by 27%, surpassing the effect of a low-calorie diet in participants with NAFLD [26]. However, a recent meta-analysis of 34 observational studies revealed that total carbohydrate intake was not significantly associated with MASLD, suggesting that different types and sources of carbohydrates may have varying effects on liver fat accumulation [28]. Specifically, fructose has been associated with a higher incidence of MASLD, whereas fiber and whole-grain starch may have protective effects [28,29]. This was observed in our study where cluster 3 participants with a lower risk of MASLD consumed fewer carbohydrates in total with more dietary fiber.

Our analysis revealed that carbohydrate intake has the greatest influence on MASLD risk; however, the exact reason re-

mains unclear. Moreover, this may be a result of the particularly high-carbohydrate intake in Korea, especially among middle-aged and older adults, due to traditional dietary patterns rich in carbohydrate-heavy foods such as rice and noodles [30,31]. Future research should focus on examining the effects of different carbohydrate types and sources on MASLD through longitudinal studies, and investigate the underlying mechanisms involved.

High-fat diets traditionally influence MASLD development because they promote the accumulation of fat in the liver through increased lipogenesis and impaired fatty acid oxidation [16,32]. However, the type of fat consumed is crucial. In our study, fat intake was identified as the second most significant factor influencing MASLD risk after carbohydrate intake. The CART analysis revealed that participants with lower fat intake had a distinct risk profile. Those in cluster 3—characterized by lower carbohydrate and higher fat intake, particularly with a favorable n-6 to n-3 ratio—showed the lowest incidence of new-onset MASLD. Similar to our findings, findings from a 6-month randomized controlled trial in Denmark revealed that a low-carbohydrate and high-fat diet resulted in greater improvements in NAFLD activity scores than a high-carbohydrate and low-fat diet [27]. These results highlight the importance of fat quality over quantity in preventing MASLD and suggest that balancing fat intake and focusing on healthier fats may be key in MASLD management.

There is a complex relationship between protein intake and MASLD, with some studies suggesting that higher protein intake may help protect against sarcopenia, which is a risk factor for MASLD [19]. However, the outcomes vary based on the type and quality of protein. Plant-based proteins have different effects when compared with animal proteins; however, further research is needed to clarify this in the context of MASLD [19,33]. Our results revealed that cluster 1, characterized by lower protein intake, was significantly associated with a higher risk of MASLD than cluster 3, supporting the idea that protein intake may influence MASLD development. We did not differentiate between protein types in this study; however, a study on older adults in Korea reported that 69% of their protein intake came from plant sources [34], and our findings also reveal that adequate protein intake, particularly in aging populations, may be beneficial for MASLD management. Similarly, a community-based case-control study of Chinese older adults aged ≥ 65 years reported that a daily protein intake of 58.7 to 70.7 g/day was associated with a reduced risk of NAFLD compared

with intake levels below 45.8 g/day [35].

The impact of total caloric intake versus macronutrient composition on MASLD development is still being argued. Some studies suggest that caloric deficit alone can reduce liver fat [36]; however, others highlight that the type of fats and carbohydrates consumed significantly influences MASLD, regardless of calorie intake [14]. Our study supports the latter, highlighting that carbohydrate intake has the greatest influence on MASLD risk, followed by fat, protein, and total calories. These findings emphasize that managing the quality and balance of macronutrients may be more crucial than reducing calories alone. Thus, a dietary approach focusing on macronutrient composition rather than reducing calories may be more effective in preventing and managing MASLD.

This study has several limitations. First, relying on self-reported dietary data introduces potential recall bias. Second, dietary intake was not assessed at every follow-up period in the KoGES study, making it impossible to comprehensively analyze longitudinal dietary changes. Additionally, as only baseline dietary intake data were used, we acknowledge that changes in dietary intake over time may have influenced study outcomes, thereby limiting our ability to evaluate temporal variations in dietary intake. Future studies should consider investigating the impact of dietary intake changes on the incidence of MASLD. Third, unmeasured confounding factors could have influenced the results. Fourth, the lack of detailed nutrient composition analysis, such as distinguishing between refined and unrefined carbohydrates or plant- and animal-based proteins, limits the depth of our findings. Fourth, focusing on a predominantly Korean adult population may affect the generalizability of the results to other ethnicities or regions. Fifth, the use of the NAFLD liver fat score to diagnose MASLD may be less accurate than imaging or biopsy-based methods. While serum-based biomarkers are valuable tools for large-scale epidemiological studies, their diagnostic accuracy remains suboptimal for clinical application, limiting their reliability in precisely identifying MASLD cases [37]. Lastly, the CART algorithm has some limitations, such as converting continuous variables into discrete categories and being sensitive to data variations, which may affect reproducibility [38,39].

Our study findings offer valuable strengths despite its limitations. We used decision trees, random forest models, and clustering analysis to thoroughly assess the impact of macronutrient intake and dietary patterns on MASLD risk. This comprehensive approach makes our study one of the first to explore

these dietary influences in-depth, providing essential insights for developing targeted dietary interventions for MASLD prevention and management.

In conclusion, our study results reveal the critical role of macronutrient composition, particularly carbohydrate intake, in new-onset MASLD development. The findings suggest that optimizing macronutrient quality and balance, rather than simply reducing calorie levels, may be more effective for prevention and management. Specifically, strategies that limit refined carbohydrates and emphasize healthier fats could lead to better clinical outcomes. Therefore, further research is needed to investigate the underlying mechanisms and refine these dietary approaches to improve MASLD management.

SUPPLEMENTARY MATERIALS

Supplementary materials related to this article can be found online at <https://doi.org/10.4093/dmj.2025.0026>.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

AUTHOR CONTRIBUTIONS

Conception or design: Y.C.L., Y.J.K., J.W.L.

Acquisition, analysis, or interpretation of data: all authors.

Drafting the work or revising: Y.C.L., H.S.L., Y.J.K., J.W.L.

Final approval of the manuscript: all authors.

ORCID

Yea-Chan Lee <https://orcid.org/0000-0003-0425-7359>

Hye Sun Lee <https://orcid.org/0000-0001-6328-6948>

Yu-Jin Kwon <https://orcid.org/0000-0002-9021-3856>

Ji-Won Lee <https://orcid.org/0000-0002-2666-4249>

FUNDING

This research received funding from the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPEF) through a high-value-added food technology development program supported by the Ministry of Agriculture, Food, and Rural Affairs (MAFRA) (321030051HD030). It

was also funded by a National Research Foundation of Korea (NRF) grant from the Korean government (MSIT) (RS-2024-00354524).

ACKNOWLEDGMENTS

We would like to thank the Biostatistics Collaboration Unit, Department of Research Affairs, Yonsei University College of Medicine, MID (Medical Illustration & Design), as a member of the Medical Research Support Services of Yonsei University College of Medicine, providing excellent support with medical illustration.

REFERENCES

1. Kanwal F, Neuschwander-Tetri BA, Loomba R, Rinella ME. Metabolic dysfunction-associated steatotic liver disease: update and impact of new nomenclature on the American Association for the Study of Liver Diseases practice guidance on nonalcoholic fatty liver disease. *Hepatology* 2024;79:1212-9.
2. Suzuki H, Tsutsumi T, Kawaguchi M, Amano K, Kawaguchi T. Changing from NAFLD to MASLD: prevalence and progression of ASCVD risk are similar between NAFLD and MASLD in Asia. *Clin Mol Hepatol* 2024;30:577-9.
3. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *J Hepatol* 2019;70:151-71.
4. Li R, Li H, Ye X, Qin J. Metabolic dysfunction-associated steatotic liver disease is becoming the leading driver of the burden of cirrhosis in China: results from the Global Burden of Disease Study 2019. *J Clin Gastroenterol* 2025;59:592-7.
5. Kim D, Li AA, Gadiparthi C, Khan MA, Cholankeril G, Glenn JS, et al. Changing trends in etiology-based annual mortality from chronic liver disease, from 2007 through 2016. *Gastroenterology* 2018;155:1154-63.
6. Fahed G, Aoun L, Bou Zerdan M, Allam S, Bou Zerdan M, Bouferraou Y, et al. Metabolic syndrome: updates on pathophysiology and management in 2021. *Int J Mol Sci* 2022;23:786.
7. Feng G, Targher G, Byrne CD, Yilmaz Y, Wai-Sun Wong V, Adithya Lesmana CR, et al. Global burden of metabolic dysfunction-associated steatotic liver disease, 2010 to 2021. *JHEP Rep* 2024;7:101271.
8. Lee NH, Jeong SJ, Wang JH, Choi YJ, Oh HM, Cho JH, et al. The clinical diagnosis-based nationwide epidemiology of metabolic dysfunction-associated liver disease in Korea. *J Clin Med* 2023;12:7634.

9. Dongiovanni P, Valenti L. A nutrigenomic approach to non-alcoholic fatty liver disease. *Int J Mol Sci* 2017;18:1534.
10. Meroni M, Longo M, Rustichelli A, Dongiovanni P. Nutrition and genetics in NAFLD: the perfect binomium. *Int J Mol Sci* 2020;21:2986.
11. Sandireddy R, Sakthivel S, Gupta P, Behari J, Tripathi M, Singh BK. Systemic impacts of metabolic dysfunction-associated steatotic liver disease (MASLD) and metabolic dysfunction-associated steatohepatitis (MASH) on heart, muscle, and kidney related diseases. *Front Cell Dev Biol* 2024;12:1433857.
12. Parry SA, Hodson L. Influence of dietary macronutrients on liver fat accumulation and metabolism. *J Investig Med* 2017;65: 1102-15.
13. Chan WK, Chuah KH, Rajaram RB, Lim LL, Ratnasingam J, Vethakkan SR. Metabolic dysfunction-associated steatotic liver disease (MASLD): a state-of-the-art review. *J Obes Metab Syndr* 2023;32:197-213.
14. Nemer M, Osman F, Said A. Dietary macro and micronutrients associated with MASLD: analysis of a national US cohort database. *Ann Hepatol* 2024;29:101491.
15. Schwimmer JB, Ugalde-Nicalo P, Welsh JA, Angeles JE, Cordeiro M, Harlow KE, et al. Effect of a low free sugar diet vs usual diet on nonalcoholic fatty liver disease in adolescent boys: a randomized clinical trial. *JAMA* 2019;321:256-65.
16. Vacca M, Kamzolas I, Harder LM, Oakley F, Trautwein C, Hatting M, et al. An unbiased ranking of murine dietary models based on their proximity to human metabolic dysfunction-associated steatotic liver disease (MASLD). *Nat Metab* 2024;6: 1178-96.
17. Khoo J, Hsiang J, Taneja R, Law NM, Ang TL. Comparative effects of liraglutide 3 mg vs structured lifestyle modification on body weight, liver fat and liver function in obese patients with non-alcoholic fatty liver disease: a pilot randomized trial. *Diabetes Obes Metab* 2017;19:1814-7.
18. Properzi C, O'Sullivan TA, Sherriff JL, Ching HL, Jeffrey GP, Buckley RF, et al. Ad libitum Mediterranean and low-fat diets both significantly reduce hepatic steatosis: a randomized controlled trial. *Hepatology* 2018;68:1741-54.
19. Wong R, Yuan LY. Sarcopenia and metabolic dysfunction associated steatotic liver disease: time to address both. *World J Hepatol* 2024;16:871-7.
20. Kotronen A, Peltonen M, Hakkarainen A, Sebastianova K, Bergholm R, Johansson LM, et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* 2009;137:865-72.
21. Ahn Y, Kwon E, Shim JE, Park MK, Joo Y, Kimm K, et al. Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *Eur J Clin Nutr* 2007;61: 1435-41.
22. The Korean Nutrition Society. Recommended dietary allowances for Koreans. 7th ed. Seoul: The Korean Nutrition Society; 2000.
23. Gillespie J. "You are what you eat": the role of dietary macronutrients and micronutrients in MAFLD. *Clin Liver Dis (Hoboken)* 2021;18:67-71.
24. Romero-Gomez M. NAFLD and NASH: biomarkers in detection, diagnosis and monitoring. Cham: Springer; 2020. Chapter 4, Geometry of nutrition: nutrients and NAFLD progression; p49-67.
25. Afsharfar M, Salimi Z, Aminnezhad Kavkani B, Shekari S, Abbastorki S, Majidi N, et al. Association of nonalcoholic fatty liver disease with the different types of dietary carbohydrates: a cross-sectional study. *J Diabetes Metab Disord* 2023;22:1139-43.
26. Browning JD, Baker JA, Rogers T, Davis J, Satapati S, Burgess SC. Short-term weight loss and hepatic triglyceride reduction: evidence of a metabolic advantage with dietary carbohydrate restriction. *Am J Clin Nutr* 2011;93:1048-52.
27. Hansen CD, Gram-Kampmann EM, Hansen JK, Hugger MB, Madsen BS, Jensen JM, et al. Effect of calorie-unrestricted low-carbohydrate, high-fat diet versus high-carbohydrate, low-fat diet on type 2 diabetes and nonalcoholic fatty liver disease: a randomized controlled trial. *Ann Intern Med* 2023;176:10-21.
28. Orliacq J, Perez-Cornago A, Parry SA, Kelly RK, Koutoukidis DA, Carter JL. Associations between types and sources of dietary carbohydrates and liver fat: a UK Biobank study. *BMC Med* 2023;21:444.
29. Sebastianova K, Santos A, Kotronen A, Hakkarainen A, Makkonen J, Silander K, et al. Effect of short-term carbohydrate overfeeding and long-term weight loss on liver fat in overweight humans. *Am J Clin Nutr* 2012;96:727-34.
30. Yoo H, Jo U, Park K. Carbohydrate intake levels and the risk of metabolic syndrome in Korean populations: a prospective study. *Nutrients* 2024;16:2440.
31. Song Y, Joung H. A traditional Korean dietary pattern and metabolic syndrome abnormalities. *Nutr Metab Cardiovasc Dis* 2012;22:456-62.
32. Szudzik M, Hutsch T, Chabowski D, Zajdel M, Ufnal M. Normal caloric intake with high-fat diet induces metabolic dysfunction-associated steatotic liver disease and dyslipidemia without

obesity in rats. *Sci Rep* 2024;14:22796.

33. Singh A, Buckholz A, Kumar S, Newberry C. Implications of protein and sarcopenia in the prognosis, treatment, and management of metabolic dysfunction-associated steatotic liver disease (MASLD). *Nutrients* 2024;16:658.
34. Lee YS, Lee Y. Regional differences in protein intake and protein sources of Korean older adults and their association with metabolic syndrome using the 2016–2019 Korea National Health and Nutrition Examination Surveys: a cross-sectional study. *Korean J Community Nutr* 2024;29:173-88.
35. Li X, Ma W, Yang T, Wang C, Zhang W, Li H, et al. Higher intakes of lysine, threonine and valine are inversely associated with non-alcoholic fatty liver disease risk: a community-based case-control study in the Chinese elderly. *Food Sci Hum Well-*
ness 2024;13:191-7.
36. Haufe S, Engeli S, Kast P, Bohnke J, Utz W, Haas V, et al. Randomized comparison of reduced fat and reduced carbohydrate hypocaloric diets on intrahepatic fat in overweight and obese human subjects. *Hepatology* 2011;53:1504-14.
37. Reinson T, Buchanan RM, Byrne CD. Noninvasive serum biomarkers for liver fibrosis in NAFLD: current and future. *Clin Mol Hepatol* 2023;29(Suppl):S157-70.
38. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *J Clin Epidemiol* 2001;54:603-9.
39. Henrard S, Speybroeck N, Hermans C. Classification and regression tree analysis vs. multivariable linear and logistic regression methods as statistical tools for studying haemophilia. *Haemophilia* 2015;21:715-22.

SUPPLEMENTARY METHODS

Covariates

Body mass index (BMI) was defined as the ratio of body weight (kg) to height (squared m²). Furthermore, waist circumference was measured at the middle of the line between the lowest rib and the topline of the iliac crest after the participants exhaled and were in a relaxed state without applying pressure to the skin. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were estimated as the mean value after measuring three times at 2 to 3 cm above the antecubital fossa, with the participants calm for approximately 5 minutes before the measurement. Serum levels of total cholesterol, triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), fasting plasma glucose (FPG), glycosylated hemoglobin (HbA1c), fasting insulin, aspartate aminotransferase, and alanine aminotransferase were analyzed using the blood sample test obtained after 8 hours of overnight fasting. Additionally, serum low-density lipoprotein cholesterol (LDL-C) level was derived by calculation using the Friedewald formula: Total cholesterol (mg/dL) minus HDL-C (mg/dL) minus one-fifth of TG (mg/dL) [1].

Participants who have smoked more than 100 cigarettes in their lifetime and are currently smoking at least one cigarette per day are defined as current smoking status. Furthermore, current drinking was defined as the status in which a person currently drinks alcohol at least once a month. Physical activity was measured in terms of metabolic equivalent of task (MET)-hours per week (MET-hr/week) by multiplying the weekly hours reported for each type of activity by the respective MET values: sedentary (0 METs), very light (1.5 METs), light (3 METs), moderate

(5 METs), and vigorous (7 METs) [2]. The three categories of physical activity variable were defined as low for less than 7.5 MET-hr/week, moderate for 7.5–30 MET-hr/week, and high for over 30 MET-hr/week, respectively. We defined obesity as BMI ≥ 25 kg/m² according to the Asia-Pacific classification [3]. Additionally, diabetes mellitus was defined based on at least one of the following criteria; FPG level of ≥ 126 mg/dL, serum glucose level of ≥ 200 mg/dL at 2-hour 75 g oral glucose tolerance test, serum HbA1c level of $\geq 6.5\%$, and current treatment with anti-diabetic agents or insulin therapy. Hypertension was characterized as SBP of ≥ 140 mm Hg DBP of ≥ 90 mm Hg, or current treatment with antihypertensive agents. Lastly, dyslipidemia was defined as the presence of at least one of the following criteria: total cholesterol ≥ 240 mg/dL, TG ≥ 200 mg/dL, HDL-C < 40 mg/dL, LDL-C ≥ 160 mg/dL, or current use of dyslipidemia medications.

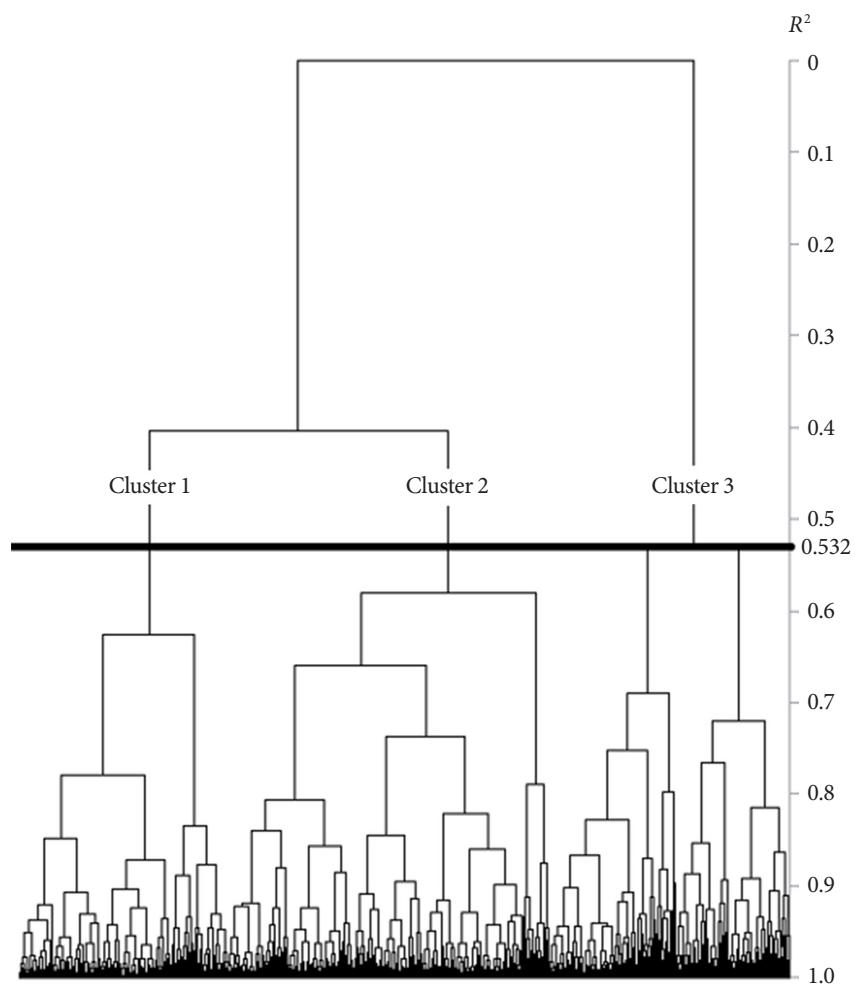
SUPPLEMENTARY REFERENCES

1. Roberts WC. The Friedewald-Levy-Fredrickson formula for calculating low-density lipoprotein cholesterol, the basis for lipid-lowering therapy. *Am J Cardiol* 1988;62:345-6.
2. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000; 32(9 Suppl):S498-504.
3. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;363:157-63.

Supplementary Table 1. Determination of the optimal number of clusters

No. of clusters	Pseudo t^2 statistics ^a	R^2 ^b	Semipartial R^2 ^c
2	1,997.7	0.404	0.128
3	545.8	0.532	0.050
4	820.8	0.582	0.044
5	1,115.0	0.626	0.030

^aPseudo t^2 statistic identifies an optimal cluster number of clusters when a minimal value is followed by a notably larger one, ^b R^2 values estimate the heterogeneity among the clusters in each level of cluster solution, and the higher value of R^2 means more heterogeneous one another in the cluster solution, ^cSemipartial R^2 was calculated as subtraction of R^2 value at a given step from R^2 value at next step cluster solution. A larger value of semipartial R^2 means that the value is obtained from merging of more heterogenous clusters.



Supplementary Fig. 1. Dendrogram presenting the hierarchical cluster analysis made by the agglomerative technique, which began with each subject being a cluster by itself and merged together continuously based on similarity between clusters. When the number of clusters was 3, R^2 value, which estimated the heterogeneity of the cluster solution formed at a given step, was 0.532.